Università degli Studi di Padova

Sede Amministrativa: Università degli Studi di Padova
Dipartimento di Matematica "Tullio Levi-Civita"

Corso di dottorato di ricerca in Scienze Matematiche
Curriculum: Matematica Computazionale
XXXVIII ciclo

Topics in random graph theory: limit theorems and mixing times of random walks

Coordinatore:
Ch.mo Prof. Giovanni Colombo

Supervisora:
Prof.ssa Alessandra Bianchi

Co-Supervisore:
Dr. Matteo Quattropani

Dottorando:
Giacomo Passuello

Anno Accademico 2024/2025

Abstract

This doctoral dissertation is concerned with the study of static and dynamical observables of random graphs, with a particular attention to concentration phenomena and phase transitions. The thesis is divided in two parts.

The first part of the thesis is devoted to the asymptotic analysis of two types of static graph observables. In the first chapter, we analyze the edge-triangle model, a random graph exhibiting dependence among edges. We prove concentration of the triangle count and, for some approximations of the model, we obtain more refined results including standard and non-standard central limit theorems, depending on the value of the parameters. In a mean-field setting, our results are supported by simulations. The results rely on large deviation principles and the analyticity properties of the free energy of the model. In the second chapter we consider a family of inhomogeneous directed random graphs, which includes the Chung–Lu directed graph and stochastic block models, and we consider their adjacency matrices. We establish the existence of eigenvalues outside the bulk of their spectrum, for which we prove Gaussian fluctuations. The results are based on the trace method and a perturbative analysis.

In the second part of the thesis, we consider the simple random walk (SRW) on directed random graphs, we characterize its mixing properties and we give particular emphasis to the cutoff phenomenon. In the third chapter we analyze the SRW on the Chung–Lu directed graph. For this dynamics we prove the occurrence of the cutoff phenomenon at entropic time. We characterize the size of the cutoff window, in which the total variation profile converges to a universal Gaussian shape, independent of the parameters. Finally, the fourth chapter is concerned with the mixing properties of the SRW on a directed graph exhibiting a community structure, corresponding to a directed version of the stochastic block model. For this model we show the occurrence of a mixing trichotomy, related to the strength of the community structure. In particular, we identify three mixing regimes, where the cutoff survives to the bottleneck perturbation or is substituted by an exponential relaxation, and we provide a first-order characterization of the total variation profile. A substantial part of the analysis is given by a control on the homogenization of the random environment.

Sommario

Questa tesi dottorale si occupa dello studio asintotico di osservabili statiche e dinamiche nei modelli di grafi aleatori, con particolare attenzione ai fenomeni di concentrazione e alle transizioni di fase. La tesi è suddivisa in due parti.

La prima parte è dedicata all'analisi asintotica di due tipi di osservabili su grafi. Nel primo capitolo analizziamo il modello edge-triangle, un grafo casuale che presenta dipendenze tra gli archi. Dimostriamo la concentrazione del numero di triangoli e, per alcune approssimazioni del modello, otteniamo risultati più raffinati, che includono teoremi di limite centrale, sia standard che non standard, a seconda del valore dei parametri. In un contesto di tipo mean-field, i nostri risultati sono supportati da simulazioni. L'analisi si basa su principi di grandi deviazioni e sulle proprietà di analiticità dell'energia libera del modello. Nel secondo capitolo consideriamo una famiglia di grafi casuali diretti inomogenei, che include il grafo di Chung–Lu e lo stochastic block model, e ne studiamo la matrici di adiacenza: dimostriamo la presenza di autovalori al di fuori del nucleo principale, per i quali otteniamo fluttuazioni Gaussiane. I risultati si basano sul metodo della traccia e su un'analisi perturbativa.

La seconda parte della tesi riguarda l'evoluzione della passeggiata aleatoria semplice su grafi casuali diretti. In particolare, ne caratterizziamo le proprietà di mescolamento, dando enfasi al fenomeno del cutoff. Nel terzo capitolo analizziamo la passeggiata aleatoria semplice sul grafo diretto di Chung–Lu. Per questa dinamica dimostriamo l'occorrenza del cutoff a tempo entropico. Caratterizziamo inoltre la dimensione della finestra di cutoff, all'interno della quale il profilo della variazione totale converge a una forma gaussiana universale. Infine, il quarto capitolo riguarda le proprietà di mescolamento della passeggiata aleatoria semplice su un grafo diretto con comunità, corrispondente a una versione diretta dello stochastic block model. Per questo modello, mostriamo l'esistenza di una tricotomia nel comportamento di mescolamento, legata all'intensità della struttura di comunità. In particolare, identifichiamo tre regimi distinti, in cui il cutoff sopravvive alla perturbazione oppure cede il posto a un rilassamento esponenziale, e forniamo una caratterizzazione al primo ordine del profilo della variazione totale. Una parte sostanziale dell'analisi è data da un controllo sull'omogeneizzazione dell'ambiente casuale.

Contents

Useful notation

$S^{\complement}$	complement of set S
$\mathbf{1}_S$	indicator function of set S
δ_x	delta distribution concentrated at point x
$\mathrm{Ber}(p)$	Bernoulli distribution of parameter p
$\mathrm{Bin}(n,p)$	binomial distribution of parameters n, p
$\mathrm{Pois}(\lambda)$	Poisson distribution of parameter $\lambda > 0$
$\mathrm{Exp}(\lambda)$	exponential distribution of parameter $\lambda > 0$
$\mathcal{N}(\mu, \sigma^2)$	Gaussian distribution of mean μ and variance σ^2

$a_n = o(b_n)$ or $a_n \ll b_n$	$a_n/b_n \xrightarrow[n\to+\infty]{} 0$
$a_n = O(b_n)$ or $a_n \lesssim b_n$	$\exists\ \ell \in (0,+\infty)$ s.t. $a_n/b_n \leq \ell$, for large n
$a_n = \Omega(b_n)$	$b_n = O(a_n)$
$a_n = \Theta(b_n)$	$a_n = O(b_n)$ and $b_n = O(a_n)$
$a_n \gg b_n$	$b_n = o(a_n)$
$a_n \sim b_n$	$a_n/b_n \xrightarrow[n\to+\infty]{} 1$
$a_n \asymp b_n$	$\exists\ \ell \in (0+\infty)$ s.t. $a_n \sim \ell \cdot b_n$

$\mathbb{P}$	probability measure associated to graph environment
$\mathbb{E}$	expectation w.r.t. graph environment
A_n holds w.h.p.	$\mathbb{P}(A_n^{\complement}) = o(1)$, 2
A_n holds w.v.h.p.	$\exists\ \eta > 1$ s.t. $\mathbb{P}(A_n^{\complement}) \leq e^{-\log(n)^\eta}$, 46
$X_n = o_{\mathbb{P}}(Y_n)$	$\mathbb{P}(\lvert X_n/Y_n\rvert > \delta) = o(1)$ for every $\delta > 0$, 46
$X_n = O_{\mathbb{P}}(Y_n)$	$\mathbb{P}(\lvert X_n/Y_n\rvert > K) = o(1)$ for some $K > 0$, 46
$X_n = o_{v.h.\mathbb{P}}(Y_n)$	$\exists\ \eta > 1$ s.t. $\mathbb{P}(\lvert X_n/Y_n\rvert > \delta) \leq e^{-\log(n)^\eta}$ for every $\delta > 0$, 46
$X_n = O_{v.h.\mathbb{P}}(Y_n)$	$\exists\ \eta > 1$ s.t. $\mathbb{P}(\lvert X_n/Y_n\rvert > K) \leq e^{-\log(n)^\eta}$ for some $K > 0$, 46
$G = (V,E)$	realization of graph environment
$\mathbf{P}^G$	quenched measure associated to simple random walk, 83
$\mathbb{P}^{\mathrm{an}} = \mathbb{E}[\mathbf{P}^G]$	annealed measure 83
$P^t(x,\cdot)$	law of simple random walk started at x, 83

$t(H,G)$	homomorphism density of H in G, 9
$E_n(G)$	number of edges in G, 10
$T_n(G)$	number of triangles in G, 10
$\mathcal{H}_{n;\alpha,h}$	Hamiltonian of edge-triangle model, 10
$f_{n;\alpha,h}$, $f_{\alpha,h}$	finite- and infinite-size free energy, 10
$\mathcal{U}^{rs}$	unicity region in the replica symmetric regime, 11
$\mathcal{M}^{rs}$	multiplicity region in the replica symmetric regime, 11
w_x^-, w_x^+	in- and out-weight in the Chung–Lu graph
$\boldsymbol{w}$	total weight
$p_{x,y}$	connection probability of edge (x,y)
D_x^-, D_x^+	in- and out-degree of a vertex x
δ_+, Δ_+	minimum and maximum out-degree
$\lambda_1(M_n),\dots,\lambda_n(M_n)$	eigenvalues of M_n in decreasing modulus order
$\sigma_1(M_n),\dots,\sigma_n(M_n)$	singular values of M_n in decreasing order
$\mathcal{B}(x,\varepsilon)$	complex ball of center x and radius ε
μ_{M_n}	empirical spectral distribution of M_n, 66
$\boxplus$, $\boxtimes$	free additive and multiplicative convolution, 69
$\mathbf{m}(\mathfrak{p})$	probability mass of a path $\mathfrak{p}$, 83
$\lVert\cdot\rVert_{TV}$	total variation distance, 84
$\boldsymbol{t}^{(x)}_{\text{mix}}$	mixing time of parameter $\varepsilon>0$ starting form x, 84
μ_{in}	in-degree distribution, 84
H	entropy, 85
$\boldsymbol{t}_{\text{ent}}$	entropic time, 86
$\mathbf{w}_{\text{ent}}$	cutoff window, 86
$\mathcal{B}^-(h)$, $\mathcal{B}^+(h)$	discrete in- and out-neighborhood of radius h, 97
α	re-wiring parameter, 122
I_x^+, O_x^+	internal and external out-degree of a vertex x, 123
$\mathcal{G}_i$	set of gates in i-th community, 140
τ_{jump}	first community-jump time, 135
τ_S	hitting time of a set S, 141
$\mu^\star$	quasi-stationary distribution, 142

Introduction

Complex networks represent a ubiquitous and modern tool for modeling real-world systems. They are natural to describe, and they are perfect for applications in biological, social sciences, computer science, and economics. The enormous list of available examples includes collaboration or social networks, webpage and internet architectures, microscopical structures describing proteins and polymers, transport and energy grids (see the reviews Albert and Barabási [2002], Newman [2003]). All these instances are given by a finite but large number of elements. More remarkably, although they come from completely different environments, they often share relevant common features, such as inhomogeneity, scale-freeness, and small-world properties.

These common tendencies have led to the possibility of modeling a large class of networks through simplified probabilistic models. In this sense, random graphs represent a young and versatile field of research that occupies a central role in the modern mathematical framework. In the last three decades, there has been an explosion in the number of contributions in this field, and its growth rate has been substantially stable in recent years. Many models have been designed to reproduce various asymptotic features as the size of the network grows. For example: Stochastic Block Models exhibit community structures and are used for community detection purposes (see Abbe [2018]); Chung-Lu graphs (see Chung and Lu [2002b]) and other inhomogeneous graphs provide networks with huge hubs; preferential attachment models describe small worlds (see Albert and Barabási [1999]); exponential random graphs exhibit prescribed expected local densities and capture clustering (see Strauss [1986]). See the books Bollobás [2001], van der Hofstad [2016, 2024], Newman [2010] for an introduction to random graphs.

A crucial aspect for most of these models lies in the dependence on external parameters, which can originate many non-trivial and interesting behaviors. In fact, the large size of random networks allows them to exhibit an important feature, called complexity. A system with a finite number of agents is said to be complex if, despite a simple formulation, the behavior of its macroscopic observables does not depend linearly on external parameters and cannot be easily determined as the number of agents grows. In particular, in the thermodynamic limit, where the "volume" of the systems diverges, slight variations in the parameters can produce huge qualitative and quantitative changes. Such thresholds in the nature of a system are referred to as phase transitions, and in the literature, there are plenty of probabilistic models that present such behavior. Their study has become a great source of mathematical problems, giving birth to an actual branch of discrete mathematics lying at the interface between probability and theoretical physics. Among the most famous examples in statistical physics, it is worth mentioning lattice systems such as the Ising model and percolation. Analogous threshold phenomena arise naturally in the context of random graphs regarding global properties, such as the size of connected components component, typical distances, local features, such as the density of certain substructures, and even dynamical observables.

In general, a random graph is a sequence of random variables, say $(\mathbb{G}_n)_{n\in\mathbb{N}}$, defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For each $n \in \mathbb{N}$, $\mathbb{G}_n$ takes values in a suitable space $\mathcal{G}_n$ of graphs with n vertices and satisfies suitable consistency conditions. Equivalently, a random graph can be seen as a choice, for every $n \in \mathbb{N}$, of a probability distribution on $\mathcal{G}_n$. This also defines a probability measure $\mathbb{P}_n$. Graphs can be simple or have multi-edges, they can be undirected or oriented, and the right specification of the graph set allows random graphs to cover a wide range of applications.

Although a rigorous formalism is more than satisfactory, it is sometimes more intuitive to define random graphs via algorithmic stochastic procedures. The easiest example of a random graph is the Erdős–Rényi random graph. In this model a parameter $p \in [0,1]$ is fixed and edges are included in an empty graph with vertex set $[n] \coloneqq \{1, 2, \ldots, n\}$, independently and with probability p. This procedure results in a given simple graph G with $|E|$ edges with probability $p^{|E|}(1-p)^{\binom{n}{2}-|E|}$. Although very simple to define and interesting to study, the Erdős–Rényi model constitutes a pure mathematical abstraction rather than a physically descriptive model.

Firstly, the Erdős–Rényi graph exhibits edge independence. This property facilitates the achievement of exact limit statements, but it is not present in real–work networks. This can be easily motivated, for example, in the context of social networks; if two people share a common friend or interest, it is more likely that they form a connection themselves. Exponential random graphs, which are part of the contents of this dissertation, go beyond this constraint, allowing edge dependence and allowing the probability density associated with the graph to depend on prescribed fixed subgraph densities.

Secondly, the Erdős–Rényi graph is a homogeneous random graph, which is another unrealistic property for most networks. For this reason, many generalizations of this model have been defined over the years, which keep the independence of edges. An example, which will be central in this thesis, is the Chung–Lu random graph, which was proposed and analyzed in Lu [2002] and Chung and Lu [2002b,c, 2003, 2006a,b], and which is defined as follows. Given a sequence of positive weights $(\widetilde{w}_x)_{x\in[n]}$, corresponding to the choice of expected degrees, two distinct vertices $x, y \in [n]$ are connected by an edge, independently and with probability

$$p_{xy}^{(\mathrm{CL})} = p_{yx}^{(\mathrm{CL})} = \frac{\widetilde{w}_x \widetilde{w}_y}{\sum_{z\in[n]} \widetilde{w}_z} \wedge 1.$$

This choice allows to model inhomogeneity, since weights correspond to the degree tendency of vertices. The model is easily adapted to a directed framework, which is more suitable for applications to the World Wide Web. Here, two sequences $(\widetilde{w}_x^-)_{x\in[n]}$, $(\widetilde{w}_x^+)_{x\in[n]}$ with equal sum are taken, and the connection probability is set to

$$p_{xy}^{(\mathrm{DCL})} = \frac{\widetilde{w}_x^+ \widetilde{w}_y^-}{\sum_{z\in[n]} \widetilde{w}_z^+} \wedge 1.$$

In this dissertation, we present some contributions to the study of random graphs and the convergence to the equilibrium of random dynamics in such environments. The thesis is articulated in two parts that tackle problems related to scaling limits, static and dynamical phase transitions of random graphs. Both parts are preceded by an introduction and each chapter is self-contained. Although the main mathematical objects have a consistent notation in the whole manuscript, we warn the reader that there may be some changes and adaptations from chapter to chapter. For instance, this will be the case for the notation of the Chung–Lu directed graph.

Part I is focused on the asymptotic properties of the two above generalizations of the Erdős–Rényi random graph, which are shaped by two different constraints specifications. In both cases, we are interested in the study of the asymptotics of relevant observables.

In Chapter 1, we consider a model belonging to the family of Exponential Random Graphs. These graphs generalize the Erdős–Rényi graph by introducing a dependency between edges, which is modeled by a Gibbs distribution. In this setting, the probability of a specific graph display is a parametric function of certain subgraph counts. This corresponds to fixing the expected value of certain subgraph densities and making certain local structures more or less probable. For this model, we study the asymptotic behavior of the triangle density, providing a law of large numbers. Then, we consider two approximations of the model, exhibiting the same phase diagram, for which we prove a central limit theorem for the triangle density.

In Chapter 2 we consider a family of inhomogeneous directed random graphs, which includes the Chung-Lu directed random graph and stochastic block models. For these models, in a non-sparse setting, we study the spectrum of the adjacency matrix, establishing the presence and the typical scaling the eigenvalues outside the bulk. Then, we provide a CLT for their fluctuations.

In Part II we are interested in studying random dynamics on random graphs. In particular, we analyze the mixing properties of simple random walks on random digraphs (i.e., directed graphs).

In Chapter 3, we study the mixing time of the simple random walk on the Chung–Lu directed graph. In this setting, we prove that with high probability, the simple random walk exhibits a cutoff. Moreover, we determine the presence of a window with Gaussian shape. Both statements hold uniformly in the starting position, namely for the worst-case and best-case scenarios.

In Chapter 4, we study the mixing time of the simple random walk on a directed stochastic block model. This random graph is inspired by the classical stochastic block model and exhibits a community structure. Building on the results coming from Chapter 3, and some technical analysis of the first community jump time, we establish a mixing trichotomy, namely the existence of three mixing regimes. In particular, depending on the strength of the community structure, there is a fast mixing phase where the random walk exhibits a cutoff at entropic time, a slow mixing phase, where the random walk has a limit profile with exponential shape, and a critical phase, where the two mechanisms enter in competition.

We conclude the introduction with a few comments on the chosen regimes.
The specification of exponential random graph that we consider is suitable for the dense setting. The graphon formalism and the overall approach that will be used later is meaningful for graphs such that parameters and average edge densities are bounded away from zero. Sparse instances of this model may be considered by letting parameters vanishing as the number of vertices grows, or passing to a micro-canonical setting with a conditioned number of edges. However, these are typically harder to study.
For the analysis of the directed Chung–Lu model, we consider a non-sparse regime where the average degree grows at least polylogarithmically in the size of the graph. This of course includes the dense regime. This condition is required for the appearance of outliers in the spectrum of the adjacency matrix and to provide a control on error terms, in accordance with the undirected setting. For what concerns the mixing analysis, a dense regime would constitute a huge speeding factor, since the random walk is expected to mix in a constant number of steps. On the other side, to have a well defined problem, the simple random walk has to be irreducible, and in a directed setting this happens when average degrees are at least logarithmic, which is our assumption.

Part I

Asymptotic properties of random graphs

Introduction to Part I

Given an instance of a large real-world network, identifying the best theoretical model that captures its features is a crucial task. Understanding the asymptotic properties of a random graph is then very useful for applicative purposes.

As their size grows, graph statistics, despite their discrete nature, can converge to continuous limits, and there is high interest in establishing their existence and features, and the same can happen for graphs themselves. For instance, dense graphs can converge to graphons, namely functions $g : [0,1]^2 \to [0,1]$, which can be endowed with a metric structure (see Borgs et al. [2008, 2012] and Lovász [2012]), while sparse ones are well described by local weak limits, which make use of rooted infinite trees to describe the neighborhood of a fixed vertex in the limiting structure (see Benjamini and Schramm [2001] and van der Hofstad [2024]).

Concerning graph observables, many issues can be considered and examples involve: metric properties, such as diameter and typical distances; combinatorial properties, related to chromatic numbers and matching issues; global properties, linked to connected components, expansion, and bottlenecks (which can influence dynamics); local properties, such as degree distributions, clustering coefficients, and subgraph counts. The interest is usually quantifying their asymptotic order and the nature of their fluctuations around the mean. Laws of large numbers, Central limit theorems, and large deviation principles constitute the main target of the analysis. Sometimes, some graph properties do not hold asymptotically almost surely. In that case, we may be interested in proving that a statement holds with high probability. Given, for every $n \in \mathbb{N}$, a subset $A_n \subseteq \mathcal{G}_n$, we say that a family of events, or simply an event, $(A_n)_{n\in\mathbb{N}}$ happens with high probability (shortly w.h.p.) if

$$\mathbb{P}(G_n \in A_n^{\complement}) \xrightarrow[n\to+\infty]{} 0,$$

where $A_n^{\complement}$ denotes the complement set. In Part I we will analyze spectral observables and subgraph densities in two different random graph ensembles.

Phase transitions in random graphs

As mentioned, the study of limit properties of networks constitutes an extremely wide research area, where phase transitions are very likely to take place. The most famous example is, with no doubt, the one undergone by the connectivity properties of the Erdős–Rényi graph. In that setting, a relevant scale for the connection probability is given by the choice $p = n^{-a}$, for $a > 0$. Certain values of $a > 0$ have been shown to be thresholds for certain (monotone) properties. In particular, $a = 1$ has been shown to be the threshold for the presence, with high probability, of triangles and cycles. Moreover, in the regime $p = \lambda/n$, where $\lambda > 0$ is constant, $\lambda = 1$ has been shown to be a threshold for the existence, with high probability, of a connected component of linear size Erdős and Rényi [1960]. See Frieze and Karoński [2016], van der Hofstad [2016] for a description of this probabilistic threshold.

A large deviation principle (LDP), is one of the tools that can reveal the presence of a phase transition. A LDP corresponds to the determination of a functional, called rate function, which provides exponential asymptotics for the probability of a family of events. Studying the minima of the rate function, it is possible to achieve information on typical realizations. Changes in the number of minima, or their form, are a sign that a threshold has been attained. See Chatterjee [2016] for an introduction to LDPs for random graphs. An important contribution in this direction has been given in Chatterjee and Varadhan [2011], where a LDP for the Erdős–Rényi random graph is derived. Here, as a byproduct, a LDP for the number of triangles is provided, and the presence of a double transition for the minima of the rate function can be read. We also mention other instances of notable LDPs: in Andreis et al. [2021, 2023] a LDP for the empirical measure of the size of connected components in sparse random graphs is derived; in Chakrabarty et al. [2022] a LDP for the largest eigenvalue of a dense inhomogeneous Erdős–Rényi graph is given.

Subgraph densities in exponential random graphs

One question among sociologists is to understand how local communities can affect the overall network structure. For this reason, the study of subgraph densities constitutes a usual target in the context of random graphs, which provides good insights on the network. For example, a high number of triangles implies a high local density and that the network is clustered. The presence of other structures, such as squares without diagonal elements, could be typical of more regular lattice geometries.

Exponential random graphs (ERGs) provide a framework where expected subgraph densities are controlled by external parameters, leading to a flexible modeling. In this class of random graphs, introduced in Strauss [1986], the probability measure over the state space is designed to enhance or decrease the probability of certain structures, biasing their occurrence. Following a statistical mechanics approach, the bias is encoded by a function called Hamiltonian, contained in an exponential term, and the probability measure is then a Gibbs distribution. In particular, let $\mathcal{G}_n$ denote the set of simple graphs with n vertices. For a fixed number $k \in \mathbb{N}$ of simple graphs $H_1, \dots, H_k$ and a vector of (possibly negative) parameters $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k) \in \mathbb{R}^k$, the associated exponential random graph has law given by a probability distribution on $\mathcal{G}_n$ defined by

$$\mathbb{P}_n(G) = \mathbb{P}_{n;\boldsymbol{\beta}}(G) \propto \exp\left(n^2 \sum_{i=1}^{k} \beta_i t(H_i, G)\right), \qquad G \in \mathcal{G}_n,$$

where for $i = 1, \dots, k$, $t(H_i, G)$ is the homomorphism density of H in G (see Subsection 1.1.1). Due to their versatility, exponential random graphs are widely used for the analysis and modeling of social multi-layered networks (see Caimo and Gollini [2023], Harris

[2013], Lusher et al. [2012]). However, statistical estimation is a difficult theoretical problem. See Fischer et al. [2025], Xu and Reinert [2021], based on Reinert and Ross [2019], where Stein's method is employed to this aim. Many important and rigorous results have been obtained on this model so far in Chatterjee and Dey [2010], Chatterjee and Diaconis [2013], sometimes imposing constraints on subgraph densities Kenyon and Yin [2017], Neeman et al. [2023]. In particular, in Chatterjee and Diaconis [2013] the authors derive a LDP for the model, showing that in a certain regime of parameters, called "replica symmetric regime", minimizers of the rate function are given by constant graphons. Using this result, Radin and Yin [2013] characterized the phase diagram of the model, showing the presence of a parametric curve, ending in a critical second-order point, where the rate function admits two different minima.

There is an active line of literature that is concerned with proving limit theorems for sums of dependent variables. In the the framework of ERGs, a Central Limit Theorem (CLT) for the edge density was first established in Mukherjee and Xu [2023] for the two-star model, a class of ERGs in which edge dependencies arise from the presence of two-stars (i.e., subgraphs consisting of three vertices with two edges sharing a common vertex). For this model, correlation inequalities have been derived in Bianchi et al. [2022]. Subsequently, Bianchi et al. [2024] obtained the first CLT for the edge-triangle model, a specific class of ERGs whose Hamiltonian depends exclusively on the edge and triangle densities. Their approach employs the analyticity properties of the free energy density. A LLN and a CLT for the edge density were derived, together with a non-standard CLT at the critical point for a mean-field approximation of the model. Their result was later strengthened in the work of Fang et al. [2025a], which not only generalizes, via Stein's method (see Stein [1972]), the result of Bianchi et al. [2024] to a broader class of ERGs, but also makes the CLT quantitative in terms of both the Kolmogorov and the Wasserstein distance. In the latter, the authors apply Stein's method Stein [1972] to prove a CLT for general subgraph counts. This is a powerful technique, also used for spin systems in Ellis and Newman [1978], Eichelsbacher and Löwe [2010], which provides a normal approximation to the distribution of dependent random variables. This approach often requires higher-order concentration inequalities for controlling the error terms. The analysis was initially restricted to a specific parameter regime known as Dobrushin's uniqueness region, and was only recently extended to include the so-called subcritical regime, or high-temperature regime (see for instance Bhamidi et al. [2011], at the very end of Subsec. 1.1 for a precise definition of this regimes). The work Winstein [2025] goes further, covering the supercritical regime, or low-temperature regime, including the critical curve, albeit with certain caveats applying exclusively along that curve. Both works restrict to the case where the coefficients of subgraph densities (other than the edge density) are nonnegative. We also point out the very recent work of Fang et al. [2025b], which studies the asymptotic distribution of the number of two-stars in a model of ERG where the number of edges is conditioned to satisfy some constraint. Chapter 1, is concerned with the asymptotics of the triangle density in edge-triangle models.

Overview of chapter 1: Density of triangles in edge-triangle models

This chapter contains the results of the works Magnanini and Passuello [2025a,b]. We consider the edge-triangle model, where $k = 2$, H_1 corresponds to a single edge, and H_2 corresponds to a triangle, and two related models.

The models are presented in Section 1.1. In Subsection 1.1.1 we introduce the exponential random graph family and we recall results present in the literature, highlighting that in the dense setting, the free energy density of the model can be written in terms of a one-dimensional maximization problem. In Subsection 1.1.2 we specialize to the edge-triangle model. The Hamiltonian in this case can be expressed, for a given simple graph $G \in \mathcal{G}_n$ as

$$\mathcal{H}_{n;\alpha,h}(G) = \alpha \frac{T_n(G)}{n} + hE_n(G),$$

where α and h are real parameters and $T_n(G)$ and $E_n(G)$ represent, respectively, the number of triangles and the number of edges in G. We describe the phase diagram of the model in the replica symmetric regime, where there exists a curve ending in a critical point, where the variational problem describing the free energy admits two distinct solutions. Subsection 1.1.3 is devoted to the mean-field approximation of the model. In Subsection 1.1.4 we fix the notation and show how the model can be seen as system of interacting spins. In Subsection 1.1.5 we define an approximated model, where the integer part of the triangle count is considered.

In Section 1.2 we state the results. The main result for the edge-triangle model is a strong law of large numbers (SLLN), Theorem 1.4, valid for the triangle density. In the mean-field setting, we consider an approximated triangle count, for which we can prove a SLLN, a phase coexistence result at the critical curve, Theorem 1.8, a standard CLT out of criticality, Theorem 1.9, a non-standard CLT at the critical point, Theorem 1.10, and a result on the rate on these convergences. These results hold also on the critical curve if we consider some conditional measures, as Theorem 1.14 states. Finally in Subsection 1.2.3 we consider the integer part model, for which we state a standard CLT out of criticality, Theorem 1.18, valid for the triangle density. Our results allow the parameters of the model to be negative.

Section 1.3 is devoted to the proofs. In Subsection 1.3.1, we provide some preliminary background on the graphon formalism. Subsection 1.3.2 focuses on the proof of Theorem 1.5, which provides a concentration result valid on the critical curve. Here the known LDP for the Erdős–Rényi graph is employed, after having noticed that the probability measure of the edge-triangle model can be represented as a tilted measure on the space of graphons. This is later used to prove Theorem 1.4. Subsections 1.3.3–1.3.7 are devoted to the mean-field model. In particular, in Subsection 1.3.3 we present some additional helpful notation and we state Lemma 1.23, which provides a very useful expression of the partition function of the model in terms of tractable Riemann sums. In Subsection 1.3.8, the proof of Theorem 1.18 is given. This is based on the possibility to express the partition function of the model as a polynomial, which implies uniform convergence for the derivatives of the free energy.

Finally, in Sec. 1.4 we outline why we expect that the fluctuations observed for the approximated triangle density should also hold for the original triangle density in the edge-triangle model, supporting our heuristic argument with simulations.

Spectra of random graphs

Other statistics that are useful to investigate the connectivity structure of a graph concern its spectrum. The spectrum of the adjacency matrices of a graph can give information on the number of its connected components, or reveal the presence of a community structure. The eigenvalues of its associated Laplacian matrix highlight the presence of bottlenecks. Moreover, the normalized Laplacian of undirected graphs provides information on the convergence rate of Markov dynamics.

The study of spectral properties of large matrices has been a central theme in probability theory and mathematical physics for decades, motivated by applications to theoretical physics and data analysis. Given a $n \times n$ matrix A_n, for large n, one of the typical object of interest is its empirical spectral distribution (ESD), defined by

$$\mu_{A_n} = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(A_n)},$$

where $(\lambda_i(A_n))_{i \leq n}$ represents the set of eigenvalues of A. In the classical Hermitian setting, under quite general hypotheses, the ESD of random matrices converges to limits that do not depend on the specific random matrix realization nor on the specific model. The most famous result in this direction is the celebrated Wigner semicircular law, observed for the first time in Wigner [1955]: symmetric random matrices with i.i.d. entries and bounded second moment, when rescaled by $n^{-1/2}$ have ESD converging to the so-called semicircular distribution with density $(2\pi)^{-1}\sqrt{4-x^2}\mathbf{1}_{[-2,2]}(x)\,dx$. The analysis can also regard the edge of the spectrum, which is usually Tracy-Widom distributed Tracy and Widom [1993, 1996]. See the books Mehta [1991], Anderson et al. [2009] and Erdős and Yau [2017] for an introduction to random matrices.

A strong interest relies in studying the spectrum of low-rank perturbation of random matrices Baik et al. [2005], Benaych-Georges and Nadakuditi [2011], Pizzo et al. [2013], Tao [2013]. For instance, in a homogeneous Erdős–Rényi graph with appropriate scaling, almost all eigenvalues of the adjacency matrix A_n lie near the semicircular bulk, except the largest one, that is typically closely tied to the average degree Erdős et al. [2013]. In this setting, the matrix $\mathbb{E}[A_n]$ corresponds to a rank-1 perturbation added to the matrix $A_n - \mathbb{E}[A_n]$, which constitutes a centered i.i.d. noise.

In real-world networks, edge probabilities are inhomogeneous, varying across vertex pairs. Recent works have characterized the spectrum of an inhomogeneous Erdős–Rényi graph. Chakrabarty et al. [2021] identified the limiting ESD of the adjacency and Laplacian matrix, and Chakrabarty et al. [2020] analyzed the eigenvalues outside the bulk (outliers) in such model. These studies confirm that connection probabilities exhibiting low-rank deterministic structures can lead to a handful of eigenvalues straying away from the bulk of the spectrum. In Cipriani et al. [2025], the limiting adjacency ESD for a kernel-based generalization of scale-free percolation is characterized.

Regarding general inhomogeneous random matrices, we mention Brailovskaya and van Handel [2024], where a general universality principle for extreme singular values is proved. Extreme eigenvalues for rectangular arrays were studied in Götze and Tikhomirov [2023] where sparse i.i.d. matrices with a logarithmic number of non-zero entries are studied and in Dumitriu and Zhu [2024], where inhomogeneous matrices are considered. The largest eigenvalues of directed graphs with fixed degrees have been studied in Bordenave [2020] and Coste [2021].

Overview of chapter 2: Spectral properties of the directed Chung–Lu graph

In this chapter, we present the results of an ongoing work by the author and R. S. Hazra. We consider a directed inhomogeneous Erdős–Rényi graph. We study the spectrum of its adjacency matrix A_n, focusing on its largest eigenvalues. In particular, we consider two models, the directed Chung–Lu graph, and a more general model with higher (but finite) rank. In this setting, connection probabilities are given in additive form as the sum of rank one matrices, corresponding to eigenvectors of $\mathbb{E}[A_n]$.

In Section 2.1 the setup is described. In Section 2.1.1 the objects of the analysis are introduced. Subsection 2.1.2 is devoted to the definition of the models. In Subsection 2.1.3 the results are stated. We prove that the eigenvalues of the random matrix A_n correspond to small perturbations of those of $\mathbb{E}[A_n]$. In particular, provided that the average degree uniformly grows as a sequence $s_n = \Omega(\log(n)^4)$, their distance from their deterministic counterparts is at most $O(\sqrt{s_n})$. This implies the existence of outliers at the scale of s_n, stated in Theorems 2.5 and 2.7. In Theorem 2.6 we provide a similar result for the transition matrix of the simple random walk. Theorems 2.8 and 2.9 provide a refinement of the previous theorems: we show that outliers exhibit Gaussian fluctuations around their mean. Subsection 2.1.4 contains a discussion of the methods.

In Section 2.2 we study the outlier problem for the rank-one model. We use Bauer–Fike Theorem (Subsection 2.2.1) to provide a bound on the spectral norm of the matrix $A_n - \mathbb{E}[A_n]$ (Subsection 2.2.2). We then repeat the analysis for the transition matrix (Subsection 2.2.3). after that fluctuations of the outliers, are characterized (Subsection 2.2.4). This is done by identifying, via concentration inequalities, a fixed-point equation satisfied by the outlier with very high probability, and invoking Lindeberg CLT.

Section 2.3 contains the adaptation of Section 2.2 to the higher-rank framework. In particular, Subsection 2.3.1 employs again Bauer–Fike Theorem in a different fashion, to prove the bound on the spectral norm of the centered matrix, and Subsection 2.3.1 introduces a $r \times r$ matrix whose eigenvalues are exactly the outliers of A_n, and then establishes the desired convergence in distribution.

In Section 2.4 we prove some technical lemmata and finally, in Section 2.5 we state a conjecture regarding the limiting empirical spectral distribution of the matrix $\frac{1}{\sqrt{s_n}} A_n$.

Chapter 1

Density of triangles in edge-triangle models

Exponential random graphs (ERGs) are a widely studied class of models that aim to incorporate typical tendencies, such as clustering, commonly observed in real networks. As a generalization of the Erdős-Rényi model Erdős and Rényi [1961], ERGs allow for dependencies between edges. Their probability distribution is obtained by tilting the Erdős-Rényi measure by an exponential weight that contains different subgraph densities. This is done by introducing an Hamiltonian, parametrized by real coefficients, to bias the probability measure over the space of graphs, enhancing or penalizing the density of specific subgraph counts. From a statistical mechanics perspective, ERGs can be interpreted as finite spin systems, where each edge corresponds to a spin in $\{0, 1\}$.

In this chapter, building on Bianchi et al. [2024], we consider the edge-triangle model and study two approximations. We first consider a mean-field approximation, where the triangle density in the Hamiltonian is substituted by a normalized power of the edge density, and then a model, where number of triangles per vertex is substituted by its integer part. Both models share the same free energy of the original model, whose analytical expression and phase diagram are known, in a region of parameters called *replica symmetric regime* Chatterjee and Diaconis [2013].

As a main advantage, the mean-field approximation allows for exact computations. Indeed, the Hamiltonian can be expressed as a function of the edge density, which plays the same role as the magnetization in the Curie-Weiss model. For an approximated triangle density, we prove a standard CLT out of criticality and a non-standard CLT at criticality. We are driven by the belief (supported by simulations) that such approximation asymptotically behaves the same as the original model.

On the other hand, the integer part model allows for a polynomial expression of the free energy. Then, the theory developed in Lee and Yang [1952] provides a uniform convergence of its derivatives, which is the crucial point to prove, for the triangle density, a standard CLT out of criticality.

1.1 Models and background

1.1.1 Exponential random graphs

Given H and G in $\mathcal{G}_n$, let $|\hom(H,G)|$ denote by the number of homomorphism of H in G, namely mappings $\varphi : H \to G$ that are edge-preserving: if u, w are adjacent in H, then $\varphi(u), \varphi(w)$ are adjacent in G. For example, if H is a triangle, then $|\hom(H,G)| = 6T_n(G)$, where $T_n(G)$ denotes the number of triangles in G. If H is a two-star (or wedge) and G is a triangle, then $|\hom(H,G)| = 3 \cdot 2^2$; indeed, there are three copies of H in G (one for each root) and for each of them 4 possible homorphisms. We define the homomorphism density as

$$t(H,G) := \frac{|\hom(H,G)|}{|V(G)|^{|V(H)|}},$$

where the notation $V(\cdot)$ denotes the vertex set of a graph. For a fixed $k \in \mathbb{N}$, we consider $H_1, H_2, \dots, H_k$ pre-chosen finite simple graphs (such as edges, stars, triangles, cycles, ...) weighted by a collection of real parameters contained in the vector $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)$. The Hamiltonian is a function $\mathcal{H}_{n;\boldsymbol{\beta}} : \mathcal{G}_n \to \mathbb{R}$ defined as

$$\mathcal{H}_{n;\boldsymbol{\beta}}(G) := n^2 \sum_{i=1}^{k} \beta_i t(H_i, G)\,, \qquad \text{for } G \in \mathcal{G}_n. \tag{1.1}$$

As probability measure on the space $\mathcal{G}_n$ we take the Gibbs probability density

$$\mu_{n;\boldsymbol{\beta}}(G) := \frac{\exp\left(\mathcal{H}_{n,\boldsymbol{\beta}}(G)\right)}{Z_{n;\boldsymbol{\beta}}}, \quad \text{with } Z_{n;\boldsymbol{\beta}} := \sum_{G \in \mathcal{G}_n} \exp\left(\mathcal{H}_{n;\boldsymbol{\beta}}(G)\right), \tag{1.2}$$

where the normalizing constant $Z_{n;\boldsymbol{\beta}}$ is called partition function. Random graphs whose distribution is a Gibbs measure of the form (1.2) are called exponential random graphs. We will denote the related Gibbs measure and average by $\mathbb{P}_{n;\boldsymbol{\beta}}$ and $\mathbb{E}_{n;\boldsymbol{\beta}}$, respectively. Two crucial functions for studying the model are the finite-size and infinite-size free energy:

$$f_{n;\boldsymbol{\beta}} := \frac{1}{n^2} \ln Z_{n;\boldsymbol{\beta}} \qquad \text{and} \qquad f_{\boldsymbol{\beta}} := \lim_{n \to +\infty} f_{n;\boldsymbol{\beta}}\,.$$

An explicit expression of this function has been obtained in Chatterjee and Diaconis [2013] when the vector of parameters $\boldsymbol{\beta}$ lies in a specific region called replica symmetric regime (term borrowed from spin glasses theory). As stated in [Chatterjee and Diaconis, 2013, Thm. 4.1], if $\beta_2, \dots, \beta_k$ are non-negative, then

$$f_{\boldsymbol{\beta}} = \sup_{0 \le u \le 1} \left(\sum_{i=1}^{k} \beta_i \, u^{|E(H_i)|} - \frac{1}{2} I(u) \right), \tag{1.3}$$

where $|E(H_i)|$ denotes the number of edges in H_i and $I(u) := u \ln u + (1-u)\ln(1-u)$. Despite this result covers only non-negative values of the parameters, the replica symmetric regime can be slightly extended including (not too big) negative values of

$\beta_2, \ldots, \beta_k$ (see Chatterjee and Diaconis [2013], Thm. 4.2). More precisely, (1.3) holds whenever $\beta_2, \ldots, \beta_k$ are such that

$$\sum_{i=2}^{k} |\beta_i||E(H_i)|(|E(H_i)| - 1) < 2\,. \tag{1.4}$$

Notice that the quantity above does not depend on the number of edges $|E(H_1)|$ and if we are not interested in considering the density of edges as a relevant statistics, it suffices to take $\beta_1 = 0$.

1.1.2 Edge-triangle model

The edge-triangle or *Strauss* model Strauss [1986] is obtained by considering only the contribution of edges and triangles in the Hamiltonian (1.1). By convention we assume H_1 to be a single edge and H_2 to be a triangle. More precisely, by setting $\beta_3 = \cdots = \beta_k = 0$ in (1.1), we get

$$\mathcal{H}_{n;\boldsymbol{\beta}}(G) = n^2 \left[\beta_1 t(H_1, G) + \beta_2 t(H_2, G)\right] \qquad G \in \mathcal{G}_n\,.$$

Let $E_n(G)$ (resp. $T_n(G)$) denote the number of edges (resp. triangles) in G. By recalling the definition of homomorphism density, we have

$$t(H_1, G) = \frac{2E_n(G)}{n^2} \quad \text{and} \quad t(H_2, G) = \frac{6T_n(G)}{n^3}. \tag{1.5}$$

Therefore, by performing the change of variable $h := 2\beta_1$; $\alpha := 6\beta_2$, we can equivalently consider

$$\mathcal{H}_{n;\alpha,h}(G) = \frac{\alpha}{n} T_n(G) + hE_n(G) \quad \text{with } \alpha, h \in \mathbb{R}\,. \tag{1.6}$$

We will denote by $\mathbb{P}_{n;\alpha,h}$ the Gibbs measure related to this Hamiltonian, and by $\mathbb{E}_{n;\alpha,h}$ the corresponding expectation. Notice that in this setting condition (1.4) reads $|\alpha| = 6|\beta_2| < 2$, and, therefore, the replica symmetric regime coincides with the region $\alpha > -2$, $h \in \mathbb{R}$. The free energy (1.3) reduces then to

$$f_{\alpha,h} = \sup_{0 \le u \le 1} \left(\frac{\alpha}{6} u^3 + \frac{h}{2} u - \frac{1}{2} I(u) \right) = \frac{\alpha}{6}(u^*)^3 + \frac{h}{2} u^* - \frac{1}{2} I(u^*), \tag{1.7}$$

where $I(u)$ is defined below (1.3) and $u^* = u^*(\alpha, h)$ is a maximizer that solves the fixed-point equation

$$\frac{e^{\alpha u^2 + h}}{1 + e^{\alpha u^2 + h}} = u\,. \tag{1.8}$$

A numerical investigation of the optimizers of the free energy when α is negative and $|\alpha|$ is large has been done in Giardinà et al. [2021]. Equation (1.8) can admit more than one solution at which the supremum in (1.7) is attained, and this denotes the presence of a phase transition inside the replica symmetric regime. When the parameters α, h are chosen in this region, the edge-triangle model, when n goes to infinity, becomes

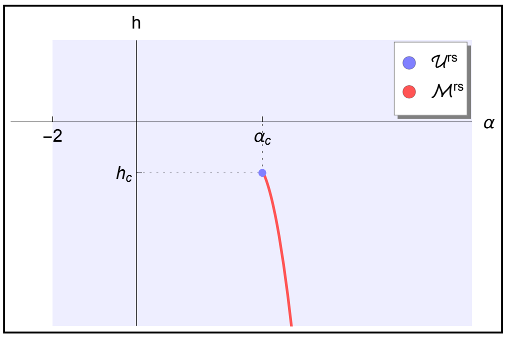

Figure 1.1: Illustration of the phase in replica symmetric regime taken from Bianchi et al. [2024]. The curve $\mathcal{M}^{rs}$ (1.9) represents the region of (α, h) where the optimization problem (1.7) admits two solutions. Inside the blue region, that includes the critical point (α_c, h_c), the scalar problem (1.7) admits a unique solution.

indistinguishable from an Erdős-Rényi graph with connection probability u^* (we refer the reader to Sec. 1.3.1, where these notions are made precise). This remains true even when the supremum is not unique; in this case the parameter u^* is randomly chosen according to some (unknown) probability distribution on the set of solutions of (1.7) (see Chatterjee and Diaconis [2013], Thm. 4.2). The effect of the phase transition is then a jump between very different values of the edge density u^* of the limiting object. We refer to this as a phase transition as it entails a lack of analyticity of the free energy, even though the qualitative structure of the limiting object remains the same.

Phase diagram. We recall that the limiting free energy $f_{\alpha,h}$ is well defined on the whole replica symmetric regime $\alpha > -2$, $h \in \mathbb{R}$. However, the fixed point equation (1.8) can admit more than one solution, and this is strictly related to the loss of analyticity of $f_{\alpha,h}$. It has been proved (see [Radin and Yin, 2013, Prop. 3.2]) that (1.7) has exactly one optimizer on the whole replica symmetric regime except for a certain critical curve $\mathcal{M}^{rs}$ that starts at the critical point $(\alpha_c, h_c) := \left(\frac{27}{8}, \ln 2 - \frac{3}{2}\right)$ and that can be written as $h = q(\alpha)$ for a (non-explicit) continuous and strictly decreasing function q:

$$\mathcal{M}^{rs} := \{(\alpha, h) \in (\alpha_c, +\infty) \times (-\infty, h_c) : h = q(\alpha)\}. \tag{1.9}$$

It is worth noting that along this critical curve, the scalar problem (1.7) admits multiple maximizers (precisely two), as the notation $\mathcal{M}^{rs}$ is meant to suggest. The free energy is analytic on the region $\mathcal{U}^{rs} \setminus \{(\alpha_c, h_c)\}$, where

$$\mathcal{U}^{rs} := ((-2, +\infty) \times \mathbb{R}) \setminus \mathcal{M}^{rs}.$$

Here, in contrast, we use the notation $\mathcal{U}^{rs}$ to denote the region where the scalar problem (1.7) has a unique maximizer. Moreover, at the critical point (α_c, h_c) the second order partial derivatives of $f_{\alpha,h}$ diverge (see Radin and Yin [2013], Thm. 2.1), while along the curve $\mathcal{M}^{rs}$ the first order partial derivatives of $f_{\alpha,h}$ have jump discontinuities. Fig. 1.1 provides a qualitative representation of the phase diagram.

1.1.3 A mean-field model

As a consequence of the convergence of the ERG to the Erdős-Rényi graphon with parameter u^* (which holds in probability w.r.t. the so-called *cut distance*, see [Chatterjee and Diaconis, 2013, Thm. 4.2]), we can heuristically approximate the triangle density (as well as other graph-statistics) in the large n limit. The Erdős-Rényi random graph with parameter u^* and n vertices has, on average, $u^{*3}\binom{n}{3}$ triangles and $u^*\binom{n}{2}$ edges. We observe that

$$u^{*3}\binom{n}{3} \approx \frac{4}{3n^3}\left(u^*\binom{n}{2}\right)^3.$$

What we expect is that the same holds, within the replica symmetric regime and when n is large, for the ERG. Thus, we introduce the approximated count of triangles

$$T_n(G) \approx \frac{4}{3n^3} E_n(G)^3 =: \bar{T}_n(G). \tag{1.10}$$

We can equivalently say that we approximate the number of triangles $T_n(G) = \frac{n^3 t(H_2,G)}{6}$ (see (1.5)) with $\bar{T}_n(G) = \frac{n^3 t^3(H_1,G)}{6}$. Definition (1.10) leads to the following mean-field approximation, originally introduced in Bianchi et al. [2024], of the edge-triangle Hamiltonian (1.6):

$$\bar{\mathcal{H}}_{n;\alpha,h}(G) := \frac{\alpha}{n}\bar{T}_n(G) + hE_n(G)\,, \qquad \text{for } G \in \mathcal{G}_n. \tag{1.11}$$

We borrow this terminology from statistical mechanics, due to the similarities with the Curie–Weiss model (see e.g. [Friedli and Velenik, 2017, Chap. 2]), which we are going to highlight further in the next paragraph. The big advantage of Hamiltonian (1.11) is that it is just a function of the one dimensional parameter $t(H_1, G) = \frac{2E_n(G)}{n^2}$, taking values in $\Gamma_n := \left\{0, \frac{2}{n^2}, \ldots, 1 - \frac{1}{n}\right\}$, as it is for the Erdős–Rényi graph. We denote by $\bar{\mathbb{P}}_{n;\alpha,h}$ and $\bar{\mathbb{E}}_{n;\alpha,h}$ the corresponding measure and expectation, respectively. Moreover, as usual, we define the finite size free energy as

$$\bar{f}_{n;\alpha,h} := \frac{1}{n^2} \ln \bar{Z}_{n;\alpha,h}.$$

A crucial property of this approximated model is the following (see [Bianchi et al., 2024, Thm. 8.2]). Let $(\alpha, h) \in (-2, +\infty) \times \mathbb{R}$ and let $f_{\alpha,h}$ as in (1.7). Then

$$\lim_{n\to+\infty} \bar{f}_{n;\alpha,h} = f_{\alpha,h}\,. \tag{1.12}$$

In other words, the edge-triangle model and this mean-field approximation share the same infinite volume free energy; this result will be extensively used in the proofs.

1.1.4 Notation and preliminaries

We denote by $\mathcal{E}_n$ the edge set of the complete graph on n vertices, with elements labeled from 1 to $\binom{n}{2}$ and we set $\mathcal{A}_n := \{0,1\}^{\mathcal{E}_n}$. We observe that there is a one-to-one correspondence between $\mathcal{A}_n$ and the set of $n \times n$ symmetric adjacency matrices with zeros on the diagonal and the graphs in $\mathcal{G}_n$. As a consequence, to each graph $G \in \mathcal{G}_n$ we can associate an element $x = (x_i)_{i \in \mathcal{E}_n} \in \mathcal{A}_n$ where $x_i = 1$ if the edge i is present in G, and $x_i = 0$ otherwise. With an abuse of nomenclature, we can write $E_n(x) = E_n(G)$, $T_n(x) = T_n(G)$ and $\bar{T}_n(x) = \bar{T}_n(G)$ whenever $x \in \mathcal{A}_n$ is the adjacency matrix of a graph $G \in \mathcal{G}_n$. This representation allows for the following equivalent formulation of the Hamiltonians (1.6)–(1.11), as functions on $\mathcal{A}_n$:

$$\mathcal{H}_{n;\alpha,h}(x) = \frac{\alpha}{n} \sum_{\{i,j,k\} \in \mathcal{T}_n} x_i x_j x_k + h \sum_{i \in \mathcal{E}_n} x_i, \tag{1.13}$$

$$\bar{\mathcal{H}}_{n;\alpha,h}(x) = \frac{4\alpha}{3n^4} \Big(\sum_{i \in \mathcal{E}_n} x_i \Big)^3 + h \sum_{i \in \mathcal{E}_n} x_i, \tag{1.14}$$

where $\mathcal{T}_n := \{\{i,j,k\} \subset \mathcal{E}_n : \{i,j,k\} \text{ is a triangle}\}$. The Gibbs probability $\mathbb{P}_{n;\alpha,h}$ (resp. $\bar{\mathbb{P}}_{n;\alpha,h}$) will act consequently on $\mathcal{A}_n$.

Remark 1.1. *The sequence of measures $(\mathbb{P}_{n;\alpha,h})_{n\geq 1}$ (as well as $(\bar{\mathbb{P}}_{n;\alpha,h})_{n\geq 1}$) satisfies proper consistency conditions allowing for the application of Kolmogorov Existence Theorem (see, for example, Appendix A.7 in Ellis [1985]). As a consequence, there exists a unique probability measure $\mathbb{P}_{\alpha,h}$ on the space $\big(\{0,1\}^{\mathbb{N}}, \mathcal{B}(\{0,1\}^{\mathbb{N}})\big)$ with marginals corresponding to the measures $\mathbb{P}_{n;\alpha,h}$, for all $n \in \mathbb{N}$ (here $\mathcal{B}$ denotes the Borel $\sigma-$algebra).*

Remark 1.2. *Note that Hamiltonian (1.13) has the same form of the energy function typically used in the field of interacting particle systems. We can think of an ERG as a system where each edge is a particle having a spin (0 or 1), which interacts with its neighbors. For instance, as observed in Mukherjee and Xu [2023], the two-star model can be thought as an of Ising model on a d-regular graph with $n(n-1)/2$ nodes, where $d = 2(n-2)$. The notion of "neighbor" depends on the specific choice of the subgraphs $H_1, \ldots, H_k$; for the edge-triangle model, two edges are neighbors if they are adjacent. This interaction is not global, however, if we ignore the relative position of edges, we recover (1.14):*

$$\sum_{\{i,j,k\} \in \mathcal{T}_n} x_i x_j x_k = \sum_{i \in \mathcal{E}_n} x_i \sum_{\substack{j,k \in \mathcal{E}_n: \\ \{i,j,k\} \in \mathcal{T}_n}} x_j x_k \approx \sum_{i \in \mathcal{E}_n} x_i \sum_{j,k \in \mathcal{E}_n} \frac{4 x_j x_k}{3n^3},$$

where the factor 4 appears when we replace the number of wedges roughly with $\big(\frac{2}{n^2} \sum_{i \in \mathcal{E}_n} x_i\big)^2$, and adjusting the normalization in accordance with the choice $x = (x_i)_{i \in \mathcal{E}_n} \equiv 1$. The factor $1/3$ avoids overcounting.

We are interested in understanding the asymptotic behavior of the number of triangles inside the replica symmetric regime. We prove classical limit theorems for the sequences $(T_n)_{n\geq 1} \equiv (T_n(X))_{n\geq 1}$ and $(\bar{T}_n)_{n\geq 1} \equiv (\bar{T}_n(X))_{n\geq 1}$, where $X = (X_i)_{i \in \mathcal{E}_n} \in \mathcal{A}_n$ is a random adjacency matrix, whose law will be specified case by case. The core of our results is concerned with the sequence $(\bar{T}_n)_{n\geq 1}$ under the mean-field distribution, since the approximation encoded by the Hamiltonian (1.14) allows for explicit computations. This quantity is in principle not related to the number of triangles, but we believe that it captures the correct limiting behavior of triangle density.

Definition 1.3. *For each $n \in \mathbb{N}$, we define*

$$\bar{m}_n^{\Delta}(\alpha, h) := \frac{6\bar{\mathbb{E}}_{n;\alpha,h}\left(\frac{\bar{T}_n}{n}\right)}{n^2} \quad \text{and} \quad \bar{v}_n^{\Delta}(\alpha, h) := \partial_\alpha \bar{m}_n^{\Delta}(\alpha, h).$$

It is easy to see that $\frac{\bar{\mathbb{E}}_{n;\alpha,h}(\bar{T}_n)}{n^3} = \partial_\alpha \bar{f}_{n;\alpha,h}$ and $\frac{\bar{\mathbb{V}}\mathrm{ar}_{n;\alpha,h}(\bar{T}_n)}{n^3} = \partial_{\alpha\alpha} \bar{f}_{n;\alpha,h}$. Therefore, $\bar{m}_n^{\Delta}(\alpha, h) = 6\partial_\alpha \bar{f}_{n;\alpha,h}$ and $\bar{v}_n^{\Delta}(\alpha, h) = 6\partial_{\alpha\alpha} \bar{f}_{n;\alpha,h}$ (and the same holds for the edge-triangle model, replacing $\bar{f}_{n;\alpha,h}$ with $f_{n;\alpha,h}$). In the rest of the paper we will use the following notation to distinguish the optimizer(s) of the scalar problem (1.7), sometimes dropping the dependence on (α, h) to the sake of readability:

$$\begin{cases} u_0^*(\alpha, h) & \text{if } (\alpha, h) \in \mathcal{U}^{rs} \setminus \{(\alpha_c, h_c)\}, \\ u_1^*(\alpha, h) \text{ and } u_2^*(\alpha, h) & \text{if } (\alpha, h) \in \mathcal{M}^{rs}, \\ u_c^*(\alpha, h) = \frac{2}{3} & \text{if } (\alpha, h) = (\alpha_c, h_c) \,. \end{cases}$$

1.1.5 An integer part model

The Hamiltonian (1.6), can be modified as follows, taking into account only the integer part of the normalized number of triangles:

$$\hat{\mathcal{H}}_{n;\alpha,h}(G) := \alpha \left\lfloor \frac{T_n(G)}{n} \right\rfloor + hE_n(G). \tag{1.15}$$

We denote by $\hat{\mu}_{n;\alpha,h}$ the associated Gibbs probability density and by $\hat{\mathbb{P}}_{n;\alpha,h}$ the related measure, with normalizing partition function $\hat{Z}_{n;\alpha,h}$, and expectation $\hat{\mathbb{E}}_{n;\alpha,h}$. Finally, we indicate by

$$\hat{f}_{n;\alpha,h} := \frac{1}{n^2} \ln \hat{Z}_{n;\alpha,h} \quad \text{and} \quad \hat{f}_{\alpha,h} := \lim_{n \to +\infty} \hat{f}_{n;\alpha,h} \tag{1.16}$$

the finite-size and the limiting free energy, respectively. Importantly, $\hat{f}_{\alpha,h} = f_{\alpha,h}$. This immediately follows from the decomposition

$$\frac{T_n(G)}{n} = \left\lfloor \frac{T_n(G)}{n} \right\rfloor + \left\{ \frac{T_n(G)}{n} \right\},$$

where $\{\cdot\} \in [0, 1]$ denotes the fractional part. This model amounts to a minimal perturbation of the edge-triangle model, which will turn to be analytically tractable. Its partition function can be expressed in a polynomial in the variable $z = e^\alpha$ and this will be used to prove standard CLT for the density of triangles.

1.2 Main results

1.2.1 Edge-triangle model

Theorem 1.4 (SLLN for T_n). *For all* $(\alpha, h) \in \mathcal{U}^{rs}$,

$$\frac{6T_n}{n^3} \xrightarrow[n\to+\infty]{\text{a.s.}} u^{*3}(\alpha, h) \qquad w.r.t.\ \mathbb{P}_{\alpha,h},$$

where $u^* = \begin{cases} u_c^*, & \text{if } (\alpha, h) = (\alpha_c, h_c) \\ u_0^*, & \text{otherwise} \end{cases}$ *solves the maximization problem* (1.7).

Theorem 1.5. *For all* $(\alpha, h) \in \mathcal{M}^{rs}$ *and for all sufficiently small* $\varepsilon > 0$, *there exists a constant* $\tau = \tau(\varepsilon; \alpha, h) > 0$ *such that if*

$$J(\varepsilon) := (u_1^{*3}(\alpha, h) - \varepsilon, u_1^{*3}(\alpha, h) + \varepsilon) \cup (u_2^{*3}(\alpha, h) - \varepsilon, u_2^{*3}(\alpha, h) + \varepsilon),$$

then, for large enough n

$$\mathbb{P}_{n;\alpha,h}\left(\frac{6T_n}{n^3} \in J(\varepsilon)\right) \geq 1 - e^{-\tau n^2},$$

where $u_1^*(\alpha, h)$ *and* $u_2^*(\alpha, h)$ *are the two maximizers of the scalar problem* (1.7).

Remark 1.6. *Our proof of Thms. 1.4 and 1.5 follows the approach of Bianchi et al. [2024], and strongly relies on the LDP proved in Chatterjee and Varadhan [2011], and on [Ellis, 1985, Thm. II.7.2]. Consequently our techiniques differ from those used in [Radin and Sadun, 2023, Thm. 6], where an analogous result to Thm. 1.5, for a model with subgraph densities subject to hard constraints, has been proven.*

After having proved a SLLN, it would be natural to investigate the fluctuations of the triangle density around its mean value. In Sec. 1.4 we perform simulations and sketch some heuristics, also based on the mean-field investigation of Sec. 1.3, which suggest that this result holds. In order to establish it, a powerful tool is the Yang-Lee theorem [Lee and Yang, 1952, Thm. 2], implies uniform convergence of the derivatives of the free energy. This is applicable to our case if the partition function of the model admits a polynomial representation in $z := e^{\alpha}$. The results for the integer part model, stated in Subsection 1.2.3, follow this approach. Another tool which could be useful for proving the central limit theorem, is the Griffiths, Hurst and Sherman inequality Griffiths et al. [1970], which allows for the interchange of limits and derivatives in spin systems; for the broader class of exponential random graphs this result has not been proven (although it was recently established for the two-star model Bianchi et al. [2022])

1.2.2 Mean-field approximation

Theorem 1.7 (SLLN for $\bar{T}_n$). *For all* $(\alpha, h) \in \mathcal{U}^{rs}$,

$$\frac{6\bar{T}_n}{n^3} \xrightarrow[n\to+\infty]{\text{a.s.}} u^{*3}(\alpha, h) \qquad w.r.t.\ \bar{\mathbb{P}}_{\alpha,h},$$

where $u^* = \begin{cases} u_c^*, & \text{if } (\alpha,h) = (\alpha_c,h_c) \\ u_0^*, & \text{otherwise} \end{cases}$ *solves the maximization problem* (1.7)*.*

Theorem 1.8. *For all* $(\alpha,h) \in \mathcal{M}^{rs}$*,*

$$\frac{6\bar{T}_n}{n^3} \xrightarrow[n\to+\infty]{\mathrm{d}} \kappa\delta_{u_1^{*3}(\alpha,h)} + (1-\kappa)\delta_{u_2^{*3}(\alpha,h)} \qquad w.r.t.\ \bar{\mathbb{P}}_{n;\alpha,h},$$

where u_1^**,* u_2^* *solve the maximization problem in* (1.7)*, and*

$$\kappa := \frac{\sqrt{\left[1-2\alpha\left(u_1^*\right)^2\left(1-u_1^*\right)\right]^{-1}}}{\sqrt{\left[1-2\alpha\left(u_1^*\right)^2\left(1-u_1^*\right)\right]^{-1}}+\sqrt{\left[1-2\alpha\left(u_2^*\right)^2\left(1-u_2^*\right)\right]^{-1}}}.$$

Theorem 1.9 (CLT for $\bar{T}_n$)**.** *If* $(\alpha,h) \in \mathcal{U}^{rs} \setminus \{(\alpha_c,h_c)\}$*,*

$$\sqrt{6}\,\frac{\frac{\bar{T}_n}{n} - \frac{n^2}{6}\bar{m}_n^{\Delta}(\alpha,h)}{n} \xrightarrow[n\to+\infty]{\mathrm{d}} \mathcal{N}(0,\bar{v}_0^{\Delta}(\alpha,h)) \qquad w.r.t.\ \bar{\mathbb{P}}_{n;\alpha,h},$$

where $\mathcal{N}(0,\bar{v}_0^{\Delta}(\alpha,h))$ *is a centered Gaussian distribution with variance*

$$\bar{v}_0^{\Delta}(\alpha,h) := \frac{3u_0^{*4}(\alpha,h)}{4c_0}, \tag{1.17}$$

being $c_0 \equiv c_0(\alpha,h) := \frac{1-2\alpha[u_0^*(\alpha,h)]^2[1-u_0^*(\alpha,h)]}{4u_0^*(\alpha,h)[1-u_0^*(\alpha,h)]}$*.*

Theorem 1.10 (Non-standard CLT for $\bar{T}_n$)**.** *If* $(\alpha,h) = (\alpha_c,h_c)$*,*

$$6\,\frac{\frac{\bar{T}_n}{n} - \frac{n^2}{6}\bar{m}_n^{\Delta}(\alpha_c,h_c)}{n^{3/2}} \xrightarrow[n\to+\infty]{\mathrm{d}} \bar{Y} \qquad w.r.t.\ \bar{\mathbb{P}}_{n;\alpha_c,h_c},$$

where $\bar{Y}$ *is a generalized Gaussian random variable with Lebesgue density* $\bar{\ell}^c(y) \propto e^{-\frac{3^8}{2^{14}}y^4}$*.*

Proposition 1.11. *For all* $(\alpha,h) \in \mathcal{U}^{rs} \setminus \{(\alpha_c,h_c)\}$*,*

$$\lim_{n\to+\infty} n \cdot \bar{\mathbb{E}}_{n;\alpha,h}\left(\left|\frac{6\bar{T}_n}{n^3} - u_0^{*3}(\alpha,h)\right|\right) = \mathbb{E}(|\bar{X}|)\,, \tag{1.18}$$

where $\bar{X}$ *is a centered Gaussian random variable with variance* $6\bar{v}_0^{\Delta}(\alpha,h) = \frac{9u_0^{*4}(\alpha,h)}{2c_0}$*, being* $c_0 \equiv c_0(\alpha,h) = \frac{1-2\alpha[u_0^*(\alpha,h)]^2[1-u_0^*(\alpha,h)]}{4u_0^*(\alpha,h)[1-u_0^*(\alpha,h)]} > 0$*. Moreover, at the critical point*

$$\lim_{n\to+\infty} \sqrt{n} \cdot \bar{\mathbb{E}}_{n;\alpha_c,h_c}\left(\left|\frac{6\bar{T}_n}{n^3} - u^{*3}(\alpha_c,h_c)\right|\right) = \mathbb{E}(|\bar{Y}|)\,,$$

where $\bar{Y}$ *is a generalized Gaussian random variable with Lebesgue density* $\bar{\ell}^c(y) \propto e^{-\frac{3^8}{2^{14}}y^4}$*.*

Corollary 1.12. *For all* $(\alpha, h) \in \mathcal{U}^{rs} \setminus \{(\alpha_c, h_c)\}$, *we have*

$$\lim_{n\to+\infty} n \cdot (\bar{m}_n^{\Delta}(\alpha, h) - u_0^{*3}(\alpha, h)) = 0\,, \tag{1.19}$$

while for $(\alpha, h) = (\alpha_c, h_c)$

$$\lim_{n\to+\infty} \sqrt{n} \cdot (\bar{m}_n^{\Delta}(\alpha_c, h_c) - u^{*3}(\alpha_c, h_c)) = 0\,. \tag{1.20}$$

Remark 1.13 (Generalization to a clique graph)**.** *As we approximated the number of triangles* T_n *with* $\bar{T}_n$, *we can similarly approximate the number of cliques* K_n *with* $\ell \geq 3$ *vertices by*

$$\bar{K}_n := \left(\frac{2E_n}{n^2}\right)^{\binom{\ell}{2}} \cdot \frac{n^\ell}{\ell!} = \left(\frac{2\sum_{i\in\mathcal{E}_n} X_i}{n^2}\right)^{\binom{\ell}{2}} \cdot \frac{n^\ell}{\ell!}.$$

Using the techniques from Thms. 1.7, 1.8, 1.9, and 1.10, similar results can be derived for the approximated clique density. Specifically, within the uniqueness region, the almost sure limit of $\ell!\bar{K}_n/n^\ell$ *will be* $(u_0^*(\alpha, h))^{\binom{\ell}{2}}$, *where* $u_0^*(\alpha, h)$ *solves the maximization problem in* (1.7)*. In this region, one can further prove both Gaussian and non-Gaussian fluctuations, respectively for the terms*

$$\sqrt{\ell!}\,\frac{\frac{\bar{K}_n}{n^{\ell-2}} - \bar{\mathbb{E}}_{n;\alpha,h}\left(\frac{\bar{K}_n}{n^{\ell-2}}\right)}{n} \qquad \textit{and} \qquad \ell!\,\frac{\frac{\bar{K}_n}{n^{\ell-2}} - \bar{\mathbb{E}}_{n;\alpha_c,h_c}\left(\frac{\bar{K}_n}{n^{\ell-2}}\right)}{n^{3/2}}.$$

Conditional measures. When (α, h) lies in the multiplicity curve $\mathcal{M}^{rs}$, where the solution of (1.8) is not unique, we can still characterize the limiting behavior of the triangle density in the mean-field approximation, provided that we constraint the edge density to be close to one of the maximizers of the scalar problem (1.7). To this aim, we consider a conditioned model, as follows. For $(\alpha, h) \in \mathcal{M}^{rs}$, let $u_i^*(\alpha, h)$ $(i = 1, 2)$ be the solutions of the scalar problem (1.7). For $n \in \mathbb{N}$ and any fixed $\delta \in (0, 1)$, consider the event

$$B_{u_i^*} \equiv B_{u_i^*}(n, \delta) := \left\{ x \in \mathcal{A}_n \,:\, \left| \frac{2E_n(x)}{n^2} - u_i^*(\alpha, h) \right| \leq n^{-\delta} \right\}, \tag{1.21}$$

and define the conditional probability measures

$$\hat{\mathbb{P}}^{(i)}_{n;\alpha,h}(\,\cdot\,) \equiv \hat{\mathbb{P}}^{(i),\delta}_{n;\alpha,h}(\,\cdot\,) := \bar{\mathbb{P}}_{n;\alpha,h}\left(\,\cdot\,\middle| B_{u_i^*}(n, \delta)\right), \qquad \text{for } i = 1,\, 2\,. \tag{1.22}$$

We denote the corresponding averages by $\hat{\mathbb{E}}^{(i)}_{n;\alpha,h}$ and we set $\hat{m}_n^{(i)}(\alpha, h) := \hat{\mathbb{E}}^{(i)}_{n;\alpha,h}\left(\frac{6\bar{T}_n}{n^3}\right)$. The next statements represent the analog of the results presented in Subsec. 1.2.2, when the parameters belong to the multiplicity region $\mathcal{M}^{rs}$. The next statements are valid for every $\delta \in (0, 1)$ and for this reason the dependence is omitted.

Theorem 1.14 (Conditional SLLN and CLT)**.** *For* $i = 1, 2$ *and for all* $(\alpha, h) \in \mathcal{M}^{rs}$,

$$\frac{6\bar{T}_n}{n^3} \xrightarrow[n\to+\infty]{\text{a.s.}} u_i^*(\alpha, h) \qquad \textit{w.r.t. } \hat{\mathbb{P}}^{(i)}_{n;\alpha,h}, \tag{1.23}$$

and

$$\sqrt{6}\,\frac{\frac{\bar{T}_n}{n}-\frac{n^2}{6}\hat{m}_n^{(i)}(\alpha,h)}{n}\xrightarrow[n\to+\infty]{\mathrm{d}}\mathcal{N}(0,\bar{v}_i^{\Delta}(\alpha,h))\qquad w.r.t.\ \hat{\mathbb{P}}^{(i)}_{n;\alpha,h},\tag{1.24}$$

where $\mathcal{N}(0,\bar{v}_i^{\Delta}(\alpha,h))$ *is a centered Gaussian distribution with variance* $\bar{v}_i^{\Delta}(\alpha,h) := \frac{3u_i^{4*}(\alpha,h)}{4c_i}$, *being* $c_i \equiv c_i(\alpha,h) := \frac{1-2\alpha[u_i^*(\alpha,h)]^2[1-u_i^*(\alpha,h)]}{4u_i^*(\alpha,h)[1-u_i^*(\alpha,h)]}$.

Proposition 1.15. *For* $i=1,2$ *and for all* $(\alpha,h)\in\mathcal{M}^{rs}$,

$$\lim_{n\to+\infty} n\cdot\hat{\mathbb{E}}^{(i)}_{n;\alpha,h}\left(\left|\frac{6\bar{T}_n}{n^3}-u_i^{*3}(\alpha,h)\right|\right)=\mathbb{E}\left(\left|\bar{X}^{(i)}\right|\right),$$

$\bar{X}^{(i)}$ *is a centered Gaussian random variable with variance* $6\bar{v}_i^{\Delta}(\alpha,h)=\frac{9u_i^*(\alpha,h)^4}{2c_i}$, *being* $c_i\equiv c_i(\alpha,h)=\frac{1-2\alpha[u_i^*(\alpha,h)]^2[1-u_i^*(\alpha,h)}{4u_i^*(\alpha,h)[1-u_i^*(\alpha,h)]}$.

Corollary 1.16. *For* $i=1,2$ *and for all* $(\alpha,h)\in\mathcal{M}^{rs}$, *we have*

$$\lim_{n\to+\infty} n\cdot\left(\hat{m}_n^{(i)}(\alpha,h)-u_i^{*3}(\alpha,h)\right)=0,$$

where we recall that $\hat{m}_n^{(i)}(\alpha,h):=\hat{\mathbb{E}}^{(i)}_{n;\alpha,h}\left(\frac{6\bar{T}_n}{n^3}\right)$.

Remark 1.17 (Beyond the edge-triangle model)**.** *Theorems 1.4 and 1.5 can be easily extended to the Gibbs measure* $\mathbb{P}_{\beta_k,\beta_1}$, *whose associated Hamiltonian is obtained by setting* $\beta_i=0$ *for all* $i\neq 1,k$ *(with* $k>2$*) in (1.1), i.e.*

$$\mathcal{H}_{n;\beta_k,\beta_1}(G):=n^2\left[\beta_k t(H_k,G)+\beta_1 t(H_1,G)\right],\qquad for\ G\in\mathcal{G}_n,$$

where H_k, *which takes the place of* H_2, *is a simple graph with* $p\geq 3$ *edges. The limiting free energy is given by the scalar maximization problem (1.3) and the replica symmetric region is defined by the condition* $\beta_k\geq 0$ *combined with (1.4). The phase diagram has been fully characterized in [Radin and Yin, 2013, Prop. 3.2] and is completely analogous to those illustrated in Fig. 1.1 for the edge-triangle model. More precisely, in the replica symmetric regime, there is still a region* $\mathcal{U}^{rs}$ *where the solution to the maximization problem (1.3) is unique, and a critical curve* $\mathcal{M}^{rs}$, *ending in the critical point* $(\beta_k^c,\beta_1^c)=\frac{1}{2}\left(\frac{p^{p-1}}{(p-1)^p},\log(p-1)-\frac{p}{(p-1)}\right)$, *where the maximization problem (1.3) has two distinct solutions.*

1.2.3 Integer part model

Theorem 1.18 (CLT for T_n w.r.t. $\hat{\mathbb{P}}_{n;\alpha,h}$)**.** *For all* $(\alpha,h)\in\mathcal{U}^{rs}\setminus\{(\alpha_c,h_c)\}$

$$\sqrt{6}\,\frac{T_n/n-\hat{\mathbb{E}}_{n;\alpha,h}(T_n/n)}{n}\xrightarrow[n\to+\infty]{\mathrm{d}}\mathcal{N}(0,\hat{v}_0^{\Delta}(\alpha,h))\qquad w.r.t.\ \hat{\mathbb{P}}_{n;\alpha,h},$$

where $\hat{v}_0^{\Delta}(\alpha,h):=3u_{\alpha,h}^{*2}\partial_\alpha u_{\alpha,h}^*$ *and* $\mathcal{N}(0,\hat{v}_0^{\Delta}(\alpha,h))$ *is a centered Gaussian distribution.*

Extension to three parameters. The above theorem immediately extends to a 3-parameter setting. Assume that H_1 is a single edge, H_2 has p edges, and H_3 has q edges, with $2 \leq p \leq q \leq 5p-1$. Let $f_{\beta_1,\beta_2,\beta_3}$ the limiting free energy arising from the Hamiltonian (1.1) by setting $\beta_k = 0$ for all $k \geq 4$. Such function, inside the domain $\mathcal{D}^{rs} := \{(\beta_1, \beta_2, \beta_3) : \beta_2 \geq 0, \beta_3 \geq 0, \beta_1 \in \mathbb{R}\}$ (again, by [Chatterjee and Diaconis, 2013, Thm. 4.1]) exists and equals [1]

$$f_{\beta_1,\beta_2,\beta_3} = \sup_{0\leq u\leq 1} \left(\beta_3 u^q + \beta_2 u^p + \beta_1 u - \frac{1}{2}I(u)\right) = \beta_3 u^{*q} + \beta_2 u^{*p} + \beta_1 u^* - \frac{1}{2}I(u^*),$$

where u^* solves

$$u = \frac{e^{2\beta_3 q u^{q-1} + 2\beta_2 p u^{p-1} + 2\beta_1}}{1 + e^{2\beta_3 q u^{q-1} + 2\beta_2 p u^{p-1} + 2\beta_1}}. \tag{1.25}$$

In this setting, the phase diagram, studied in [Yin, 2013, Thm. 1], is also known. The free energy $f_{\beta_1,\beta_2,\beta_3}$ is analytic in $\mathcal{D}^{rs}$ except for a certain continuous surface S which includes three bounding curves C_1, C_2, and C_3, and that can be characterized as follows:

- the surface S approaches the plane $\beta_1 + \beta_2 + \beta_3 = 0$ as $\beta_1 \to -\infty$, $\beta_2, \beta_3 \to \infty$;
- the curve C_1 is the intersection of S with the plane $\{(\beta_1, \beta_2, \beta_3) : \beta_3 = 0\}$;
- the curve C_2 is the intersection of S with the plane $\{(\beta_1, \beta_2, \beta_3) : \beta_2 = 0\}$;
- the curve C_3 is a critical curve, and is given parametrically by

$$\begin{aligned}
\beta_1(u) &= \frac{1}{2}\ln\frac{u}{1-u} - \frac{1}{2(p-1)(1-u)} + \frac{pu-(p-1)}{2(p-1)(q-1)(1-u)^2},\\
\beta_2(u) &= \frac{qu-(q-1)}{2p(p-1)(p-q)u^{p-1}(1-u)^2},\\
\beta_3(u) &= \frac{pu-(p-1)}{2q(q-1)(q-p)u^{q-1}(1-u)^2},
\end{aligned}$$

  where we take $\frac{p-1}{p} \leq u \leq \frac{q-1}{q}$ to meet the non-negativity constraints on β_2, β_3.

Consider now the 3-parameter Hamiltonian obtained from (1.1) by taking H_1 a single edge, H_2 a triangle, and H_3 a simple subgraph with $q \in [3, 14]$ edges, and setting $\beta_k = 0$ for all $k \geq 4$. Similarly to what we did for the edge-triangle case, we denote by

$$\hat{\mathcal{H}}_{n;\beta_1,\beta_2,\beta_3}(x) = \beta_3 n^2 t(H_3, x) + \beta_2 \lfloor n^2 t(H_2, x)\rfloor + \beta_1 n^2 t(H_1, x), \tag{1.26}$$

and by $\hat{\mathbb{P}}_{n;\beta_1,\beta_2,\beta_3}$ the associated Gibbs measure. Then, the following generalization of Thm. 1.18 holds.

[1] We stress that, from [Chatterjee and Diaconis, 2013, Thm. 6.1], $\mathcal{D}^{rs}$ is actually a subset of the region where the free energy is known. However, the phase diagram is characterized only to that restriction.

Theorem 1.19 (CLT for T_n w.r.t. $\hat{\mathbb{P}}_{n;\beta_1,\beta_2,\beta_3}$)**.** *For all* $(\beta_1, \beta_2, \beta_3) \in \mathcal{D}^{rs} \setminus S$

$$\sqrt{6}\,\frac{T_n/n - \hat{\mathbb{E}}_{n;\beta_1,\beta_2,\beta_3}(T_n/n)}{n} \xrightarrow[n\to+\infty]{\mathrm{d}} \mathcal{N}(0, v(\beta_1,\beta_2,\beta_3)) \qquad w.r.t.\ \hat{\mathbb{P}}_{n;\beta_1,\beta_2,\beta_3},$$

where $v(\beta_1,\beta_2,\beta_3) := 18 u^{*2}_{\beta_1,\beta_2,\beta_3} \partial_{\beta_2} u^*_{\beta_1,\beta_2,\beta_3}$ *and* u^* *solves* (1.25).

Note that, unlike in the definition of α, h (see below (1.13)), this generalization retains a constant factor within the homomorphism density. This explains why the two variances in Thm. 1.19 and Thm. 1.18 differ by a factor of 6.

Remark 1.20. *Some remarks are in order.*

1. *Based on the result valid for the mean-field version of the edge-triangle model, we conjecture that the variance* $\hat{v}_0^{\Delta}(\alpha, h)$ *in Theorem 1.18 is* $\frac{3u^{*4}_{\alpha,h}}{4c_0}$, *where* c_0 *as in Theorem 1.9.*
2. *The choice of including triangles in the statistics of* (1.15) *and* (1.26) *is crucial, as it allows us to connect the expectation of* $\lfloor T_n/n \rfloor$ *and hence the scaled cumulant generating function defined in* (1.73) *below, to the derivative of the finite-size free energy. This will be a key step in the proof of Thm. 1.18.*
3. *The integer part of the normalized triangle count plays a crucial role when we represent the partition function as a polynomial (see Subsec. 1.3.8). The other subgraph counts collected in the Hamiltonian (see e.g.* (1.26)*), can be taken without such integer value, as they contribute only to the coefficients of the polynomial, and do not affect the validity of the representation.*
4. *Our main theorem covers the replica symmetric regime (except for the critical curve), without any further restrictions. Moreover, unlike the setting considered in [Fang et al., 2025a, Cor. 3.1] or [Winstein, 2025, Cor. 1.2], our framework allows the parameter controlling the triangle density to take slightly negative values.*
5. *The technique of Thm. 1.18 can be easily extended to other subgraph counts but also, in principle, to more general families of ERGs, provided that the phase diagram of the free energy is known.*

1.3 Proofs

We first we provide a short overview of the main notions and results on graph limits theory, relevant to the proof of Thm. 1.5. We refer the reader to Borgs et al. [2008, 2012], Lovász [2012] for a thorough description of these concepts.

1.3.1 Key results on graph limits

Let $(G_n)_{n\geq 1}$ be a sequence of simple, dense graphs whose number of vertices tends to infinity; the limit object of this sequence is a symmetric measurable function on the unitary square called *graphon*. A crucial step for understanding where this definition comes from, is to introduce the notion of *checkerboard graphon* Lovász [2012].

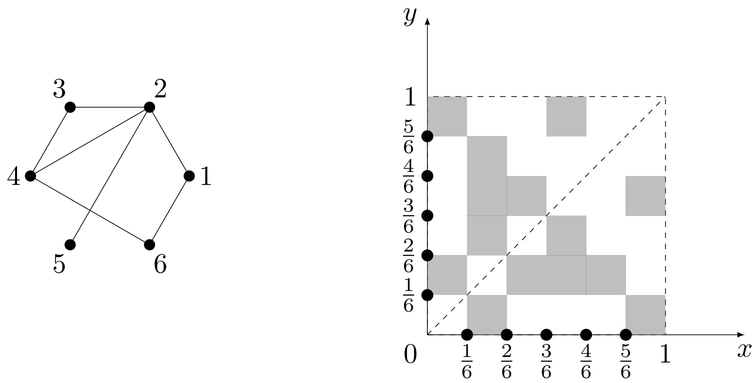

Figure 1.2: Graph H with $m = 6$ vertices on the left and its checkerboard graphon g^H on the right. The gray regions are constantly equal to one, whereas the white regions are constantly equal to zero (example from den Hollander et al. [2018]).

Let H be a finite simple graph H with vertex set $[m]$. The checkerboard graphon g^H, corresponding to H, is defined by

$$g^H(x,y) := \begin{cases} 1 & \text{if } \{\lceil mx\rceil, \lceil my\rceil\} \text{ is an edge in } H \\ 0 & \text{otherwise} \end{cases},$$

where $(x, y) \in [0, 1]^2$. In other words, g^H is a step function corresponding to the adjacency matrix of H (see Fig. 1.2). It is important to stress that any finite simple graph admits a graphon representation, therefore the sequence $(G_n)_{n\geq 1}$ can be equivalently turned into a sequence of checkerboard graphons $(g^{G_n})_{n\geq 1}$. Very intuitively, if we imagine to assign a black pixel to each block constantly equal to 1 appearing in the step function g^H (and conversely a white pixel to each block constantly equal to 0), then, as n gets large, pixels become finer and finer and the density of black pixels can be expressed as a number between 0 and 1. It is then more natural to see that the limit of $(g^{G_n})_{n\geq 1}$ can be represented by a measurable and symmetric function $g : [0, 1]^2 \to [0, 1]$ (called, indeed, graphon). The set of all graphons is denoted by $\mathcal{W}$; also notice that checkerboard graphons allow to represent all simple graphs as elements of the space $\mathcal{W}$. The definition of convergence is formalized by making use of the notion of homomorphism density $t(H, G_n)$ and its continuous analog, the so-called subgraph density

$$t(H, g) := \int_{[0,1]^m} \prod_{\{i,j\}\in\mathcal{E}(H)} g(x_i, x_j)\, dx_1 \dots dx_m, \tag{1.27}$$

where $\mathcal{E}(H)$ denotes the edge set of H and we recall that m is the number of vertices of H. A sequence of graphs $(G_n)_{n\geq 1}$ is said to converge to the graphon g if, for every finite simple graph H,

$$\lim_{n\to+\infty} t(H, G_n) = t(H, g).$$

Any sequence of graphs that converges in the appropriate way has a graphon as limit, and vice-versa every graphon arises as the limit of an appropriate graph sequence. Intuitively, the interval $[0,1]$ represents a continuum of vertices and $g(x_i, x_j)$ corresponds to the probability of drawing the edge $\{x_i, x_j\}$. For instance, the limit of a sequence of dense Erdős-Rényi graphs is represented by the function that is constantly equal to p on the unit square. In order to take into account the arbitrary labeling of the vertices when they are embedded in the unit interval, one needs to introduce an equivalence relation on $\mathcal{W}$. Let Σ be the space of all bijections $\sigma : [0,1] \to [0,1]$ preserving the Lebesgue measure. We say that the functions $g_1, g_2 \in \mathcal{W}$ are equivalent, and we write $g_1 \sim g_2$, if $g_2(x,y) = g_1(\sigma(x), \sigma(y))$ for some $\sigma \in \Sigma$. The quotient space under $\sim$ is denoted by $\widetilde{\mathcal{W}}$ and $\tau : g \mapsto \widetilde{g}$ is the natural mapping associating a graphon with its equivalence class. The space $\widetilde{\mathcal{W}}$ can be equipped with the so-called *cut distance* that turns $\widetilde{\mathcal{W}}$ into a compact metric space (see Lovász and Szegedy [2007], Thm. 5.1). On the space $(\widetilde{\mathcal{W}}, \delta_\square)$ a large deviation principle for the sequence of measures $(\widetilde{\mathbb{P}}^{\mathrm{ER}}_{n;p})_{n\geq 1}$ of a dense Erdős-Rényi random graph has been proved by Chatterjee and Varadhan in Chatterjee and Varadhan [2011]. Here $\widetilde{\mathbb{P}}^{\mathrm{ER}}_{n;p}$ denotes the probability distribution on $\widetilde{\mathcal{W}}$ induced by the Erdős-Rényi graph $G = G(n,p)$ via the mapping $G \mapsto g^G \mapsto \widetilde{g}^G$. We report below the large deviation principle:

Theorem 1.21 (Chatterjee and Varadhan [2011], Thm. 2.3)**.** *For each fixed $p \in (0,1)$, the sequence $(\widetilde{\mathbb{P}}^{ER}_{n;p})_{n\geq 1}$ satisfies a large deviation principle on the space $(\widetilde{\mathcal{W}}, \delta_\square)$, with speed n^2 and rate function*

$$\mathcal{I}_p(\widetilde{g}) = \frac{1}{2}\int_0^1\int_0^1 I_p(g(x,y))\,dx\,dy,$$

where g is any representative element of the equivalence class $\widetilde{g}$ and, for $u \in [0,1]$, we set $I_p(u) = u \ln \frac{u}{p} + (1-u)\ln\frac{1-u}{1-p}$.

We will strongly rely on Thm. 1.21 for the proof of Thm. 1.5.

1.3.2 Exponential convergence for the edge-triangle model

Proof of Theorem 1.5. The proof consists in showing that the sequence of the laws of the triangle density w.r.t. the measure $\mathbb{P}_{n;\alpha,h}$ is exponentially tight; this is made by representing the measure $\mathbb{P}_{n;\alpha,h}$ as a tilted probability measure on the space of graphons, that has as a priori measure the Erdős-Rényi one. Let H_2 be a triangle. Note that the homomorphism density is then $t(H_2, G) = \frac{6T_n(G)}{n^3}$. An important property that we are going to use is that $t(\,\cdot\,, G) = t(\,\cdot\,, \tilde{g})$, where $\widetilde{g}$ is the image in $\widetilde{\mathcal{W}}$ of the checkerboard graphon g^G of G and $t(\cdot, \tilde{g})$ is the subgraph density (1.27). This allows us to extend the edge-triangle Hamiltonian $\mathcal{H}_{n;\alpha,h}$ to the space $\widetilde{\mathcal{W}}$ replacing homomorphism densities with subgraph densities. Indeed, for all $G \in \mathcal{G}_n$,

$$\mathbb{P}_{n;\alpha,h}(G) = \frac{\exp\left(\mathcal{H}_{n;\alpha,h}(G)\right)}{\sum_{\widetilde{g}\in\widetilde{\mathcal{W}}} \sum_{G\in[\widetilde{g}]_n} \exp\left(\mathcal{H}_{n;\alpha,h}(G)\right)} = \frac{\exp\left(\mathcal{H}_{n;\alpha,h}(\widetilde{g})\right) \mathbb{1}(g^G \in \widetilde{g})}{\sum_{\widetilde{g}\in\widetilde{\mathcal{W}}} |[\widetilde{g}]_n| \exp\left(\mathcal{H}_{n;\alpha,h}(\widetilde{g})\right)}, \quad (1.28)$$

where $[\widetilde{g}]_n := \{G \in \mathcal{G}_n : g^G \in \widetilde{g}\}$ and $|\cdot|$ denotes the cardinality of a set. Notice that thanks to the fact that we can replace the homomorphism density with the subgraph density, the internal sum $\sum_{G\in[\widetilde{g}]_n}$ in the second term of (1.28) simply reduces to the cardinality of the set $[\widetilde{g}]_n$. Since for $p = \frac{1}{2}$ the Erdős-Rényi measure becomes uniform on $\mathcal{G}_n$, we can equivalently write $|[\widetilde{g}]_n| = 2^{n(n-1)/2}\mathbb{P}^{\mathrm{ER}}_{n;\frac{1}{2}}([\widetilde{g}]_n)$. As a consequence, from (1.28), we obtain

$$\mathbb{P}_{n;\alpha,h}(G) = \frac{2^{-n(n-1)/2} \exp\left(\mathcal{H}_{n;\alpha,h}(\widetilde{g})\right) \mathbb{1}(g^G \in \widetilde{g})}{\sum_{\widetilde{g}\in\widetilde{\mathcal{W}}} \exp\left(\mathcal{H}_{n;\alpha,h}(\widetilde{g})\right) \widetilde{\mathbb{P}}^{\mathrm{ER}}_{n;\frac{1}{2}}(\{\widetilde{g}\})}. \quad (1.29)$$

We now express the Hamiltonian in terms of homomorphism densities; to do so, we introduce the function

$$U_{\alpha,h}(G) := \tfrac{\alpha}{6} t(H_2, G) + \tfrac{h}{2} t(H_1, G),$$

so that $\mathcal{H}_{n;\alpha,h}(G) = n^2 U_{\alpha,h}(G)$. For each $n \geq 1$ and each Borel set $\widetilde{A} \subseteq \widetilde{\mathcal{W}}$, we define the probabilities

$$\widetilde{\mathbb{Q}}_{n;\alpha,h}(\widetilde{A}) := \frac{\sum_{\widetilde{g}\in\widetilde{A}} \exp\left(n^2 U_{\alpha,h}(\widetilde{g})\right) \widetilde{\mathbb{P}}^{\mathrm{ER}}_{n;\frac{1}{2}}(\widetilde{g})}{\sum_{\widetilde{g}\in\widetilde{\mathcal{W}}} \exp\left(n^2 U_{\alpha,h}(\widetilde{g})\right) \widetilde{\mathbb{P}}^{\mathrm{ER}}_{n;\frac{1}{2}}(\widetilde{g})}. \quad (1.30)$$

Since $U_{\alpha,h}$ is a continuous and bounded function on the metric space $(\widetilde{\mathcal{W}}, \delta_\square)$ (see Borgs et al. [2008, 2012]), by [Ellis, 1985, Thm. II.7.2], the sequence $\{\widetilde{\mathbb{Q}}_{n;\alpha,h}\}_{n\geq 1}$ of probability measures satisfies a large deviation principle with speed n^2 and rate function

$$\mathcal{I}_{\alpha,h}(\widetilde{g}) := \mathcal{I}_{\frac{1}{2}}(\widetilde{g}) - U_{\alpha,h}(\widetilde{g}) - \inf_{\widetilde{g}\in\widetilde{\mathcal{W}}} \left\{ \mathcal{I}_{\frac{1}{2}}(\widetilde{g}) - U_{\alpha,h}(\widetilde{g}) \right\}. \quad (1.31)$$

The function $\mathcal{I}_{\frac{1}{2}}$ is lower semicontinuous (see [Chatterjee and Varadhan, 2011, Lem. 2.1]), and, therefore $\mathcal{I}_{\alpha,h}$ is lower semicontinuous as well, as it is the sum of lower semicontinuous functions; thus it admits a minimizer on the compact space $\widetilde{\mathcal{W}}$. In particular, for $(\alpha, h) \in \mathcal{M}^{rs}$ the minimizers of (1.31) are given by the set $\widetilde{C}^* = \{\widetilde{u}_1^*, \widetilde{u}_2^*\}$, where $\widetilde{u}_1^*$ and $\widetilde{u}_2^*$ are the constant graphons in $\widetilde{\mathcal{W}}$ with density given by the solutions u_1^*, u_2^* to the scalar problem (1.8) (we know that they are exactly two thanks to [Radin and Yin, 2013, Prop. 3.2]). For all sufficiently small $\varepsilon > 0$, we define the open interval

$$J(\varepsilon) := (u_1^{*3} - \varepsilon, u_1^{*3} + \varepsilon) \cup (u_2^{*3} - \varepsilon, u_2^{*3} + \varepsilon)$$

and we consider the sets

$$\widetilde{C}^*_\varepsilon := \{\tilde{g} \in \widetilde{\mathcal{W}} : t(H_2, \widetilde{g}) \notin J(\varepsilon)\} \quad \text{and} \quad C^*_\varepsilon := \left\{ G \in \mathcal{G}_n : \frac{6T_n(G)}{n^3} \notin J(\varepsilon) \right\}.$$

It is important to observe that, due to (1.29) and (1.30), $\widetilde{\mathbb{Q}}_{n;\alpha,h}(\widetilde{C}^*_\varepsilon) = \mathbb{P}_{n;\alpha,h}(C^*_\varepsilon)$. Moreover, $\widetilde{C}^*_\varepsilon$ does not contain any element of $\widetilde{C}^*$, indeed, for the constant graphons $\widetilde{u}^*_i, i \in \{1,2\}$, it holds $t(H_2, \widetilde{u}^*_i) = u^{*3}_i \in J(\varepsilon) \Rightarrow \widetilde{C}^*_\varepsilon \cap \widetilde{C}^* = \emptyset$. Hence, since $\widetilde{C}^*_\varepsilon$ is closed and does not contain any minimizer of the rate functional (1.31), Thm. II.7.2(b) in Ellis [1985] guarantees that, for sufficiently large n, there is some positive constant $\tau = \tau(\widetilde{C}^*_\varepsilon) > 0$, corresponding to the infimum of the rate functional on $\widetilde{C}^*_\varepsilon$, such that $\widetilde{\mathbb{Q}}_{n;\alpha,h}(\widetilde{C}^*_\varepsilon) \le e^{-n^2\tau}$. The thesis follows since

$$\mathbb{P}_{n;\alpha,h}\left(\frac{6T_n}{n^3} \in J(\varepsilon)\right) = 1 - \widetilde{\mathbb{Q}}_{n;\alpha,h}(\widetilde{C}^*_\varepsilon) \ge 1 - e^{-n^2\tau}. \qquad \square$$

When $(\alpha, h) \in \mathcal{U}^{rs}$, namely we work in the uniqueness regime, the proof above can be carried out exactly in the same way, and it gives exponential convergence of the sequence $\left(6T_n/n^3\right)_{n\ge1}$ to u^{*3} (being $u^* = u^*_0$ or $u^* = u^*_c$, depending on (α, h)). Indeed in $\mathcal{U}^{rs}$, the set of minimizers of (1.31) coincides with the singleton $\widetilde{C}^* = \{\widetilde{u}^*\}$, where $\widetilde{u}^*$ is the image in $\widetilde{\mathcal{W}}$ of the unique solution u^* to the scalar problem (1.7). As pointed out in [Bianchi et al., 2024, Thm. 1.5], as a byproduct of this proof, we obtain an LDP for $\widetilde{\mathbb{Q}}_{n;\alpha,h}$:

Remark 1.22 (Bianchi et al. [2024], Rem. 7.5)**.** *The sequence $(\widetilde{\mathbb{Q}}_{n;\alpha,h})_{n\ge1}$ obeys a large deviation principle on the space $(\widetilde{W}, \delta_\square)$, with speed n^2 and rate function $\mathcal{I}_{\alpha,h}$.*

We are now ready to prove the strong law of large numbers stated in Thm. 1.4.

Proof of Theorem 1.4. The thesis immediately follows as a consequence of Borel-Cantelli lemma, since exponential convergence provided by Thm. 1.5 implies almost sure convergence (see Ellis [1985], Thm. II.6.4). We stress that this almost sure convergence holds w.r.t. a probability measure $\mathbb{P}_{\alpha,h}$ on the space $\left(\{0,1\}^{\mathbb{N}}, \mathcal{B}(\{0,1\}^{\mathbb{N}})\right)$, with marginals corresponding to the measures $\mathbb{P}_{n;\alpha,h}$, for all $n \in \mathbb{N}$ (see Rem. 1.1). $\square$

1.3.3 Preliminaries for the mean-field analysis

We start from some preliminaries that are preparatory to the proofs of this section. First, we observe that the Hamiltonian $\bar{\mathcal{H}}_{n;\alpha,h}$ (given in (1.14)), which is defined on $\mathcal{A}_n$, is actually a function of the edges density $m \equiv m(x) := 2E_n(x)/n^2$, $x \in \mathcal{A}_n$, taking values in the set $\Gamma_n := \left\{0, \frac{2}{n^2}, \frac{4}{n^2}, \frac{6}{n^2}, \dots, 1 - \frac{1}{n}\right\}$. In particular, for all $x \in \mathcal{A}_n$ such that $\frac{2E_n(x)}{n^2} = m$, we have

$$\bar{\mathcal{H}}_{n;\alpha,h}(x) = \bar{\mathcal{H}}_{n;\alpha,h}(m) = n^2\left(\frac{\alpha}{6}m^3 + \frac{h}{2}m\right).$$

As a consequence, we can also write, with a little abuse of notation,

$$\bar{\mathbb{P}}_{n;\alpha,h}(A) = \sum_{m\in A} \frac{\mathcal{N}_m e^{n^2(\frac{\alpha}{6}m^3+\frac{h}{2}m)}}{\bar{Z}_{n;\alpha,h}}, \qquad \text{for } A \subseteq \Gamma_n, \tag{1.32}$$

where $\mathcal{N}_m := \binom{\frac{n(n-1)}{2}}{\frac{n(n-1)m}{2}}$ coincides with the number of adjacency matrices in $\mathcal{A}_n$ with edge density m. From representation (1.32) arises another important function, that we call energy function; for any $(\alpha, h) \in \mathbb{R}^2$, it is defined as

$$g_{\alpha,h}(m) := \frac{\alpha}{6}m^3 + \frac{h}{2}m - \frac{I(m)}{2}, \qquad \text{for } m \in \Gamma_n. \tag{1.33}$$

The first two terms coincide with the exponent in (1.32), whereas the entropic term $I(m) = m\ln m + (1-m)\ln(1-m)$, defined below (1.3), comes from the Stirling approximation of the binomial coefficient $\mathcal{N}_m$. Let $f_{\alpha,h}$ be the infinite volume free energy of the edge-triangle model, as given in (1.7), and let u_i^*, $i = 0, 1, 2$, or u_c^* be a solution of (1.8), depending on the range of (α, h). We stress that, by construction, $g_{\alpha,h}(u_i^*) = f_{\alpha,h}$ if $(\alpha, h) \neq (\alpha_c, h_c)$, and $g_{\alpha_c,h_c}(u_c^*) = f_{\alpha_c,h_c}$.

Neighborhoods of the maximizer(s). Fix $0 < \delta < 1$. We will mainly work in the following neighborhoods of the maximizer(s) (whose definition was anticipated in (1.21)):

$$B_{u_i^*} \equiv B_{u_i^*}(n, \delta) = \{m \in \Gamma_n \,:\, |m - u_i^*| \leq n^{-\delta}\}, \quad i = 0, 1, 2 \tag{1.34}$$

$$B_{u_c^*} \equiv B_{u_c^*}(n, \delta) = \{m \in \Gamma_n \,:\, |m - u_c^*| \leq n^{-\delta}\} \tag{1.35}$$

making extensive use of the following Taylor expansions from Bianchi et al. [2022]:

$$\begin{cases} g_{\alpha,h}(m) - g_{\alpha,h}(u_i^*) = -c_i(m-u_i^*)^2 + k_i(m-u_i^*)^3 & \text{if } (\alpha,h) \neq (\alpha_c,h_c) \\ g_{\alpha,h}(m) - g_{\alpha,h}(u_c^*) = -\frac{81}{64}(m-u_c^*)^4 + k_c(m-u_c^*)^5 & \text{if } (\alpha,h) = (\alpha_c,h_c), \end{cases} \tag{1.36}$$

where

$$c_i := -\frac{g''_{\alpha,h}(u_i^*)}{2} = \frac{1 - 2\alpha(u_i^*)^2(1-u_i^*)}{4u_i^*(1-u_i^*)} > 0, \tag{1.37}$$

$$k_i := g'''_{\alpha,h}(\tilde{u}_i)/6, \quad k_c := g^{(v)}_{\alpha_c,h_c}(\tilde{u}_c)/5!, \tag{1.38}$$

for some $\tilde{u}_i, \tilde{u}_c$ such that $|\tilde{u}_i - u_i^*| < n^{-\delta}$, $|\tilde{u}_c - u_c^*| < n^{-\delta}$.

Lattice sets. As a result of suitable changes of variables, obtained as fluctuations of $m \in \Gamma_n$, we will need to consider the following integration ranges:

$$R_{i,\delta}^{(n)} := \left(-n^{1-\delta}, n^{1-\delta}\right) \cap \left\{-nu_i^*, -nu_i^* + \frac{2}{n}, \ldots, -nu_i^* + (n-1)\right\}, \tag{1.39}$$

$$R_{c,\delta}^{(n)} := \left(-n^{\frac{1}{2}-\delta}, n^{\frac{1}{2}-\delta}\right) \cap \left\{-\sqrt{n}u_c^*, -\sqrt{n}u_c^* + \frac{2}{n^{\frac{3}{2}}}, \ldots, -\sqrt{n}u_c^* + \frac{n-1}{\sqrt{n}}\right\}, \tag{1.40}$$

where $i = 0, 1, 2$. The following lemma shows how the main contribution to $\bar{Z}_{n;\alpha,h}$ is given by sums over these sets.

Lemma 1.23 (Bianchi et al. [2024], Lemma 8.1)**.** *Let $(\alpha, h) \in (-2, +\infty) \times \mathbb{R}$ and let $f_{\alpha,h}$ be the infinite volume free energy of the edge-triangle model given in (1.7).*

Fix $\delta \in (0, 1)$. If $(\alpha, h) \neq (\alpha_c, h_c)$ and c_i, k_i as in (1.37)–(1.38), let

$$D_i^{(n)} := \sum_{x \in R_{i,\delta}^{(n)}} \frac{2}{n} \frac{e^{-c_i x^2 + \frac{k_i}{n} x^3}}{\sqrt{\left(u_i^* + \frac{x}{n}\right)\left(1 - u_i^* - \frac{x}{n}\right)}}, \quad \text{for } i = 0, 1, 2. \tag{1.41}$$

Fix $\delta \in \left(0, \frac{3}{8}\right)$. If $(\alpha, h) = (\alpha_c, h_c)$ and k_c as in (1.38), let

$$D_c^{(n)} := \sum_{x \in R_{c,\delta}^{(n)}} \frac{2}{n^{3/2}} \frac{e^{-\frac{81}{64} x^4 + \frac{k_c}{\sqrt{n}} x^5}}{\sqrt{\left(u_c^* + \frac{x}{\sqrt{n}}\right)\left(1 - u_c^* - \frac{x}{\sqrt{n}}\right)}}. \tag{1.42}$$

Then, as $n \to +\infty$,

$$\bar{Z}_{n;\alpha,h} = \frac{e^{n^2 f_{\alpha,h}}}{2\sqrt{\pi}} \left(D^{(n)}(\alpha, h)\right)(1 + o(1)), \tag{1.43}$$

where

$$D^{(n)}(\alpha, h) := \begin{cases} D_0^{(n)} & \text{if } (\alpha, h) \in \mathcal{U}^{rs} \setminus \{(\alpha_c, h_c)\} \\ D_1^{(n)} + D_2^{(n)} & \text{if } (\alpha, h) \in \mathcal{M}^{rs} \\ \sqrt{n} D_c^{(n)} & \text{if } (\alpha, h) = (\alpha_c, h_c) \end{cases}.$$

Remark 1.24. *Lem. 1.23 directly proves (1.12).*

Remark 1.25. *The quantities defined in (1.41)–(1.42) are Riemann sums with volume elements respectively given by $2/n$ and $2/n^{3/2}$. Indeed the points $x \in R_i^{(n)}$, $i = 0, 1, 2$ (resp. inside $R_c^{(n)}$) are evenly spaced with gaps of length $2/n$ (resp. $2/n^{3/2}$). It is possible to show that on the ranges $R_{i,\delta}^{(n)}$ it holds $e^{-c_i x^2 + \frac{k_i}{n} x^3} \leq e^{-(c_i - k_i n^{-\delta}) x^2}$ which is in turn bounded by $e^{-0.99 c_i x^2}$ for sufficiently large n. With this domination and point-wise convergence, we get (similar bounds can be done at criticality),*

$$D_i^{(n)} \xrightarrow{n \to +\infty} D_i := 2\sqrt{\pi \left[1 - 2\alpha \left(u_i^*\right)^2 \left(1 - u_i^*\right)\right]^{-1}}, \qquad i \in \{0, 1, 2\} \tag{1.44}$$

$$D_c^{(n)} \xrightarrow{n \to +\infty} D_c := \frac{3}{\sqrt{2}} \int_{-\infty}^{\infty} e^{-\frac{81}{64} x^4} dx \approx 3.63\,, \tag{1.45}$$

where for (1.45) we used $u_c^* = \frac{2}{3}$. These terms will play the role of normalization weights.

1.3.4 Phase coexistence on the critical curve and SLLN

Proof of Theorem 1.7. In this part, we denote by u^* both terms u_0^* or u_c^*, indistinguishably (recall that the region $(\alpha, h) \in \mathcal{U}^{rs}$ includes the critical point (α_c, h_c)). Fix $\varepsilon > 0$ and set $K(\varepsilon) := \{m \in \Gamma_n : |m - u^*| \geq \varepsilon/3\}$. We prove almost sure convergence via the exponential convergence of $\frac{6\bar{T}_n}{n^3}$ to u^{*3} (similarly to [Ellis, 1985, Thm. IV.4.1]). Indeed,

$$\begin{aligned}\bar{\mathbb{P}}_{n;\alpha,h}\left(\left|\frac{6\bar{T}_n}{n^3} - u^{*3}\right| \geq \varepsilon\right) &\leq \sum_{m\in K(\varepsilon)} \frac{\mathcal{N}_m e^{n^2\left(\frac{\alpha}{6}m^3+\frac{h}{2}m\right)}}{\bar{Z}_{n;\alpha,h}} \\ &\leq Cn^3 e^{-n^2(f_{\alpha,h}-\max_{m\in K(\varepsilon)} g_{\alpha,h}(m))}\frac{2\sqrt{\pi}}{D^{(n)}(\alpha,h)}(1+o(1)) \\ &= Cn^3 e^{-n^2 \min_{m\in K(\varepsilon)}(f_{\alpha,h}-g_{\alpha,h}(m))}\frac{2\sqrt{\pi}}{D^{(n)}(\alpha,h)}(1+o(1)),\end{aligned}$$

where in the second inequality we used (1.43), together with the following Stirling approximation (see [Friedli and Velenik, 2017, Chap. 2, Eq. (2.11)]):

$$cn^{-1}e^{-\frac{n^2}{2}I(m)} \leq \mathcal{N}_m \leq Cne^{-\frac{n^2}{2}I(m)}, \tag{1.46}$$

where c and C are positive constants. As stated in [Radin and Yin, 2013, Prop. 3.2], when $(\alpha, h) \in \mathcal{U}^{rs}$ and for sufficiently small ε, the function $f_{\alpha,h} - g_{\alpha,h}(m)$ is positive, convex and admits u^* as unique zero. Therefore, the minimum appearing at the exponent is strictly positive, and we obtain the exponential convergence. Finally, almost sure convergence follows as a consequence of Borel-Cantelli lemma (Ellis [1985], Thm. II.6.4). □

Proof of Theorem 1.8. We will determine the limit of

$$\bar{\mathbb{E}}_{n;\alpha,h}\left[\varphi\left(6\bar{T}_n/n^3\right)\right],$$

for any continuous and bounded real function φ. First we observe that, since $m \equiv m(x) = \frac{2E_n(x)}{n^2}$, we obtain $\bar{T}_n(x) = \frac{n^3m^3}{6}$. Then, using (1.32), we get:

$$\bar{\mathbb{E}}_{n;\alpha,h}\left[\varphi\left(\frac{6\bar{T}_n}{n^3}\right)\right] = \sum_{m\in\Gamma_n} \varphi(m^3)\frac{\mathcal{N}_m e^{n^2\left(\frac{\alpha}{6}m^3+\frac{h}{2}m\right)}}{\bar{Z}_{n;\alpha,h}}. \tag{1.47}$$

We split the sum in (1.47) over the sets $B_{u_1^*}$, $B_{u_2^*}$ given in (1.34), and $\mathcal{C} \equiv \mathcal{C}(n,\delta) := \Gamma_n \setminus \left(B_{u_1^*} \cup B_{u_2^*}\right)$, considering the three contributions separately. First we observe that whenever we work inside the sets $B_{u_i^*}$, $i = 1, 2$, the bounds in (1.46) can be made more precise, because $n^{-2} \ll m \ll 1 - n^{-2}$ and, consequently, $n^2 m \to \infty$ and $n^2(1-m) \to \infty$, as $n \to \infty$. Hence, the following Stirling approximation is valid

$$\mathcal{N}_m = \frac{e^{-\frac{n^2}{2}I(m)}}{n\sqrt{\pi m(1-m)}}(1+o(1)). \tag{1.48}$$

This, together with Lem. 1.23, yields the following representation:

$$\sum_{m\in B_{u_i^*}} \varphi(m^3)\frac{\mathcal{N}_m e^{n^2(\frac{\alpha}{6}m^3+\frac{h}{2}m)}}{\bar{Z}_{n;\alpha,h}} = \sum_{m\in B_{u_i^*}} \frac{2}{n}\frac{\varphi(m^3)}{\sqrt{m(1-m)}}\frac{e^{-n^2\left(f_{\alpha,h}-g_{\alpha,h}(m)\right)}}{D_1^{(n)}+D_2^{(n)}}(1+o(1))\,,$$

where $i\in\{1,2\}$ and $g_{\alpha,h}$ is the energy function defined in (1.33). By performing the change of variable $x=n(m-u_i^*)$, and using the Taylor expansion (1.36), we obtain:

$$\begin{aligned}\sum_{m\in B_{u_i^*}} \varphi(m^3)\frac{\mathcal{N}_m e^{n^2(\frac{\alpha}{6}m^3+\frac{h}{2}m)}}{\bar{Z}_{n;\alpha,h}} &= \sum_{x\in R_{i,\delta}^{(n)}} \frac{2}{n}\frac{\varphi((u_i^*+\frac{x}{n})^3)}{\sqrt{(u_i^*+\frac{x}{n})(1-u_i^*-\frac{x}{n})}}\frac{e^{-c_i x^2+\frac{k_i}{n}x^3}}{D_1^{(n)}+D_2^{(n)}}(1+o(1))\\ &\xrightarrow{n\to+\infty} \varphi(u_i^{*3})\frac{D_i}{D_1+D_2}\,,\end{aligned}$$

where D_i is defined in (1.44) and $R_{i,\delta}^{(n)}$, given in (1.39), represents the range of values of x. To conclude the analysis, we show that the sum over the remaining set $\mathcal{C}$ provides no contribution in the limit. Outside the sets $B_{u_i^*}$, $i=1,2$, the Stirling approximation (1.48) is not valid anymore. However, from (1.46) we deduce:

$$\sum_{m\in\mathcal{C}} \mathcal{N}_m e^{n^2(\frac{\alpha}{6}m^3+\frac{h}{2}m)} \le Cn\sum_{m\in\mathcal{C}} e^{n^2 g_{\alpha,h}(m)} < \frac{C}{2}n^3 e^{n^2\max_{m\in\mathcal{C}} g_{\alpha,h}(m)}, \tag{1.49}$$

where the last inequality is due to the fact that $\mathcal{C}$ contains at most $\binom{n}{2}$ points. Since $f_{\alpha,h}=g_{\alpha,h}(u_i^*)$, $i=1,2$ we obtain

$$e^{n^2\max_{m\in\mathcal{C}} g_{\alpha,h}(m)} = e^{n^2 f_{\alpha,h}}e^{-n^2(g_{\alpha,h}(u_i^*)-\max_{m\in\mathcal{C}} g_{\alpha,h}(m))} \le e^{n^2 f_{\alpha,h}}e^{-kn^{2-2\delta}}, \tag{1.50}$$

where $k>0$ is a constant that does not depend on n and $\delta\in(0,1)$. The last inequality follows from the Taylor expansion (1.36), exploiting the fact that $|m-u_i^*|>n^{-\delta}$ for $i=1,2$ and for all $m\in\mathcal{C}$. As a consequence, from the rough Stirling approximation (1.46) and the consequent rough bound $\bar{Z}_{n;\alpha,h}\ge \min_{m\in\Gamma_n}\mathcal{N}_m e^{n^2(\frac{\alpha}{6}m^3+\frac{h}{2}m)}\ge cn^{-1}e^{n^2 f_{\alpha,h}}$, we get

$$\frac{\sum_{m\in\mathcal{C}}\mathcal{N}_m e^{n^2(\frac{\alpha}{6}m^3+\frac{h}{2}m)}}{\bar{Z}_{n;\alpha,h}} < c^{-1}Cn^4 e^{-kn^{2-2\delta}} \xrightarrow{n\to+\infty} 0, \tag{1.51}$$

being $2-2\delta>0$ by assumption. In conclusion,

$$\lim_{n\to+\infty}\bar{\mathbb{E}}_{n;\alpha,h}\left[\varphi\left(\frac{6\bar{T}_n}{n^3}\right)\right] = \varphi(u_1^{*3})\frac{D_1}{D_1+D_2}+\varphi(u_2^{*3})\frac{D_2}{D_1+D_2}\,.$$

This proves the thesis. □

1.3.5 Rate of convergence of triangle density

Proof of Proposition 1.11. The proof implements the same machinery of Thm. 1.8. Let $(\alpha, h) \in \mathcal{U}^{rs}$; we analyze the case $(\alpha, h) \in \mathcal{U}^{rs} \setminus \{(\alpha_c, h_c)\}$ first, and then we move to the critical point. Note that

$$\bar{\mathbb{E}}_{n;\alpha,h}\left(\left|\frac{6\bar{T}_n}{n^3} - u_0^{*3}\right|\right) = \sum_{m\in\Gamma_n} |m^3 - u_0^{*3}| \frac{\mathcal{N}_m e^{n^2(\frac{\alpha}{6}m^3+\frac{h}{2}m)}}{\bar{Z}_{n;\alpha,h}}. \tag{1.52}$$

We split the average in (1.52) in two parts, one over $B_{u_0^*}$ given in (1.34), and the other over $\mathcal{C} \equiv \mathcal{C}(n,\delta) := \Gamma_n \setminus B_{u_0^*}$. The contribution of the average over $\mathcal{C}$ is negligible, exploiting exactly the same argument in (1.49)–(1.51) with u_0^* in place of u_i^* $(i = 1, 2)$, and bounding, very roughly, $|m^3 - u_0^{*3}| = |m - u_0^*|(m^2 + mu_0^* + u_0^{*2})$ by the constant 3. We now focus on the sum over $B_{u_0^*}$. By Lem. 1.23 and the Stirling approximation (1.48), we obtain

$$\begin{aligned}\bar{\mathbb{E}}_{n;\alpha,h}\left(\left|\frac{6\bar{T}_n}{n^3} - u_0^{*3}\right|\right) &= \sum_{m\in B_{u_0^*}} |m^3 - u_0^{*3}| \frac{\mathcal{N}_m e^{n^2(\frac{\alpha}{6}m^3+\frac{h}{2}m)}}{\bar{Z}_{n;\alpha,h}}(1+o(1)) \\ &= \sum_{m\in B_{u_0^*}} \frac{2}{n}\frac{|m^3 - u_0^{*3}|}{\sqrt{m(1-m)}} \frac{e^{-n^2(f_{\alpha,h}-g_{\alpha,h}(m))}}{D_0^{(n)}}(1+o(1)),\end{aligned}$$

where we recall that $\bar{T}_n(x) = \frac{n^3m^3}{6}$ and $B_{u_0^*}$ is defined in (1.34). From the Taylor expansion (1.36) we get

$$\bar{\mathbb{E}}_{n;\alpha,h}\left(\left|\frac{6\bar{T}_n}{n^3} - u_0^{*3}\right|\right) = \sum_{m\in B_{u_0^*}} \frac{2}{n}\frac{|m^3 - u_0^{*3}|}{\sqrt{m(1-m)}} \frac{e^{-n^2c_0(m-u_0^*)^2+n^2k_0(m-u_0^*)^3}}{D_0^{(n)}}(1+o(1)). \tag{1.53}$$

We now perform the change of variable $x = n(m - u_0^*)$, and we use the identity

$$m^3 - u_0^{*3} = \left(u_0^* + \frac{x}{n}\right)^3 - u_0^{*3} = 3u_0^{*2}\frac{x}{n} + 3u_0^*\frac{x^2}{n^2} + \frac{x^3}{n^3}, \tag{1.54}$$

thus obtaining

$$n\cdot\bar{\mathbb{E}}_{n;\alpha,h}\left(\left|\frac{6\bar{T}_n}{n^3} - u_0^{*3}\right|\right) = \sum_{x\in R_{0,\delta}^{(n)}} \frac{2}{n}\frac{|3u_0^{*2}x + 3u_0^*\frac{x^2}{n} + \frac{x^3}{n^2}|\cdot e^{-c_0x^2+\frac{k_0}{n}x^3}}{\sqrt{\left(u_0^*+\frac{x}{n}\right)\left(1-u_0^*-\frac{x}{n}\right)}\cdot D_0^{(n)}}(1+o(1)), \tag{1.55}$$

where $R_{0,\delta}^{(n)}$ is defined in (1.39) and the constants c_0 and k_0 are given in (1.37)–(1.38). Consider the term

$$\sum_{x\in R_{0,\delta}^{(n)}} \frac{2}{n}\frac{|3u_0^{*2}x|\cdot e^{-c_0x^2+\frac{k_0}{n}x^3}}{\sqrt{\left(u_0^*+\frac{x}{n}\right)\left(1-u_0^*-\frac{x}{n}\right)}\cdot D_0^{(n)}}(1+o(1)),$$

together with the sequence of probability densities

$$\ell_n(x) := \frac{2}{n} \frac{e^{-c_0 x^2 + \frac{k_0}{n} x^3}}{\sqrt{\left(u_0^* + \frac{x}{n}\right)\left(1 - u_0^* - \frac{x}{n}\right) \cdot D_0^{(n)}}} \mathbb{1}_{R_0^{(n)}}(x)\,, \quad x \in \mathbb{R}\,, \tag{1.56}$$

where $D_0^{(n)}$ is the normalization weight defined in (1.41). If, for every $n \in \mathbb{N}$, X_n is a random variable with density ℓ_n, then

$$\sum_{x \in R_{0,\delta}^{(n)}} \frac{2}{n} \frac{|3u_0^{*2} x| \cdot e^{-c_0 x^2 + \frac{k_0}{n} x^3}}{\sqrt{\left(u_0^* + \frac{x}{n}\right)\left(1 - u_0^* - \frac{x}{n}\right) \cdot D_0^{(n)}}} (1 + o(1)) = \mathbb{E}(|3u_0^{*2} X_n|)(1 + o(1)).$$

Collecting all contributions, we can upper and lower bound the sum in (1.55) using the following chain of inequalities: $|a| - |b| \le ||a| - |b|| \le |a + b| \le |a| + |b|$, $a, b \in \mathbb{R}$, with $a := 3u_0^{*2} x$ and $b := 3u_0^* \frac{x^2}{n} + \frac{x^3}{n^2}$. We have:

$$\mathbb{E}(|3u_0^{*2} X_n|)(1 + o(1)) - \left(\int_{\mathbb{R}} \left| 3u_0^* \frac{x^2}{n} + \frac{x^3}{n^2} \right| \ell_n(x) dx \right) (1 + o(1)) \tag{1.57}$$

$$\le n \cdot \bar{\mathbb{E}}_{n;\alpha,h}\left(\left| \frac{6\bar{T}_n}{n^3} - u_0^{*3} \right| \right) \le$$

$$\mathbb{E}(|3u_0^{*2} X_n|)(1 + o(1)) + \left(\int_{\mathbb{R}} \left| 3u_0^* \frac{x^2}{n} + \frac{x^3}{n^2} \right| \ell_n(x) dx \right) (1 + o(1)). \tag{1.58}$$

Notice that, reasoning as for the convergence $D_0^{(n)} \xrightarrow{n \to +\infty} D_0$ (see Rem. 1.25) and the Scheffé Lemma, we obtain $X_n \xrightarrow{d} X$, where X is a Gaussian random variable with density

$$\ell(x) = \sqrt{\frac{c_0}{\pi}} e^{-c_0 x^2}\,, \quad x \in \mathbb{R}\,. \tag{1.59}$$

Moreover, the random variables X_n have finite exponential moments for any sufficiently large n. Therefore, by the dominated convergence theorem, applied to both bounds in (1.57)–(1.58), we obtain

$$n \cdot \bar{\mathbb{E}}_{n;\alpha,h}\left(\left| \frac{6\bar{T}_n}{n^3} - u_0^{*3} \right| \right) \xrightarrow{n \to +\infty} 3u_0^{*2} \mathbb{E}(|X|).$$

Indeed, the second summand in (1.57)–(1.58) vanishes, being $3u_0^* \frac{x^2}{n} + \frac{x^3}{n^2} = o(1)$, for fixed x. Setting $\bar{X} := 3u_0^{*2} X$, and noticing that X has variance $(2c_0)^{-1}$, we recover (1.18).
We now move to the critical case, so we consider $(\alpha, h) = (\alpha_c, h_c)$ and $u_c^* = u^*(\alpha_c, h_c)$. Here, the proof works exactly the same. We split the average in (1.52) in two parts, one over $B_{u_c^*}$ given in (1.35), and the other over $\mathcal{C} \equiv \mathcal{C}(n, \delta) := \Gamma_n \setminus B_{u_c^*}$. The contribution of the average over $\mathcal{C}$ is negligible, exploiting the same argument in (1.49)–(1.51). This time, by injecting the Taylor expansion (1.36) at the critical point in (1.50), and using the fact that $|m - u_c^*|^4 > n^{-4\delta}$ for $m \in \mathcal{C}$, we get

$$\frac{\sum_{m \in \mathcal{C}} \mathcal{N}_m e^{n^2 \left(\frac{\alpha_c}{6} m^3 + \frac{h_c}{2} m\right)}}{\bar{Z}_{n;\alpha_c,h_c}} < c^{-1} C n^4 e^{-k n^{2 - 4\delta}} \xrightarrow{n \to +\infty} 0, \qquad k > 0,$$

since $\delta < 3/8$ by assumption. We now restrict the average (1.52) to a sum in $B_{u_c^*}$. In place of (1.53) we get:

$$\bar{\mathbb{E}}_{n;\alpha_c,h_c}\left(\left|\frac{6\bar{T}_n}{n^3} - u_c^{*3}\right|\right) = \sum_{m\in B_{u_c^*}} \frac{2}{n^{\frac{3}{2}}} \frac{|m^3 - u_c^{*3}|e^{-n^2\frac{81}{64}(m-u_c^*)^4 + n^2 k_c (m-u_c^*)^5}}{\sqrt{m\,(1-m)}\cdot D_c^{(n)}}(1+o(1)),$$

where $B_{u_c^*}$ is defined in (1.35), and the constant k_c is given in (1.38). Notice that here the Taylor expansion (1.36) provides the fourth-order term $(m - u_c^*)^4$ at the exponent, while Lem. 1.23 brings the normalization weight $D_c^{(n)}$. After the change of variable $y := \sqrt{n}(m - u_c^*)$, recalling that $u_c^* = \frac{2}{3}$ and by means of the identity

$$m^3 - u_c^{*3} = \left(u_c^* + \frac{y}{\sqrt{n}}\right)^3 - u_c^{*3} = \frac{4}{3}\frac{y}{\sqrt{n}} + 2\frac{y^2}{n} + \frac{y^3}{n^{\frac{3}{2}}},$$

we obtain

$$\sqrt{n}\cdot\bar{\mathbb{E}}_{n;\alpha_c,h_c}\left(\left|\frac{6\bar{T}_n}{n^3} - u_c^{*3}\right|\right) = \sum_{y\in R_{c,\delta}^{(n)}} \frac{2}{n^{\frac{3}{2}}} \frac{|\frac{4}{3}y + 2\frac{y^2}{\sqrt{n}} + \frac{y^3}{n}|\cdot e^{-\frac{81}{64}y^4 + \frac{k_c}{\sqrt{n}}y^5}}{\sqrt{\left(u_c^* + \frac{y}{\sqrt{n}}\right)\left(1 - u_c^* - \frac{y}{\sqrt{n}}\right)D_c^{(n)}}}(1+o(1)),$$

where $R_{c,\delta}^{(n)}$ is given in (1.40). For every $n \in \mathbb{N}$, let Y_n be a real random variable with Lebesgue density

$$\ell_n^c(y) := \frac{2}{n^{3/2}} \frac{e^{-\frac{81}{64}y^4 + \frac{k_c}{\sqrt{n}}y^5}}{\sqrt{(u_c^* + \frac{y}{\sqrt{n}})(1 - u_c^* - \frac{y}{\sqrt{n}})\cdot D_c^{(n)}}}\, \mathbb{1}_{R_{c,\delta}^{(n)}}(y)\,, \quad y \in \mathbb{R}\,. \tag{1.60}$$

Note that $D_c^{(n)}$ provides the right normalization rate. The random variable Y_n has finite exponential moments for any sufficiently large n. By Scheffé Lemma and dominated convergence theorem, we conclude

$$\sqrt{n}\cdot\bar{\mathbb{E}}_{n;\alpha_c,h_c}\left(\left|\frac{6\bar{T}_n}{n^3} - u_c^{*3}\right|\right) \xrightarrow{n\to+\infty} \frac{4}{3}\mathbb{E}(|Y|),$$

where Y is a generalized Gaussian random variable with Lebesgue density $\ell^c(y) \propto e^{-\frac{81}{64}y^4}$. Setting $\bar{Y} := \frac{4}{3}Y$, since Y has scale parameter $\frac{2^{3/2}}{3}$, $\bar{Y}$ is a generalized Gaussian random variable with scale parameter $\frac{2^{7/2}}{3^2}$, thus proving the thesis. □

With the same strategy we can immediately prove the following corollary.

Proof of Corollary 1.12. Recall that $\bar{m}_n^{\Delta}(\alpha, h) = \bar{\mathbb{E}}_{n;\alpha,h}\left(\frac{6\bar{T}_n}{n^3}\right)$. By following the proof of Prop. 1.11, we obtain,

- for all $(\alpha, h) \in \mathcal{U}^{rs} \setminus \{(\alpha_c, h_c)\}$,

$$n\cdot\left(\bar{m}_n^{\Delta}(\alpha, h) - u_0^{*3}(\alpha, h)\right) = \mathbb{E}\left(3u_0^{*2}X_n\right)(1+o(1)) + o(1) \xrightarrow{n\to+\infty} \mathbb{E}(\bar{X}) = 0,$$

- for $(\alpha, h) = (\alpha_c, h_c)$,

$$\sqrt{n} \cdot \Big(\bar{m}_n^{\Delta}(\alpha_c, h_c) - u_c^{*3}(\alpha_c, h_c)\Big) = \mathbb{E}\Big(\frac{4}{3}Y_n\Big)(1 + o(1)) + o(1) \xrightarrow{n\to+\infty} \mathbb{E}(\bar{Y}) = 0.$$

□

1.3.6 Standard and non-standard mean-field CLT

We use Cor. 1.12 to prove Thms. 1.9 and 1.10. We start with the analysis at criticality.

Proof of Theorem 1.10. Let $u_c^* = u^*(\alpha_c, h_c)$. We consider the decomposition

$$6\,\frac{\frac{\bar{T}_n}{n} - \frac{n^2}{6}\bar{m}_n^{\Delta}(\alpha_c, h_c)}{n^{3/2}} = \bar{U}_n + \sqrt{n}\left(u_c^{*3} - \bar{m}_n^{\Delta}(\alpha_c, h_c)\right),$$

where

$$\bar{U}_n := 6\,\frac{\frac{\bar{T}_n}{n} - \frac{n^2}{6}u_c^{*3}}{n^{3/2}}.$$

By (1.20) and Slutsky theorem (see [Klenke, 2020, Thm. 13.18]), it is enough to study the convergence in distribution of the variable $6(\frac{\bar{T}_n}{n} - \frac{n^2}{6}u_c^{*3})/n^{3/2}$. We show that, for any $t \in \mathbb{R}$,

$$\bar{M}_n(t) := \bar{\mathbb{E}}_{n;\alpha_c,h_c}\left(e^{t\cdot\bar{U}_n}\right) \xrightarrow{n\to+\infty} \int_{\mathbb{R}} e^{ty}\bar{\ell}^c(y)dy, \tag{1.61}$$

where $\bar{\ell}^c$ is given in the statement. Again, we split the average in (1.61) in two parts, one over $B_{u_c^*}$ given in (1.35), and the other over $\mathcal{C} \equiv \mathcal{C}(n,\delta) := \Gamma_n \setminus (B_{u_c^*})$. We obtain:

$$\begin{aligned}\sum_{m\in\mathcal{C}} \frac{\mathcal{N}_m e^{t\sqrt{n}(m^3-u_c^{*3})+n^2(\frac{\alpha_c}{6}m^3+\frac{h_c}{2}m)}}{\bar{Z}_{n;\alpha_c,h_c}} &\overset{(1.46)}{\leq} \sum_{m\in\mathcal{C}} \frac{Ce^{3t\sqrt{n}+n^2 g_{\alpha_c,h_c}(m)}}{\bar{Z}_{n;\alpha_c,h_c}} \\ &\leq \frac{Cn^3 e^{3t\sqrt{n}+n^2 f_{\alpha_c,h_c}-n^2(g_{\alpha,h}(u_c^*)-\max_{m\in\mathcal{C}} g_{\alpha_c,h_c}(m))}}{\bar{Z}_{n;\alpha_c,h_c}} \\ &\leq c^{-1}Cn^3e^{3t\sqrt{n}-kn^{2-4\delta}} \xrightarrow{n\to+\infty} 0,\end{aligned} \tag{1.62}$$

for some constant $k > 0$. In the second to last inequality we used that $f_{\alpha_c,h_c} = g_{\alpha_c,h_c}(u_c^*)$, and that the set $\mathcal{C}$ contains at most $\binom{n}{2}$ elements. In the last inequality we used the Taylor expansion (1.36) at the critical point, the rough bound $\bar{Z}_{n;\alpha,h} \geq cn^{-1}e^{n^2 f_{\alpha,h}}$ coming from (1.46), and the fact that $|m - u_c^*|^4 > n^{-4\delta}$ for $m \in \mathcal{C}$. The assumption $\delta < 3/8$ guarantees that $2 - 4\delta > 1/2$. As a consequence of (1.62), we can reduce the average in (1.61) into a sum on $B_{u_c^*}$:

$$\begin{aligned}\bar{M}_n(t) &= \sum_{m\in B_{u_c^*}} \frac{\mathcal{N}_m e^{t\sqrt{n}(m^3-u_c^{*3})+n^2(\frac{\alpha_c}{6}m^3+\frac{h_c}{2}m)}}{\bar{Z}_{n;\alpha_c,h_c}}(1+o(1)) \\ &= \sum_{m\in B_{u_c^*}} \frac{2}{n^{\frac{3}{2}}} \frac{e^{t\sqrt{n}(m^3-u_c^{*3})-n^2(f_{\alpha_c,h_c}-g_{\alpha_c,h_c}(m))}}{\sqrt{m(1-m)}\cdot D_c^{(n)}}(1+o(1)),\end{aligned} \tag{1.63}$$

where the last identity is due to Lem. 1.23 and the Stirling approximation (1.48). Injecting in (1.63) the Taylor expansion (1.36) at the critical point, we get

$$\bar{M}_n(t) = \sum_{m \in B_{u_c^*}} \frac{2}{n^{\frac{3}{2}}} \frac{e^{t\sqrt{n}(m^3 - u_c^{*3}) - \frac{81}{64}n^2(m-u_c^*)^4 + k_c n^2 (m-u_c^*)^5}}{\sqrt{m(1-m)} \cdot D_c^{(n)}}(1+o(1)).$$

By the change of variable $y = \sqrt{n}(m - u_c^*)$, and recalling that $u_c^*(\alpha_c, h_c) = \frac{2}{3}$, we find

$$\bar{M}_n(t) = \sum_{y \in R_{c,\delta}^{(n)}} \frac{2}{n^{\frac{3}{2}}} \frac{e^{t(\frac{4}{3}y + 2\frac{y^2}{\sqrt{n}} + \frac{y^3}{n})} \cdot e^{-\frac{81}{64}y^4 + k_c \frac{y^5}{\sqrt{n}}}}{\sqrt{\left(u_c^* + \frac{y}{\sqrt{n}}\right)\left(1 - u_c^* - \frac{y}{\sqrt{n}}\right)} D_c^{(n)}}(1+o(1)).$$

Exploiting the range of $R_{c,\delta}^{(n)}$, given in (1.40), we observe that $-n^{1/2-3\delta} < 2\frac{y^2}{\sqrt{n}} + \frac{y^3}{n} < 2n^{1/2-2\delta} + n^{1/2-3\delta}$. By isolating the term

$$\bar{M}_n^*(t) := \sum_{y \in R_{c,\delta}^{(n)}} \frac{2}{n^{\frac{3}{2}}} \frac{e^{t\frac{4}{3}y} \cdot e^{-\frac{81}{64}y^4 + k_c \frac{y^5}{\sqrt{n}}}}{\sqrt{\left(u_c^* + \frac{y}{\sqrt{n}}\right)\left(1 - u_c^* - \frac{y}{\sqrt{n}}\right)} D_c^{(n)}}(1+o(1)), \tag{1.64}$$

we obtain:

$$e^{-tn^{1/2-3\delta}} \bar{M}_n^*(t) \le \bar{M}_n(t) \le e^{t\left(2n^{1/2-2\delta} + n^{1/2-3\delta}\right)} \bar{M}_n^*(t). \tag{1.65}$$

In (1.64) we recognize the probability density ℓ_n^c of the random variable Y_n, introduced in (1.60). From (1.64) we then deduce that $\bar{M}_n^*(t) = \mathbb{E}(e^{t\frac{4}{3}Y_n})(1+o(1))$. By Scheffé Lemma, Y_n converges in distribution to a generalized Gaussian random variable Y with Lebesgue density $\ell^c(y) \propto e^{-\frac{81}{64}y^4}$, therefore

$$\bar{M}_n^*(t) = \mathbb{E}(e^{t\frac{4}{3}Y_n})(1+o(1)) \xrightarrow{n \to +\infty} \mathbb{E}(e^{t\frac{4}{3}Y}).$$

With the further constraint $\frac{1}{4} < \delta < \frac{3}{8}$, from (1.65) it holds $\bar{M}_n(t) \xrightarrow{n \to +\infty} \mathbb{E}(e^{t\frac{4}{3}Y})$. By setting $\bar{Y} := \frac{4}{3}Y$ we conclude the proof. □

Proof of Theorem 1.9. The object that we want to study is in this case the random variable

$$\sqrt{6}\,\frac{\frac{\bar{T}_n}{n} - \frac{n^2}{6}\bar{m}_n^\Delta(\alpha, h)}{n} = \bar{V}_n + \frac{n}{\sqrt{6}}\left(u_0^{*3} - \bar{m}_n^\Delta(\alpha, h)\right), \tag{1.66}$$

where

$$\bar{V}_n := \sqrt{6}\,\frac{\frac{\bar{T}_n}{n} - \frac{n^2}{6}u_0^{*3}}{n}.$$

By (1.19) and Slutsky theorem, it is enough to study the moment generating function of $\bar{V}_n$, restricting again the analysis on the neighborhood $B_{u_0^*}$ (the contribution over the set $\mathcal{C} \equiv \mathcal{C}(n,\delta) := \Gamma_n \setminus B_{u_0^*}$ can be treated as in (1.62), with $0<\delta<1$). To simplify constants, we consider $\sqrt{6}\bar{V}_n$ instead of $\bar{V}_n$. We then get:

$$\bar{M}_n(t) = \sum_{m\in B_{u_0^*}} \frac{2}{n}\frac{1}{\sqrt{m(1-m)}}\frac{e^{tn(m^3-u_0^{*3})-c_0n^2(m-u_0^*)^2+k_0n^2(m-u_0^*)^3}}{D_0^{(n)}}(1+o(1)).$$

The change of variable $x = n(m-u_0^*)$, identity (1.54), and the Taylor expansion (1.36) yield

$$\bar{M}_n(t) = \sum_{x\in R_{0,\delta}^{(n)}} \frac{2}{n}\frac{e^{t(3u_0^{*2}x+3u_0^*\frac{x^2}{n}+\frac{x^3}{n^2})}\cdot e^{-c_0x^2+k_0\frac{x^3}{n}}}{\sqrt{\left(u_0^*+\frac{x}{n}\right)\left(1-u_0^*-\frac{x}{n}\right)}D_0^{(n)}}(1+o(1))\,.$$

As in the proof of Thm. 1.10, let

$$\bar{M}_n^{**}(t) := \sum_{x\in R_{0,\delta}^{(n)}} \frac{2}{n}\frac{e^{3tu_0^{*2}x}\cdot e^{-c_0x^2+k_0\frac{x^3}{n}}}{\sqrt{\left(u_0^*+\frac{x}{n}\right)\left(1-u_0^*-\frac{x}{n}\right)}D_0^{(n)}}(1+o(1)). \tag{1.67}$$

Furthermore, by exploiting the range of $R_{0,\delta}^{(n)}$, given in (1.39) we obtain

$$e^{-tn^{1-3\delta}}\bar{M}_n^{**}(t) \le \bar{M}_n(t) \le e^{t\left(3u_0^*n^{1-2\delta}+n^{1-3\delta}\right)}\bar{M}_n^{**}(t). \tag{1.68}$$

In (1.67) we recognize the probability density ℓ_n of the random variable X_n, introduced in (1.56); we then rewrite (1.67) as $\bar{M}_n^{**}(t) = \mathbb{E}(e^{3tu_0^{*2}X_n})(1+o(1))$. By Scheffé Lemma X_n converges in distribution to a real random variable X with Gaussian density ℓ given in (1.59), therefore

$$\bar{M}_n^{**}(t) = \mathbb{E}(e^{3tu_0^{*2}X_n})(1+o(1)) \xrightarrow{n\to+\infty} \mathbb{E}(e^{3tu_0^{*2}X}).$$

With the further constraint $\frac{1}{2}<\delta<1$, from (1.68), it holds $\bar{M}_n(t) \xrightarrow{n\to+\infty} \mathbb{E}(e^{3tu_0^{*2}X})$. Note that X is a centered Gaussian random variable with variance $(2c_0)^{-1}$, where $c_0 \equiv c_0(\alpha,h) = \frac{1-2\alpha[u_0^*(\alpha,h)]^2[1-u_0^*(\alpha,h)]}{4u_0^*(\alpha,h)[1-u_0^*(\alpha,h)]}$. In conclusion, $\bar{V}_n$ converges in distribution to the centered Gaussian random variable $3u_0^{*2}X/\sqrt{6}$, with variance $\bar{v}_0^{\Delta}(\alpha,h) = \frac{3u_0^{*4}}{4c_0}$, as wanted. □

1.3.7 Conditional measures

Proof of Proposition 1.15. Let $(\alpha,h)\in\mathcal{M}^{rs}$ and let $u_i^* = u_i^*(\alpha,h)$, $i=1,2$ the two solutions of the scalar problem (1.7). The proof of this proposition can be carried on exactly as the proof of the analog Prop. 1.11, but in the conditional setting introduced in Subsec. 1.2.2. Without loss of generality, we consider the case $i = 1$:

$$\hat{\mathbb{E}}^{(1)}_{n;\alpha,h}\left(\left|\frac{6\bar{T}_n}{n^3} - u_1^{*3}\right|\right) = \sum_{m\in B_{u_1^*}} |m^3 - u_1^{*3}| \frac{\mathcal{N}_m\, e^{n^2(\frac{\alpha}{6}m^3+\frac{h}{2}m)}}{\bar{Z}_{n;\alpha,h}(B_{u_1^*})}$$
$$= \sum_{m\in B_{u_1^*}} \frac{2}{n}\frac{|m^3 - u_1^{*3}|}{\sqrt{m(1-m)}}\frac{e^{-n^2(f_{\alpha,h}-g_{\alpha,h}(m))}}{D_1^{(n)}}(1+o(1)),$$

where $\hat{\mathbb{E}}^{(1)}_{n;\alpha,h}$ is the expectation associated with the measure $\hat{\mathbb{P}}^{(1)}_{n;\alpha,h}$ defined in (1.22), and $\bar{Z}_{n;\alpha,h}(B_{u_1^*})$ denotes the restriction of the partition function to the set $B_{u_1^*}$. The Taylor expansion (1.36) and the change of variable $x = n(m - u_1^*)$ yield

$$n\cdot\hat{\mathbb{E}}^{(1)}_{n;\alpha,h}\left(\left|\frac{6\bar{T}_n}{n^3} - u_1^{*3}\right|\right) = \sum_{x\in R^{(n)}_{1,\delta}} \frac{2}{n}\frac{|3u_1^{*2}x + 3u_1^*\frac{x^2}{n} + \frac{x^3}{n^2}|\cdot e^{-c_1x^2+\frac{k_1}{n}x^3}}{\sqrt{\left(u_1^*+\frac{x}{n}\right)\left(1-u_1^*-\frac{x}{n}\right)}\cdot D_1^{(n)}}(1+o(1)),$$

where $R^{(n)}_{1,\delta}$ is as in (1.39). If $(X_n^{(1)})_{n\geq 1}$ is a sequence of random variables with probability density

$$\ell_n^{(1)}(x) := \frac{2}{n}\frac{e^{-c_1x^2+\frac{k_1}{n}x^3}}{\sqrt{\left(u_1^*+\frac{x}{n}\right)\left(1-u_1^*-\frac{x}{n}\right)}\cdot D_1^{(n)}}\,\mathbb{1}_{R^{(n)}_{1,\delta}}(x), \quad x\in\mathbb{R}, \tag{1.69}$$

where $D_1^{(n)}$ is the normalization weight defined in (1.41), we obtain, as in (1.57)–(1.58),

$$\mathbb{E}(|3u_1^{*2}X_n^{(1)}|)(1+o(1)) - \left(\int_{\mathbb{R}}\left|3u_1^*\frac{x^2}{n} + \frac{x^3}{n^2}\right|\ell_n^{(1)}(x)dx\right)(1+o(1)) \tag{1.70}$$
$$\leq n\cdot\hat{\mathbb{E}}^{(1)}_{n;\alpha,h}\left(\left|\frac{6\bar{T}_n}{n^3} - u_1^{*3}\right|\right) \leq$$
$$\mathbb{E}(|3u_1^{*2}X_n^{(1)}|)(1+o(1)) + \left(\int_{\mathbb{R}}\left|3u_1^*\frac{x^2}{n} + \frac{x^3}{n^2}\right|\ell_n^{(1)}(x)dx\right)(1+o(1)). \tag{1.71}$$

Arguing as at the end of proof of Prop. 1.11, we conclude

$$n\cdot\hat{\mathbb{E}}^{1}_{n;\alpha,h}\left(\left|\frac{6\bar{T}_n}{n^3} - u_1^{*3}\right|\right) \xrightarrow{n\to+\infty} 3u_1^{*2}\mathbb{E}(|X^{(1)}|),$$

where $X^{(1)}$ is a standard Gaussian variable with variance $(2c_1)^{-1}$. Indeed, the second summand in (1.70)–(1.71) vanishes, being $3u_1^*\frac{x^2}{n} + \frac{x^3}{n^2} = o(1)$, for fixed x. Setting $\bar{X}^{(1)} := 3u_1^{*2}X^{(1)}$, we obtain a random variable with variance $6\bar{v}_1^{\Delta}(\alpha,h) = \frac{9u_i^*(\alpha,h)^4}{2c_1}$, as wanted. The same proof holds for the case $i = 2$. □

Proof of Corollary 1.16. The proof follows immediately, as for Cor. 1.12. □

As mentioned, the next theorem is the analog of Thm. 1.7 and Thm. 1.9, when the edge density is conditioned to take values in a neighborhood of the two maximizers of the scalar problem (1.7).

Proof of Theorem 1.14. Let $(\alpha, h) \in \mathcal{M}^{rs}$ and let $u_i^* = u_i^*(\alpha, h)$, $i = 1, 2$ the two solutions of the scalar problem (1.7). We focus on the case $i = 1$, being the case $i = 2$ completely analogous. We start proving (1.23) via exponential convergence, which again implies the a.s. convergence by a standard Borel-Cantelli argument (see Ellis [1985], Thm. II.6.4 and Rem. 1.1). Fix $\eta > 0$. We define

$$\mathcal{R} \equiv \mathcal{R}(\eta; n) := \left\{ m \in \Gamma_n : \eta/3 \leq |m - u_1^*| < n^{-\delta} \right\}.$$

Notice that for large n, the set $\mathcal{R}$ is empty. When this does not hold, we have

$$\begin{aligned}
\hat{\mathbb{P}}^{(1)}_{n;\alpha,h}\left(\left| \frac{6\bar{T}_n}{n^3} - u_1^{*3} \right| \geq \eta \right) &\leq \sum_{m \in \mathcal{R}} \frac{\mathcal{N}_m e^{n^2\left(\frac{\alpha}{6}m^3 + \frac{h}{2}m\right)}}{\bar{Z}_{n;\alpha,h}(B_{u_1^*})} \\
&\leq Cc^{-1} n^4 e^{-n^2(f_{\alpha,h} - \max_{m\in\mathcal{R}} g_{\alpha,h}(m))} \\
&\leq Cc^{-1} n^4 e^{-n^2 \min_{m\in\mathcal{R}}(f_{\alpha,h} - g_{\alpha,h}(m))}, \qquad (1.72)
\end{aligned}$$

where in the second to last passage we used the rough bound $\bar{Z}_{n;\alpha,h}(B_{u_1^*}) \geq cn^{-1}e^{n^2 f_{\alpha,h}}$ coming from Stirling approximation (1.46). As stated in [Radin and Yin, 2013, Prop. 3.2], for sufficiently large n, the function $f_{\alpha,h} - g_{\alpha,h}(m)$, restricted to the neighborhood $B_{u_1^*}$, is positive, convex and admits u_1^* as unique zero. Hence

$$\min_{m\in\mathcal{R}} (f_{\alpha,h} - g_{\alpha,h}(m)) = \min\{f_{\alpha,h} - g_{\alpha,h}(u_1^* - \eta), f_{\alpha,h} - g_{\alpha,h}(u_1^* + \eta)\} > 0.$$

When $\mathcal{R}$ is nonempty, the probability in (1.72) vanishes, as $n \to \infty$. This provides the desired exponential convergence, for every choice of $\eta > 0$. We now move to the proof of (1.24). By means of decomposition (1.66), and Cor. 1.16 we can reduce our analysis to the random variable

$$\bar{V}_n^{(1)} := \sqrt{6}\, \frac{\frac{\bar{T}_n}{n} - \frac{n^2}{6}u_1^{*3}}{n},$$

studying, for any $t \in \mathbb{R}$, its moment generating function

$$\hat{M}_n(t) := \hat{\mathbb{E}}^{(1)}_{n;\alpha,h}\left(e^{t\bar{V}_n^{(1)}} \right).$$

We consider $\sqrt{6}\bar{V}_n^{(1)}$ instead of $\bar{V}_n^{(1)}$ (to simplify constants), and we follow the same line of arguments as in the proof of Thm. 1.9. We get:

$$\hat{M}_n(t) = \sum_{m \in B_{u_1^*}} \frac{2}{n} \frac{1}{\sqrt{m(1-m)}} \frac{e^{tn(m^3 - u_1^{*3}) - c_1 n^2 (m - u_1^*)^2 + k_1 n^2 (m - u_1^*)^3}}{D_1^{(n)}} (1 + o(1)).$$

By the change of variable $x = n(m - u_1^*)$, identity (1.54), and the Taylor expansion (1.36),

$$\hat{M}_n(t) = \sum_{x \in R_{1,\delta}^{(n)}} \frac{2}{n} \frac{e^{t(3u_1^{*2}x + 3u_1^*\frac{x^2}{n} + \frac{x^3}{n^2})} \cdot e^{-c_1 x^2 + k_1 \frac{x^3}{n}}}{\sqrt{\left(u_1^* + \frac{x}{n}\right)\left(1 - u_1^* - \frac{x}{n}\right)} D_1^{(n)}} (1 + o(1)),$$

where $R_{1,\delta}^{(n)}$ is defined in (1.39). By defining

$$\hat{M}_n^{**}(t) := \sum_{x \in R_{1,\delta}^{(n)}} \frac{2}{n} \frac{e^{3tu_1^{*2}x} \cdot e^{-c_1 x^2 + k_1 \frac{x^3}{n}}}{\sqrt{\left(u_1^* + \frac{x}{n}\right)\left(1 - u_1^* - \frac{x}{n}\right)} D_1^{(n)}} (1 + o(1)),$$

we observe that $\hat{M}_n^{**}(t) = \mathbb{E}(e^{3tu_1^{*2} X_n^{(1)}})(1 + o(1))$, where, for each $n \in \mathbb{N}$, $X_n^{(1)}$ is a random variable with density $\ell_n^{(1)}(x)$ given in (1.69).
Notice that $X_n^{(1)}$ converges in distribution to a centered Gaussian random variable $X^{(1)}$ with variance $(2c_1)^{-1}$, where $c_1 \equiv c_1(\alpha, h) = \frac{1 - 2\alpha[u_1^*(\alpha,h)]^2 [1 - u_1^*(\alpha,h)]}{4u_1^*(\alpha,h)[1 - u_1^*(\alpha,h)]}$. Therefore

$$\bar{M}_n^{**}(t) = \mathbb{E}(e^{3tu_1^{*2} X_n^{(1)}})(1 + o(1)) \xrightarrow{n \to +\infty} \mathbb{E}(e^{3tu_1^{*2} X^{(1)}}).$$

With the further constraint $\frac{1}{2} < \delta < 1$, exploiting the same bounds as in (1.68), we also obtain the convergence $\hat{M}_n(t) \xrightarrow{n \to +\infty} \mathbb{E}(e^{3tu_1^{*2} X^{(1)}})$ for all $t \in \mathbb{R}$. In conclusion, $\bar{V}_n^{(1)}$ converges in distribution to the centered Gaussian random variable $3u_1^{*2} X^{(1)}/\sqrt{6}$, with variance $\bar{v}_1^{\Delta}(\alpha, h) = \frac{3u_1^{*4}}{4c_1}$, as wanted. The same proof works for $i = 2$. □

1.3.8 CLT for the integer part model

This section is dedicated to the proof of Thm. 1.18; the proof of Theorem 1.19 is omitted, as it follows exactly the same argument. To describe the fluctuations of $\lfloor T_n/n \rfloor$ around its mean value, in view of the decomposition $T_n/n = \lfloor T_n/n \rfloor + \{T_n/n\}$ combined with Slutsky's theorem (see [Klenke, 2020, Thm. 13.8]), it is enough to study the asymptotic behavior of the moment generating function of $W_n := \sqrt{6}\, \frac{\lfloor T_n/n \rfloor - \hat{\mathbb{E}}_{n;\alpha,h}(\lfloor T_n/n \rfloor)}{n}$. Specifically, we are going to relate this generating function to the second order derivative of the cumulant generating function of $\lfloor T_n/n \rfloor$, which is defined as

$$c_n(t) := 6n^{-2} \ln \hat{\mathbb{E}}_{n;\alpha,h}[\exp(t \lfloor T_n/n \rfloor)], \qquad t \in \mathbb{R}. \tag{1.73}$$

Remark 1.26. *Note that, by a direct calculation, we get*

$$c_n'(t) = \frac{6\hat{\mathbb{E}}_{n;\alpha+t,h}\left(\lfloor T_n/n \rfloor\right)}{n^2} \quad \text{and} \quad c_n''(t) = \frac{6\mathbb{V}\mathrm{ar}_{n;\alpha+t,h}(\lfloor T_n/n \rfloor)}{n^2}. \tag{1.74}$$

This comes after noticing that $c_n(t) = 6n^{-2}(\ln \hat{Z}_{n;\alpha+t,h} - \ln \hat{Z}_{n;\alpha,h})$ and from general properties of Gibbs densities.

The limit of the sequence $(c_n''(t))_{n \geq 1}$ for $t = t_n = o(1)$ will give the variance of the limiting Gaussian. The existence of this limit follows from the Yang–Lee theorem (see Thm. 1.27). To apply it, we first need a suitable representation of the partition function, which we provide as a first step. Following this, we establish some auxiliary results that will be used in the proof, which is deferred to the end of the section.

Representation of the partition function. We start from the partition function obtained by plugging (1.13) in the expression of the partition function, and then we incorporate the integer part. First, we have:

$$Z_{n;\alpha,h} = \sum_{x \in \mathcal{A}_n} e^{\frac{\alpha}{n} \sum_{\{i,j,k\} \in \mathcal{T}_n} x_i x_j x_k + h \sum_{i \in \mathcal{E}_n} x_i} .$$

Notice that there is a bijection between $\mathcal{A}_n$ and the power set $\mathcal{P}(\mathcal{E}_n)$, that maps an element $x \in \mathcal{A}_n$ to the set $S = \{i \in \mathcal{E}_n \,:\, x_i = 1\}$. We can then decompose $\mathcal{A}_n$ in disjoint subsets as

$$\mathcal{A}_n = \bigcup_{m=0}^{\binom{n}{3}} \bigcup_{\ell=0}^{\binom{n}{2}} \bigcup_{\substack{S \subseteq \mathcal{E}_n : |S| = \ell, \\ |\{\{i,j,k\} \subset S \,:\, \{i,j,k\} \in \mathcal{T}_n\}| = m}} \{x \in \mathcal{A}_n \,:\, x_i = 1 \Leftrightarrow i \in S\},$$

and write

$$Z_{n;\alpha,h} = \sum_{m=0}^{\binom{n}{3}} e^{\alpha \frac{m}{n}} \sum_{\ell=0}^{\binom{n}{2}} G^{(n)}_{m,\ell} e^{h\ell},$$

where $G^{(n)}_{m,\ell} := |\{S \subseteq \mathcal{E}_n : |S| = \ell, |\{\{i,j,k\} \subset S \,:\, \{i,j,k\} \in \mathcal{T}_n\}| = m\}|$. Setting $z := e^{\alpha}$ and $K^{(n)}_{m,h} := \sum_{\ell=0}^{\binom{n}{2}} G^{(n)}_{m,\ell} e^{h\ell}$, we obtain

$$Z_{n;\alpha,h}(z) = \sum_{m=0}^{\binom{n}{3}} K^{(n)}_{m,h} z^{\frac{m}{n}} ,$$

which is not a polynomial since $\frac{m}{n}$ is not necessarily an integer. For example, when $n = 3$ we have $G^{(3)}_{0,0} = G^{(3)}_{1,3} = 1$ and $G^{(3)}_{0,1} = G^{(3)}_{0,2} = 3$, yielding $Z_{3;\alpha,h} = (1 + 3e^h + 3e^{2h}) + z^{1/3} e^{3h}$. Instead, by taking the integer part, we obtain the following polynomial representation:

$$\hat{Z}_{n;\alpha,h} \equiv \hat{Z}_{\bar{n}}(z) := \sum_{k=0}^{\bar{n}} \widetilde{K}^{(n)}_{k,h} z^k , \tag{1.75}$$

where $\bar{n} := \lfloor \frac{(n-1)(n-2)}{6} \rfloor$, and $\widetilde{K}^{(n)}_{k,h} := \sum_{m : \lfloor \frac{m}{n} \rfloor = k} K^{(n)}_{m,h}$. Note that (1.75) can be equivalently written as

$$\hat{Z}_{\bar{n}}(z) = \widetilde{K}^{(n)}_{\bar{n},h} \sum_{k=0}^{\bar{n}} \frac{\widetilde{K}^{(n)}_{k,h}}{\widetilde{K}^{(n)}_{\bar{n},h}} z^k .$$

Let $\hat{Z}'_{\bar{n}}(z) := \sum_{k=0}^{\bar{n}} \frac{\widetilde{K}^{(n)}_{k,h}}{\widetilde{K}^{(n)}_{\bar{n},h}} z^k$. If $z_1, \ldots, z_{\bar{n}}$ are the complex roots of the polynomial $\hat{Z}'_{\bar{n}}(z)$, then we can write

$$\hat{Z}'_{\bar{n}}(z) = \prod_{j=1}^{\bar{n}} (z - z_j) = \prod_{j=1}^{\bar{n}} z_j \cdot \prod_{j=1}^{\bar{n}} \left(\frac{z}{z_j} - 1 \right)$$

and, since $\prod_{j=1}^{\bar{n}} z_j = (-1)^{\bar{n}} \frac{\widetilde{K}_{0,h}^{(n)}}{\widetilde{K}_{\bar{n},h}^{(n)}}$, we get

$$\hat{Z}'_{\bar{n}}(z) = (-1)^{\bar{n}}(-1)^{\bar{n}} \frac{\widetilde{K}_{0,h}^{(n)}}{\widetilde{K}_{\bar{n},h}^{(n)}} \prod_{j=1}^{\bar{n}} \left(1 - \frac{z}{z_j}\right) = \frac{\widetilde{K}_{0,h}^{(n)}}{\widetilde{K}_{\bar{n},h}^{(n)}} \prod_{j=1}^{\bar{n}} \left(1 - \frac{z}{z_j}\right).$$

Therefore we obtain

$$\hat{Z}_{\bar{n}}(z) = \widetilde{K}_{0,h}^{(n)} \prod_{j=1}^{\bar{n}} \left(1 - \frac{z}{z_j}\right). \tag{1.76}$$

Uniform convergence of derivatives. The following theorem can be now applied to this polynomial representation.

Theorem 1.27 (Lee and Yang [1952], Thm. 2)**.** *Let $Z_n(z) = \prod_{j=1}^{\bar{n}} \tilde{K}^{(n)} \left(1 - \frac{z}{z_j}\right)$ a the polynomial representation of a partition function as above. If there exists a region $R \subseteq \mathbb{C}$ containing a segment of the real positive axis that is always root-free then, as $n \to +\infty$ and for $z \in R$, all quantities*

$$\frac{1}{n} \ln Z_n(z), \qquad \frac{d^k}{d(\ln z)^k} \frac{1}{n} \ln Z_n(z), \quad \textit{with } k \in \mathbb{N}, \tag{1.77}$$

converge to analytic limits with respect to z. In particular, the limit and derivative operations switch in the whole region R.

Remark 1.28. *Recall that $f_{\alpha,h} = \hat{f}_{\alpha,h}$. With a slight abuse of notation, we might also denote the limiting free energy by the function $\alpha \mapsto f_\alpha^{(h)} := \lim_{n\to\infty} \frac{1}{n^2} \ln \hat{Z}_{\bar{n}}(e^\alpha)$. Since $f_\alpha^{(h)}$ is real analytic for all h such that $(\alpha, h) \in \mathcal{U}_{\alpha,h}^{rs} \setminus \{(\alpha_c, h_c)\}$ (see Radin and Yin [2013], Thms. 2.1 and 3.9), we claim that in this parameter regime the partition function* (1.76) *verifies the assumption of Theorem 1.27. Let us set $z_0 := e^{\alpha^*} \in \mathbb{R}^+$, for some α^* in the analyticity region. For finite n, z_0 can not be a zero of $\hat{Z}_{\bar{n}}(z)$. Indeed, since the polynomial* (1.76) *has strictly positive coefficients, for each fixed n it can only have non-real complex roots, which occur in conjugate pairs. As n grows, a phase transition in the system is usually associated to the presence of a real positive accumulation point of zeros of $\hat{Z}_{\bar{n}}(z)$ (see [Lee and Yang, 1952, Sec. IV, Item (2)] and [Bena et al., 2005, p. 4276]). As the phase diagram of the free energy has been completely characterized in Yin [2013], we know that, as n grows, no phase transition appears in the parameter regime under consideration. Therefore we claim, that in the limit $n \to \infty$ there exists a region R, containing the point z_0, which is always root-free, and the partition function* (1.76) *fulfills the assumption of Theorem 1.27.*

Corollary 1.29. *Let $(\alpha, h) \in \mathcal{U}_{\alpha,h}^{rs} \setminus \{(\alpha_c, h_c)\}$. Then,*

$$\lim_{n\to+\infty} \frac{6}{n^2} \partial_\alpha \hat{f}_{n;\alpha,h} = u_{\alpha,h}^{*3} \quad \textit{and} \quad \lim_{n\to+\infty} \frac{6}{n^2} \partial_{\alpha\alpha} \hat{f}_{n;\alpha,h} = 3u_{\alpha,h}^{*2} \partial_\alpha u_{\alpha,h}^* \tag{1.78}$$

Proof. The result is an immediate application of Rmk. 1.28 and Thm. 1.27, which holds true since we are working in the region $\mathcal{U}^{rs}_{\alpha,h} \setminus \{(\alpha_c, h_c)\}$, where the limiting free energy exists and is analytic. We observe that, since in the polynomial representation (1.75) we have $z = e^{\alpha}$, then $\frac{d}{d(\ln z)} \frac{1}{n^2} \ln \hat{Z}_{\bar{n}}(z) = \partial_\alpha \hat{f}_{n;\alpha,h}$ and $\frac{d^2}{d(\ln z)^2} \frac{1}{n^2} \ln \hat{Z}_{\bar{n}}(z) = \partial_{\alpha\alpha} \hat{f}_{n;\alpha,h}$. Therefore, Thm. 1.27 allows to commute limit and derivative to get

$$\lim_{n\to+\infty} \frac{6}{n^2} \partial_\alpha \hat{f}_{n;\alpha,h} = 6 \lim_{n\to+\infty} \partial_\alpha \hat{f}_{n;\alpha,h} = 6\partial_\alpha \left[\lim_{n\to+\infty} \hat{f}_{n;\alpha,h} \right] = 6\partial_\alpha f_{\alpha,h} = u^{*3}_{\alpha,h}.$$

where the last equality follows by directly differentiating (1.7) with respect to α, recalling that $u^*_{\alpha,h}$ satisfies (1.8). The second limit on the r.h.s. of (1.78) can be proved in the same way. □

Theorem 1.27 implies that derivatives of the free energy converge locally uniformly.

Proposition 1.30 ([Bianchi et al., 2024, Prop. 4.2])**.** *Under the hypotheses of Thm. 1.27, the quantities displayed in (1.77) converge locally uniformly (in n) inside the region R.*

Remark 1.31. *Recalling (1.74) and the definition (1.16) of $\hat{f}_{n;\alpha,h}$, a direct computation shows that*

$$c_n'(0) = \frac{6}{n^2} \partial_\alpha \hat{f}_{n;\alpha,h} \qquad \textit{and} \qquad c_n''(0) = \frac{6}{n^2} \partial_{\alpha\alpha} \hat{f}_{n;\alpha,h}.$$

Therefore, from (1.78),

$$\lim_{n\to\infty} c_n'(0) = u^{*3}_{\alpha,h} \qquad \textit{and} \qquad \lim_{n\to\infty} c_n''(0) = 3u^{*2}_{\alpha,h} \partial_\alpha u^*_{\alpha,h} = \hat{v}^{\Delta}_0(\alpha, h). \tag{1.79}$$

Proof of the CLT. The proof of our main theorem is then just one step further. We will rely on the analyticity of the free energy and on the uniform convergence of the sequence $(c_n''(t))_{n\geq 1}$ guaranteed by Thm. 1.27 and Prop. 1.30.

Proof of Thm. 1.18. Recall $\hat{v}^{\Delta}_0(\alpha, h) = 3u^{*2}_{\alpha,h} \partial_\alpha u^*_{\alpha,h}$ and $W_n = \sqrt{6}\, \frac{\lfloor T_n/n \rfloor - \hat{\mathbb{E}}_{n;\alpha,h}(\lfloor T_n/n \rfloor)}{n}$. We want to show that

$$\lim_{n\to+\infty} \hat{\mathbb{E}}_{n;\alpha,h}(\exp(tW_n)) = \exp\left(\tfrac{1}{2} \hat{v}^{\Delta}_0(\alpha, h) t^2\right)$$

for all $t \in [0, \eta)$ and some $\eta > 0$. We aim to express $\hat{\mathbb{E}}_{n;\alpha,h}(\exp(tW_n))$ in terms of $c_n''(t)$. Consider $t > 0$ and set $t_n := \sqrt{6}t/n$. We get

$$\begin{aligned} \ln \hat{\mathbb{E}}_{n;\alpha,h}(\exp(tW_n)) &= \ln \hat{\mathbb{E}}_{n;\alpha,h} \left(\exp(t_n \lfloor T_n/n \rfloor) \exp\left(-t_n \hat{\mathbb{E}}_{n;\alpha,h}(\lfloor T_n/n \rfloor)\right)\right) \\ &\overset{(1.73),(1.74)}{=} \frac{n^2}{6} [c_n(t_n) - t_n c_n'(0)]. \end{aligned}$$

Notice that, since $c_n(0) = 0$, the term in square brackets is the difference between the function $c_n(t_n)$ and its first order Taylor expansion at zero. Therefore, by using Taylor's theorem with Lagrange remainder, one gets

$$\ln \hat{\mathbb{E}}_{n;\alpha,h}(\exp(tW_n)) = \frac{c_n''(t_n^*) t^2}{2},$$

for some $t_n^* \in [0, \sqrt{6}t/n]$. To conclude the proof of the central limit theorem, we need to control the limiting behavior of $c_n''(t_n^*)$. To this end, we recall from (1.79) that $\lim_{n\to\infty} c_n''(0) = v(\alpha, h)$, and that, by Prop. 1.30, the derivatives of $c_n(t)$ converge locally uniformly. These two properties together yield the following result, which was first proved in a slightly different setting but applies unchanged in the present context.

Lemma 1.32 ([Bianchi et al., 2024, Lem. 6.1])**.** *For $(\alpha, h) \in \mathcal{U}^{rs}_{\alpha,h} \setminus \{(\alpha_c, h_c)\}$, there exists $\eta > 0$ such that we have $\lim_{n\to+\infty} c_n''(t_n) = \hat{v}_0^{\Delta}(\alpha, h)$ for all $t_n \in [0, \eta)$ with $\lim_{n\to+\infty} t_n = 0$.*

From the lemma above, we obtain the convergence of $c_n''(t_n^*)$, and, in turn, the convergence of the moment generating function. Therefore $W_n \xrightarrow{d} \mathcal{N}(0, \hat{v}_0^{\Delta}(\alpha, h))$ (see Billingsley [1986], Sect. 30). Finally, the convergence in distribution of

$$\sqrt{6}\,\frac{{T_n}/{n} - \hat{\mathbb{E}}_{n;\alpha,h}({T_n}/{n})}{n} = W_n + \sqrt{6}\,\frac{\{{T_n}/{n}\} - \hat{\mathbb{E}}_{n;\alpha,h}(\{{T_n}/{n}\})}{n}$$

follows from Slutsky's theorem, as $\sqrt{6}\,\frac{\{{T_n}/{n}\} - \hat{\mathbb{E}}_{n;\alpha,h}(\{{T_n}/{n}\})}{n} \to 0$ in probability, being the numerator bounded almost surely. □

Remark 1.33. *To extend Thm. 1.18 to the measure $\mathbb{P}_{n;\alpha,h}$ it remains to show*

$$\lim_{n\to+\infty} \hat{\mathbb{E}}_{n;\alpha,h}(\exp(tW_n)) = \lim_{n\to+\infty} \mathbb{E}_{n;\alpha,h}(\exp(tW_n)), \tag{1.80}$$

the latter expectation being associated with $\mathbb{P}_{n;\alpha,h}$. A natural approach is to compare the two expectations directly. In particular, with a direct computation one can show that

$$\mathbb{E}_{n;\alpha,h}(e^{tW_n}) = \frac{\hat{\mathbb{E}}_{n;\alpha,h}(e^{tW_n + \alpha\{\frac{T_n}{n}\}})}{\hat{\mathbb{E}}_{n;\alpha,h}(e^{\alpha\{\frac{T_n}{n}\}})},$$

and the same identity holds if we interchange the role of $\mathbb{E}_{n;\alpha,h}$ and $\hat{\mathbb{E}}_{n;\alpha,h}$. Therefore

$$\left|\mathbb{E}_{n;\alpha,h}\left(e^{tW_n}\right) - \hat{\mathbb{E}}_{n;\alpha,h}\left(e^{tW_n}\right)\right| = \frac{\left|\hat{\mathbb{E}}_{n;\alpha,h}\left(e^{tW_n + \alpha\{\frac{T_n}{n}\}}\right) - \hat{\mathbb{E}}_{n;\alpha,h}\left(e^{tW_n}\right)\hat{\mathbb{E}}_{n;\alpha,h}(e^{\alpha\{\frac{T_n}{n}\}})\right|}{\hat{\mathbb{E}}_{n;\alpha,h}(e^{\alpha\{\frac{T_n}{n}\}})}.$$

Proving that the variables e^{tW_n} and $e^{\alpha\{\frac{T_n}{n}\}}$ are asymptotically independent would be then sufficient to prove (1.80).

Remark 1.34. *In order to make our approximation more precise, we can introduce a parameter $r \in \mathbb{N}$ and consider the family of Hamiltonians*

$$\hat{\mathcal{H}}^{(r)}_{n;\alpha,h}(x) := \frac{\alpha}{r}\left\lfloor \frac{\sum_{\{i,j,k\}\in\mathcal{T}_n} x_i x_j x_k}{n} \cdot r \right\rfloor + h\sum_{i\in\mathcal{E}_n} x_i, \qquad r \in \mathbb{N}.$$

It is not difficult to observe that for each $r \in \mathbb{N}$ the related partition function can be represented as polynomial in the variable $z = e^{\frac{\alpha}{r}}$ and, as a consequence, noted with

$\hat{\mathbb{E}}^{(r)}_{n;\alpha,h}$ *the associated expectation, it is possible to prove a CLT which is analogous to Theorem 1.18 and valid for the random variable*

$$W_n^{(r)} = \frac{\sqrt{6}}{r}\frac{\lfloor T_n/n \cdot r\rfloor - \hat{\mathbb{E}}^{(r)}_{n;\alpha,h}(\lfloor T_n/n \cdot r\rfloor)}{n}.$$

Since $-\frac{1}{r} \le \frac{1}{r}\lfloor tr\rfloor - t \le \frac{1}{r}$ *for every* $t \in \mathbb{R}$*, this extension leads to*

$$e^{-\frac{|\alpha|}{r}} \le \frac{\hat{\mathbb{E}}^{(r)}_{n;\alpha,h}(\exp(tW_n^{(r)}))}{\hat{\mathbb{E}}_{n;\alpha,h}(\exp(tW_n^{(r)}))} \le e^{\frac{|\alpha|}{r}}.$$

As r *grows, the interval* $(e^{-\frac{|\alpha|}{r}}, e^{\frac{|\alpha|}{r}})$ *shrinks and the claim (1.80) can be made asymptotically exact up to a subsequence. This is however beyond our scopes.*

1.4 Heuristics on the mean-field model

For what concerns small deviations, the comparison between the mean-field approximation and the edge-triangle model respectively encoded by Hamiltonian (1.13) and (1.14) remains an open problem. We refer the reader to [Bianchi et al., 2024, Sec. 8.3] for a discussion on the main difficulties in proving that they asymptotically coincide. In particular we mention that, as observed there, in order to get a control on the distance of moment generating functions for centered and rescaled densities, we are required a bound on

$$\begin{aligned}
n^2 t_n\big[m_n^{\Delta}(\alpha + t_n^*, h) - \bar{m}_n^{\Delta}(\alpha + \bar{t}_n^*, h)\big]& \\
= n^2 t_n \Big[(m_n^{\Delta}(\alpha + t_n^*, h) - m_n^{\Delta}(\alpha, h)) + (m_n^{\Delta}(\alpha, h) - u^{*3}(\alpha, h))\Big]& \\
- n^2 t_n \Big[(\bar{m}_n^{\Delta}(\alpha + \bar{t}_n^*, h) - \bar{m}_n^{\Delta}(\alpha, h)) + (\bar{m}_n^{\Delta}(\alpha, h) - u^{*3}(\alpha, h))\Big]&,
\end{aligned}$$

where $t \in \mathbb{R}$, $t_n^*, \bar{t}_n^* \in (0, 6t/n^{3/2})$ at the critical point and $t_n^*, \bar{t}_n^* \in (0, \sqrt{6}t/n)$ elsewhere in $\mathcal{U}^{rs}$. For $(\alpha, h) \in \mathcal{U}^{rs} \setminus \{(\alpha_c, h_c)\}$ we do not have a control on the second term of the second line, while at criticality, even the first term of the second and third line explode, due to the second order transition. However, we conjecture that Thms. 1.8, 1.9, 1.10, proved for the mean-field model, hold true when replacing $\bar{T}_n$ by T_n. This belief stems both from heuristic computations based on the large deviation principle that we have at hand (see Rem. 1.22), combined with simulations, as well as from the fluctuations of the edge density studied in Bianchi et al. [2024].

We quickly sketch our heuristic argument for the CLT. To guarantee convexity of the rate function $\mathcal{I}_{\alpha,h}$, we restrict here to the region $(\alpha, h) \in (-2, \alpha_c] \times \mathbb{R}$ (see Radin and Yin [2013], Prop. 3.2). First, we obtain the following decomposition:

$$\sqrt{6}\,\frac{\frac{T_n}{n} - \frac{n^2}{6} m_n^{\Delta}(\alpha, h)}{n} = V_n + n(u_0^{*3}(\alpha, h) - m_n^{\Delta}(\alpha, h)), \quad \text{if } (\alpha, h) \neq (\alpha_c, h_c), \tag{1.81}$$

$$6\,\frac{\frac{T_n}{n} - \frac{n^2}{6} m_n^{\Delta}(\alpha_c, h_c)}{n^{3/2}} = U_n + \sqrt{n}(u^{*3}(\alpha_c, h_c) - m_n^{\Delta}(\alpha_c, h_c)) \tag{1.82}$$

where $V_n := \frac{n}{\sqrt{6}}\left[\frac{6T_n}{n^3} - u_0^{*3}(\alpha,h)\right]$, $U_n := \sqrt{n}\left[\frac{6T_n}{n^3} - u^{*3}(\alpha_c,h_c)\right]$ and m_n^{Δ} is defined as in Def. 1.3, but for the edge-triangle model. At the moment, we don't have control over the shift terms in (1.81)–(1.82), as we don't have the equivalent of Cor. 1.12, which is valid instead for the mean-field approximation. For the remaining terms, U_n and V_n, we claim, as $n \to \infty$

$$\mathbb{P}_{n;\alpha,h}\left(V_n \in dx\right) = \mathbb{P}_{n;\alpha,h}\left(\frac{6T_n}{n^3} \in u^{*3} + \frac{\sqrt{6}}{n}dx\right) \approx e^{-n^2 I_{\alpha,h}\left(\sqrt[3]{u^{*3}+\frac{\sqrt{6}}{n}x}\right)}dx$$
$$= e^{-\frac{I''_{\alpha,h}(u^*)x^2}{3u^{*4}}+o(x^2)}dx, \tag{1.83}$$
$$\mathbb{P}_{n;\alpha_c,h_c}\left(U_n \in dx\right) = \mathbb{P}_{n;\alpha_c,h_c}\left(\frac{6T_n}{n^3} \in u^{*3} + \frac{dx}{\sqrt{n}}\right) \approx e^{-n^2 I_{\alpha_c,h_c}\left(\sqrt[3]{u^{*3}+\frac{x}{\sqrt{n}}}\right)}dx$$
$$= e^{-\frac{3^8}{2^{14}}x^4+o(x^4)}dx, \tag{1.84}$$

where $I_{\alpha,h}(x) = \frac{1}{2}[x\ln x + (1-x)\ln(1-x)] - \frac{\alpha}{6}x^3 - \frac{h}{2}x + f_{\alpha,h}$. In the second to last passage in displays (1.83)-(1.84) we have used the LDP of Rem. 1.22, which reduces to a scalar problem whenever we work in replica symmetric regime. The rate function $\mathcal{I}_{\alpha,h}$ coincides with $I_{\alpha,h}$, which is continuous, positive, strictly convex (see Radin and Yin [2013], Prop. 3.2) and admits a unique zero at $x = u^*$. In the last passage in (1.83)-(1.84), we used

$$I_{\alpha,h}(u^* \pm \delta) = \begin{cases} \frac{I''_{\alpha,h}(u^*)}{2}\delta^2 + o(\delta^2), & \text{if } (\alpha,h) \in (-2,\alpha_c)\times\mathbb{R} \\ \frac{81}{64}\,\delta^4 + o(\delta^4), & \text{if } (\alpha,h) = (\alpha_c,h_c) = \left(\frac{27}{8}, \ln 2 - \frac{3}{2}\right) \end{cases},$$

recalling that $u^*(\alpha_c,h_c) = \frac{2}{3}$. A direct computation shows that $I''_{\alpha,h}(u^*) = 2c_0$ (where c_0 is given in (1.37)), therefore in (1.83) we recognize the density of a Gaussian random variable with variance $\frac{3u^{*4}}{2I''_{\alpha,h}(u^*)}$, as stated in Thm. 1.9 (Eq. (1.17)). Similarly, in (1.84) we can immediately recognize the same density $\bar{\ell}^c$ stated in Thm. 1.10. Importantly, the error terms appearing in (1.83)–(1.84) might be relevant, although we believe they could compensate the contribution of the shift terms in (1.81)–(1.82), thus producing the conjectured results. To support this belief, in the next section we present two simulations that show the asymptotic Gaussian fluctuations outside the critical curve.

1.4.1 Simulations

We perform a discrete-time Glauber dynamics, namely an ergodic reversible Markov chain on $\mathcal{A}_n$ with stationary distribution $\mathbb{P}_{n;\alpha,h}$. A step of the Glauber dynamics can be described as follows:

1. Uniformly sample $\ell \in \mathcal{E}_n$, and let x^+ (resp. x^-) be the adjacency matrix, with $x_\ell^+ = 1$ (resp. $x_\ell^- = 0$), that coincides with x for all elements except for x_ℓ. Let $\mathcal{W}_\ell := \{\{i,j\} : i,j \in \mathcal{E}_n, i \sim j, \{i,j,\ell\} \in \mathcal{T}_n \Leftrightarrow x_\ell = 1\}$ the set of two-stars insisting on the edge (or non-edge) x_ℓ. Here, the symbol $\sim$ denotes that the two edges i and j are neighbors.

2. Given the current state represented by $x \in \mathcal{A}_n$, the next state is obtained by performing the transition $x \mapsto x^+$ (resp. $x \mapsto x^-$) with probability

$$p_n(x,\ell) := \frac{e^{\alpha \sum_{\{i,j\}\in\mathcal{W}_\ell} x_i x_j + h}}{1 + e^{\alpha \sum_{\{i,j\}\in\mathcal{W}_\ell} x_i x_j + h}} \quad (\text{resp. } 1 - p_n(x,\ell)\,)\,. \tag{1.85}$$

The update probability (1.85) is given in [Bhamidi et al., 2011, Lem. 3] (or equivalently Bhamidi et al. [2015], pag. 18). Moreover in [Bhamidi et al., 2011, Thm. 5] it has been proved that the mixing time of the Glauber dynamics is $\Theta(n^2 \ln(n))$ whenever $(\alpha, h) \in \mathcal{U}^{rs}$. Fig. 1.3 shows a numerical simulation of the probability distribution of

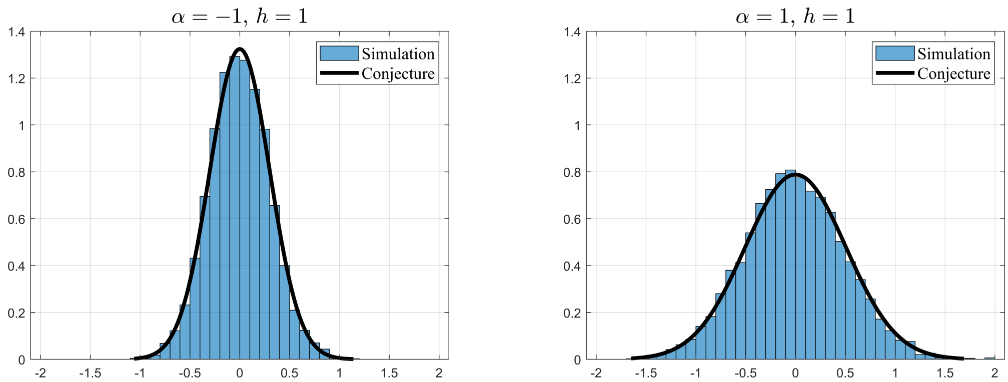

Figure 1.3: The picture displays a simulation of the distribution of $\sqrt{6}\,\frac{\frac{T_n}{n} - \frac{n^2}{6} m_n^{\Delta}(\alpha,h)}{n}$ obtained with $n = 150$, $M = 5000$ samples, and parameters $h = 1$, $\alpha = \pm 1$ (histogram), and the pdf of the Gaussian distribution introduced in Thm. 1.9 (continuous line).

$\sqrt{6}\,\frac{\frac{T_n}{n} - \frac{n^2}{6} m_n^{\Delta}(\alpha,h)}{n}$ obtained with $n = 150$ and $M = 5000$ samples, both for negative and positive values of α. The picture also displays the Gaussian probability density function given in Thm. 1.9, showing that it approximates the histogram with good accuracy.

Remark 1.35. *Note that, despite [Bhamidi et al., 2011, Thm. 5] holds in $\mathcal{U}^{rs}$, which includes (α_c, h_c), when we perform the Glauber dynamics at the critical point, the mixing time that we observe is not $\Theta(n^2 \ln(n))$, as we would expect. We believe that the proximity of the point to the critical curve $\mathcal{M}^{rs}$ where the mixing time is exponential (see [Bhamidi et al., 2011, Thm. 6]), is responsible for this behavior. As a consequence, the incredibly high computational cost prevented us from getting an equivalent simulation for supporting the non-standard CLT (Thm. 1.10).*

Chapter 2

Spectral properties of the directed Chung–Lu graph

In this chapter, we investigate the spectral properties of directed inhomogeneous random graphs. In the Chung–Lu-type setting, each vertex x is assigned a positive weight w_x^+ for its out-degree tendency and w_x^- for its in-degree tendency, and an edge from x to y is present with probability $p_{x,y}$ proportional to $w_x^+ w_y^-$ (independently for all directed pairs). The picture can be generalized to a model with higher but finite rank r. In this setting, connection probabilities are given in additive form as the sum of rank-one matrices, corresponding to eigenvectors of $\mathbb{E}[A_n]$.

Our study is motivated by the desire to bridge the gap between classical random matrix theory —which often assumes identically distributed entries—and network models that exhibit degree heterogeneity. We aim to understand how outlier phenomena extend to this inhomogeneous, non-symmetric setting.

2.1 Setup and results

2.1.1 Notation

For a given $n \times n$ complex-valued matrix M_n, we denote by $(\lambda_i(M_n))_{1\le i\le n}$ the sequence of (complex) eigenvalues of M_n, ordered so that $|\lambda_1(M_n)| \ge |\lambda_2(M_n)| \ge \cdots \ge |\lambda_n(M_n)| \ge 0$, and by $(\sigma_i(M_n))_{1\le i\le n}$ its (real) singular values, defined, for all i, by

$$\sigma_i(M_n) := \sqrt{\lambda_i(M_n M_n^*)},$$

where $M_n^* = \bar{M}_n^T$ is the adjoint of M_n. It holds $\sigma_1(M_n) \ge \sigma_2(M_n) \ge \cdots \ge \sigma_n(M_n) \ge 0$. Its spectral norm is $\|M_n\| := \sigma_1(M_n)$. If M_n is self–adjoint, then $\|M_n\| = |\lambda_1(M_n)|$.

Through the chapter, we will write

$$\mu_n \xrightarrow[n\to+\infty]{w} \mu$$

to say that a sequence of probability distributions converges weakly to some limit μ. Besides the usual Landau notation, we will also write that for two positive sequences a_n and b_n it holds $a_n \ll b_n$ (resp. $a_n \lesssim b_n$, $a_n \asymp b_n$, and $a_n \sim b_n$) if $\lim_{n\to+\infty} a_n/b_n = \ell$ and $\ell = 0$ (resp. $0 \le \ell < +\infty$, $0 < \ell < +\infty$, and $\ell = 1$). In addition, for two random variables X_n and Y_n, we will write $X_n = O_{v.h.\mathbb{P}}(Y_n)$ (the subscript will stand for *very high* probability) if there exist $K > 0$ and $\eta > 1$ such that,

$$\mathbb{P}(|X_n/Y_n| \ge K) \le e^{-(\log(n))^\eta},$$

and we will write $X_n = o_{v.h.\mathbb{P}}(Y_n)$ if, for every $\delta > 0$, there exists $\eta > 1$ such that,

$$\mathbb{P}(|X_n/Y_n| \ge \delta) \le e^{-(\log(n))^\eta}.$$

Finally, we will say that an event holds with very high probability (w.v.h.p.) if there exists $\eta > 1$ such that the probability of its complement decays as in the previous display. This strengthens the notation $O_{\mathbb{P}}(\cdot)$ and $o_{\mathbb{P}}(\cdot)$, which simply means that the probabilities on the left vanish.

2.1.2 Models

For any $n \in \mathbb{N}$, we consider the random directed graph G_n on the vertices $[n] = \{1, 2, \dots, n\}$, whose adjacency matrix A_n, has independent entries with distribution

$$A_{xy} \sim \mathrm{Ber}(p_{x,y}), \qquad x, y \in [n].$$

We consider two choices for the connection probabilities $p_{x,y}$, each defined in terms of a common scaling factor s_n. In both cases, we denote by $\mathbb{P}$ the probability measure associated with the graph.

Chung–Lu digraph. Our first model corresponds to the choice

$$p_{x,y} = \frac{w_x^+ w_y^-}{\boldsymbol{w}} s_n \wedge 1, \qquad x, y \in [n], \tag{2.1}$$

where $(w_x^+, w_x^-)_{x\in[n]}$ denote a family of non-negative bi-weights such that

$$\sum_{x\in[n]} w_x^+ = \sum_{x\in[n]} w_x^- = \boldsymbol{w}.$$

In our setting for large n it will be $p_{x,y} < 1$, so that the truncation at level 1 becomes unnecessary. We work under the following assumption, regarding the joint empirical distribution of the weights. An assumption on s_n will be given later.

Assumption 2.1. *The weights are bounded and there exists a compactly supported distribution $\rho = \rho^{+,-}$ such that*

$$\frac{1}{n} \sum_{x\in[n]} \delta_{\sqrt{\frac{n}{\boldsymbol{w}}}(w_x^+, w_x^-)} \xrightarrow[n\to+\infty]{w} \rho.$$

A higher-rank model. Our second model is more general. Let $v_1^{\pm}, \dots, v_r^{\pm} \in \mathbb{R}^n$ be a family of $2r$ bi-orthogonal vectors, i.e., such that $(v_i^+)^T v_j^- = \delta_{ij}$ for $i, j = 1, \dots, r$. We can assume without loss of generality $\|v_i^+\| = 1$ for each $i = 1, \dots, r$. Let $\theta_r > \theta_{r-1} > \cdots > \theta_1$ be positive constants, and set

$$p_{x,y} = s_n \sum_{j=1}^{r} \theta_l v_j^+(x)(v_j^-(y))^T \wedge 1, \qquad x, y \in [n], \tag{2.2}$$

where, for $x \in [n]$, $v_i^{\pm}(x)$ denotes the x-th entry of $v_i^{\pm}$. The rank-one model can be recovered by taking $r = 1$, $\theta_1 = 1$, and $v_1^{\pm}(x) = w_x^{\pm}/\sqrt{\boldsymbol{w}}$ for $x \in [n]$. Assumption 2.1 is accordingly generalized.

Assumption 2.2. *The entries of $\sqrt{n}\, v_i^{\pm}$ are uniformly bounded and there exists a compactly supported distribution $\rho = \rho_{1,\dots,r}^{+,-}$ on $\mathbb{R}_+^{2r}$ such that*

$$\frac{1}{n} \sum_{x \in [n]} \delta_{(\sqrt{n} v_1^{\pm}(x), \dots, \sqrt{n} v_r^{\pm}(x))} \xrightarrow[n \to +\infty]{w} \rho.$$

The above generalization is quite natural: every diagonalizable matrix can be decomposed in the additive form (2.2), where the $(v_i^{\pm})$ correspond to left and right eigenvectors.

Example 2.3 (Stochastic block model)**.** *Let n be even and let $a > b > 0$. If*

$$p_{x,y} = \begin{cases} a s_n/n & \text{if } \max\{x \vee y, n - x \wedge n - y\} \leq \frac{n}{2} \\ b s_n/n & \text{otherwise,} \end{cases}$$

the obtained graph falls in our hypotheses. The expected adjacency matrix has eigenvalues $\theta_1 = \frac{a-b}{2} s_n$ and $\theta_2 = \frac{b+a}{2} s_n$. The eigenvector corresponding to θ_1 contains information on the community structure, and the asymptotic behavior of its random realization can be studied as in [Chakrabarty et al., 2020, Theorems 2.4-2.5].

Any inhomogeneous graph whose expected adjacency matrix is symmetric with rank r, with eigenvalues $\theta_r s_n > \cdots > \theta_1 s_n$, falls within this framework. In this setting there exists an orthonormal basis of (unit) eigenvectors $(v_l)_{l \leq r}$ (the vectors $(v_l^+)_{l \leq r}$ and $(v_l^-)_{l \leq r}$ coincide). Of course, symmetry is not a necessary condition. To construct a non-symmetric example, if $2r < n$, we can take r orthogonal 2-dimensional subspaces $V_1, \dots, V_r$ of $\mathbb{R}^n$ and choose $v_l^{\pm} \in V_l$ on the sphere of radius 2 in such a way that $(v_l^+)^T v_l^- = 1$.

2.1.3 Main results

For both models, we will work under the following assumption, which implies that average degrees diverge sufficiently fast.

Assumption 2.4. *There exists $\xi > 4$ such that $s_n \gg \log^{\xi}(n)$.*

Under this assumption, the graph results to be w.h.p. strongly connected. Indeed, the threshold for this property is attained when the connection probability is $O(\log(n)/n)$ (see Cooper and Frieze [2012]). We are ready to state the results.

Theorem 2.5 (Existence of outlier - rank-one case)**.** *Consider the Chung–Lu model with $p_{x,y}$ as in Eq. (2.1). If Assumption 2.4 holds, then*

$$\max\{|\lambda_1(A_n) - \lambda_1(\mathbb{E}[A_n])|, |\lambda_2(A_n)|\} = O_{v.h.\mathbb{P}}(\sqrt{s_n}). \tag{2.3}$$

In particular, there exists a constant $K_0 > 0$ such that w.v.h.p.

$$K_0 s_n \le \lambda_1(A_n) \le 2K_0 s_n.$$

We again stress that the notation in (2.3) means that there exist $K > 0$ and $\eta > 1$ such that

$$\mathbb{P}\left(\max\{|\lambda_1(A_n) - \lambda_1(\mathbb{E}[A_n])|, |\lambda_2(A_n)|\} \ge K\sqrt{s_n}\right) \le e^{-(\log(n))^\eta}.$$

A similar theorem can be stated for the transition matrix of the simple random walk. As observed in Lemma 2.21, degrees are uniformly positive with high probability, so that the diagonal matrix D_n with entries $D_{xx} = D_x^+$, for $x \in [n]$, where D_x^+ is the out-degree of x, is with high probability invertible. If this is not the case, we can set

$$D_n^{-1}(x,x) = \begin{cases} 1/D_x^+ & \text{if } D_x^+ > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 2.6 (Existence of outlier, random walk)**.** *Consider the Chung–Lu model with $p_{x,y}$ as in Eq. (2.1). Let $T_n = D_n^{-1}A_n$. Then*

$$\max\{|\lambda_1(T_n) - \lambda_1(\mathbb{E}[T_n])|, |\lambda_2(T_n)|\} = O_{v.h.\mathbb{P}}\left(\frac{1}{\sqrt{s_n}}\right).$$

Theorem 2.7 (Existence of outliers - rank r case)**.** *Consider the model where $p_{x,y}$ is as in Eq. (2.2). If Assumption 2.4 holds, then*

$$\max\left\{(|\lambda_i(A_n) - \lambda_i(\mathbb{E}[A_n])|)_{i\le r}, |\lambda_2(A_n)|\right\} = O_{v.h.\mathbb{P}}(\sqrt{s_n}).$$

In particular there exists a constant $K_0 > 0$ such that w.v.h.p.

$$2K_0 s_n \ge \lambda_1(A_n) \ge \cdots \ge \lambda_r(A_n) \ge K_0 s_n.$$

Theorem 2.8 (Gaussian fluctuations, rank-one case)**.** *Consider the Chung–Lu model with $p_{x,y}$ as in Eq. (2.1). If If Assumptions 2.4 and 2.1 holds, then*

$$\sqrt{\frac{\boldsymbol{w}}{s_n}}\,(\lambda_1(A_n) - \mathbb{E}[\lambda_1(A_n)]) \xrightarrow[n\to+\infty]{\mathrm{d}} \mathcal{G},$$

where $\mathcal{G}$ is a centered Gaussian random variable with variance

$$\mathbb{V}\mathrm{ar}\,(\mathcal{G}) = \frac{\left(\int_{\mathbb{R}_+^2} (x^+)^2 x^- \, d\rho(x^+, x^-)\right)\left(\int_{\mathbb{R}_+^2} x^+ (x^-)^2 \, d\rho(x^+, x^-)\right)}{\left(\int_{\mathbb{R}_+^2} x^+ x^- \, d\rho(x^+, x^-)\right)^2}. \tag{2.4}$$

Notice the variance in (2.4) is finite, due to the compact support of ρ.

Theorem 2.9 (Gaussian fluctuations, rank-r case)**.** *Consider the model where $p_{x,y}$ is as in Eq. (2.2). Let $f : \mathbb{R}^{2r} \to (0, +\infty)$ be the function defined by $f(x_1^+, x_1^-, \ldots, x_r^+, x_r^-) = \sum_{k=1}^r \theta_k x_k^+ x_k^-$. If Assumptions 2.4 and 2.2 holds, then*

$$\sqrt{\frac{n}{s_n}}\,(\lambda_i(A_n) - \mathbb{E}[\lambda_i(A_n)])_{1\le j\le r} \xrightarrow[n\to+\infty]{\mathrm{d}} \mathcal{G},$$

where $\mathcal{G} = (\mathcal{G}_i)_{1\le i\le r}$ is a centered Gaussian random vector with covariances given, for $i, j = 1, \ldots, r$, by

$$\mathbb{C}\mathrm{ov}(\mathcal{G}_i, \mathcal{G}_j) = \frac{\int_{\mathbb{R}_+^{2r}} x_i^+ x_i^- f(x_1^\pm, \ldots, x_r^\pm) x_j^+ x_j^- \, d\rho(x_1^\pm, \ldots, x_r^\pm)}{\left(\int_{\mathbb{R}_+^{2r}} x_i^+ x_i^- \, d\rho(x_1^\pm, \ldots, x_r^\pm)\right)\left(\int_{\mathbb{R}_+^{2r}} x_j^+ x_j^- \, d\rho(x_1^\pm, \ldots, x_r^\pm)\right)}.$$

2.1.4 Methods

Let us first mention that, since the matrix A_n is not symmetric, standard self-adjoint techniques— such as the Hoffman-Wielandt inequality, which quantifies the impact of matrix perturbations on the spectrum—do not apply. In order to study the behavior of $\lambda_1(A_n)$, we will consider A_n as a random perturbation of $\mathbb{E}[A_n]$. In this non-reversible setting, we will employ Theorem 2.10, which applies to non-symmetric but diagonalizable matrices, and which will reduce the problem to establish a bound to the spectral norm of a given random matrix. To do so, we will employ the so-called *high trace method*, which was firstly developed in the context of undirected graphs with independent edges in Chung and Lu [2002a], Füredi and Komlós [1981], and then extended to directed graph with dependencies in Bordenave [2020], Coste [2021]. This provides a general way to bound the expectation of sufficiently high (diverging with n) moments of the spectral norm of the given random matrix. The analysis relies on a combinatorial estimate involving Catalan numbers and Dyck words.

When the existence of outliers living at the scale s_n is established, it is natural to investigate the fluctuations around the mean. To address this problem, we will follow the approach of Chakrabarty et al. [2020] and Erdős et al. [2013], writing the maximal eigenvalue $\lambda_1(A)$ as the solution of a fixed-point equation involving a random series.

Employing concentration results and high-probability estimates, the series can be written, up to lower order terms, as a an additive perturbation of $\mathbb{E}[\lambda_1(A)]$ by a sum of independent random variables. These, rescaled by $\sqrt{\frac{s_n}{n}}$, satisfy Lindeberg CLT, and are thus responsible for the emergence of Gaussian fluctuations. The same is done for the rank r model in (2.2), where the object to study is a random vector and there is the need to circumvent the inapplicability of Hoffmann–Wielandt inequality.

2.2 Analysis of outlier, rank-one case

2.2.1 Perturbation of non-Hermitian matrices

In order to provide a uniform control on the spectrum of a perturbed matrix it is possible to employ the following theorem, proved in Bauer and Fike [1960]. See also [Coste, 2021, Th. 4] for a modern formulation. We denote by $\mathcal{B}(\lambda, \varepsilon_n)$ the complex ball of center $\lambda \in \mathbb{C}$ and radius $\varepsilon_n > 0$.

Theorem 2.10 (Bauer–Fike, Bauer and Fike [1960])**.** *Let S_n be a $n \times n$ matrix such that $S_n = P_n D P_n^{-1}$ for an invertible matrix P_n and a diagonal matrix D. Let H_n be a $n \times n$ arbitrary matrix and $\varepsilon_n = \|P_n\|\|P_n^{-1}\|\|H_n\|$.*

(i) Then the spectrum of $S_n + H_n$ is contained in the union $\bigcup_{i=1}^{n} \mathcal{B}(\lambda_i(S_n), \varepsilon_n)$.

(ii) Moreover, if for $I \subseteq [n]$ it holds

$$\bigcup_{i \in I} \mathcal{B}(\lambda_i(S_n), \varepsilon_n) \cap \bigcup_{i \in I^{\complement}} \mathcal{B}(\lambda_i(S_n), \varepsilon_n) = \emptyset,$$

then $S_n + H_n$ has exactly $|I|$ eigenvalues inside $\bigcup_{i \in I} \mathcal{B}(\lambda_i(S_n), \varepsilon_n)$.

In the rank-one case, the previous statement can be specialized as follows.

Lemma 2.11 ([Coste, 2021, Lemma A.1])**.** *Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be two vectors of $\mathbb{R}^n$ and $S_n = \boldsymbol{x}\boldsymbol{y}^T$. Let H_n be a real $n \times n$ matrix.*

(i) The eigenvalues of the matrix $S_n + H_n$ are contained $\mathcal{B}(0, \varepsilon_n) \cup \mathcal{B}(\boldsymbol{y}^T\boldsymbol{x}, \varepsilon_n)$, where

$$\varepsilon_n = 2\|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2(\boldsymbol{y}^T\boldsymbol{x})^{-2}\|H_n\|;$$

(ii) If $\mathcal{B}(0, \varepsilon_n) \cap \mathcal{B}(\boldsymbol{y}^T\boldsymbol{x}, \varepsilon_n) = \emptyset$, then there is exactly one eigenvalue of $S_n + H_n$ inside $\mathcal{B}(\boldsymbol{y}^T\boldsymbol{x}, \varepsilon_n)$ and all the other eigenvalues of $S_n + H_n$ are contained in $\mathcal{B}(0, \varepsilon_n)$.

2.2.2 Existence of the outlier

Consider the real matrix $C_n := A_n - \mathbb{E}[A_n]$, so that $A_n = \mathbb{E}[A_n] + C_n$. We choose $H_n = C_n$, $\boldsymbol{x} = \sqrt{s_n} v^+$ and $\boldsymbol{y} = \sqrt{s_n} v^-$, where $v_x^{\pm} := \frac{w_x^{\pm}}{\sqrt{\boldsymbol{w}}}$. To prove Theorem 2.5, we only need to check that

$$\varepsilon_n = 2\|v^+\|^2\|v^-\|^2((v^-)^T v^+)^{-2}\|C_n\| = O_{v.h.\mathbb{P}}(\sqrt{s_n}).$$

Since $\|v^+\|^2\|v^-\|^2((v^-)^T v^+)^{-2}$ is bounded, it suffices to control the spectral norm of C_n, as provided by the next proposition.

Proposition 2.12. *There exists a constant $K_1 > 0$ and $\eta > 1$ such that, for large n,*

$$\mathbb{P}\left(\|C_n\| \geq K_1\sqrt{s_n}\right) \leq e^{-(\log(n))^\eta}.$$

The proof of this proposition relies on the following lemma.

Lemma 2.13. *There exists a constant $K_2 > 0$ such that, for $1 \ll m \ll \sqrt[4]{s_n}$,*

$$\mathbb{E}[\|C_n\|^{2m}] \leq 6n\left(K_2 s_n\right)^m.$$

Proof of Proposition 2.12. Taking t such that $t^{2m} = 6n\left(K_2 s_n\right)^m(1+\varepsilon)^{2m}$, where $\varepsilon = (\frac{\log(n)^\xi}{s_n})^{1/4}$, by the $2m$–moment Markov inequality we obtain

$$\mathbb{P}(\|C_n\| \geq t) \leq \frac{1}{t^{2m}}\mathbb{E}[\|C_n\|^{2m}] \leq e^{-2m\log(1+\varepsilon)} \leq e^{-(\log(n))^{\xi/4}}.$$

□

We are left with the proof of Lemma 2.13. To embark on the proof, we make some notational preliminaries. Since by definition $\|C_n\|^2 = \|C_n C_n^*\|$, we write

$$\begin{aligned}\|C_n\|^{2m} &= \lambda_1(C_n C_n^*)^m = \lambda_1((C_n C_n^*)^m) \leq \mathrm{Tr}((C_n C_n^*)^m)\\ &= \sum_{x_1,x_2,\dots,x_{2m}} C_{x_1x_2}\ C^*_{x_2x_3}C_{x_3x_4}\ C^*_{x_4x_5}\cdots C_{x_{2m-1}x_{2m}}\ C^*_{x_{2m}x_1}\\ &= \sum_{x_1,x_2,\dots,x_{2m}} C_{x_1x_2}C_{x_3x_2}C_{x_3x_4}C_{x_5x_4}\cdots C_{x_{2m-1}x_{2m}}C_{x_1x_{2m}},\end{aligned} \tag{2.5}$$

where the indices $x_1,\dots,x_{2m}$ run from 1 to n. Our aim is to provide a bound on the expectation of the latter sum. Using the notation $e^- = x$, $e^+ = y$ for any $e = (x,y) \in E$, we can define

$$\mathcal{P}_m := \left\{\mathfrak{p} = (e_1,\dots,e_{2m}) \in E^{2m} : e^+_{2i-1} = e^+_{2i},\quad e^-_{2i} = e^-_{2i+1} \text{ for all } i \in \{1,\dots,m\}\right\},$$

with the convention that $e_{2m+1} = e_1$. This set denotes a family of alternating edge-paths. For instance, if $m = 1$, elements of $\mathcal{P}_m$ will be of the form $((x_1,x_2),(x_1,x_2))$ for $x_1,x_2 \in [n]$, while for $m = 2$ we will have elements of the form $((x_1,x_2),(x_3,x_2),(x_3,x_4),(x_1,x_4))$ for $x_1,x_2,x_3,x_4 \in [n]$, and so on.

Then (2.5) reads

$$\|C_n\|^{2m} \leq \sum_{\mathfrak{p}\in\mathcal{P}_m}\prod_{s=1}^{2m} C_{e_s(\mathfrak{p})},$$

where $e_s(\mathfrak{p})$ denotes the s-th edge of $\mathfrak{p}$ and, if $e = (x,y)$, we set C_e to be the corresponding entry of C_n, C_{xy}. When taking expectation on both sides, we can restrict the sum to the smaller set of paths having each edge repeated at least twice, because $C_n = A_n - \mathbb{E}[A_n]$

has independent and centered entries. Let us denote with $\mathcal{R}_m$ such subset of $\mathcal{P}_m$. Let us also denote with $\ell(\mathfrak{p})$ the number of distinct edges in $\mathfrak{p} \in \mathcal{R}_m$ and with $E(\mathfrak{p}) = (\tilde{e}_1, \dots, \tilde{e}_{\ell(\mathfrak{p})})$ the ordered sequence of such distinct edges of $\mathfrak{p} \in \mathcal{R}_m$. We get

$$\mathbb{E}[\|C_n\|^{2m}] \le \sum_{\mathfrak{p}\in\mathcal{R}_m} \prod_{s=1}^{\ell(\mathfrak{p})} \mathbb{E}\left[C_{\tilde{e}_s(\mathfrak{p})}^{k_s(\mathfrak{p})}\right], \tag{2.6}$$

where $k_s(\mathfrak{p}) \ge 2$ denotes the multiplicity of $\tilde{e}_s(\mathfrak{p})$ in $\mathfrak{p}$. We are ready to prove Lemma 2.13.

Proof of Lemma 2.13. Given a path $\mathfrak{p} \in \mathcal{R}_m$, consider the sequence of vertices defined by the following iterative procedure. For $j = 0$, set $\tilde{v}_0 = \tilde{e}_1^-$. Then, for $j = 1, \dots, \ell(\mathfrak{p})$:

- set $\tilde{v}_j = \tilde{e}_j^+$ if the first occurrence of $\tilde{e}_j$ in $\mathfrak{p}$ occupies an odd position;
- set $\tilde{v}_j = \tilde{e}_j^-$ otherwise (if the first occurrence of $\tilde{e}_j$ in $\mathfrak{p}$ occupies an even position).

Let $V(\mathfrak{p}) = (\tilde{v}_0, \dots, \tilde{v}_{\ell(\mathfrak{p})})$. Notice that while $E(\mathfrak{p})$ has exactly $\ell(\mathfrak{p})$ distinct edges, $V(\mathfrak{p})$ has exactly $\ell(\mathfrak{p}) + 1$ vertices and maybe some of them will be repeated. Let $\#V(\mathfrak{p})$ denote the number of distinct vertices in $V(\mathfrak{p})$.

In what follows, we want to identify a subfamily of paths in $\mathcal{R}_m$ that provides the main contribution to the sum in (2.6). To this aim, for $1 \le p-1 \le l \le m$, let us define

$$\mathcal{R}_{m,l,p} := \{\mathfrak{p} \in \mathcal{R}_m \mid \ell(\mathfrak{p}) = l, \#V(\mathfrak{p}) = p\},$$

so that (2.6) becomes

$$\mathbb{E}[\|C_n\|^{2m}] \le \sum_{l=1}^{m}\sum_{p=2}^{l+1}\sum_{\mathfrak{p}\in\mathcal{R}_{m,l,p}} \prod_{s=1}^{\ell(\mathfrak{p})} \mathbb{E}\left[C_{\tilde{e}_s(\mathfrak{p})}^{k_s(\mathfrak{p})}\right]. \tag{2.7}$$

We will show that the sum over $\mathcal{R}_{m,m,m+1}$ will give the leading order for the total sum.

To see it, we first associate to each path $\mathfrak{p} = (e_1, \dots, e_{2m}) \in \mathcal{R}_m$ a code $\mathfrak{c}(\mathfrak{p}) = (\mathfrak{c}_1, \dots, \mathfrak{c}_{2m})$ of $2m$ marks, in the following way. Recall the notation $E(\mathfrak{p}) = (\tilde{e}_1, \dots, \tilde{e}_{\ell(\mathfrak{p})})$. For $j = 1, \dots, \ell(\mathfrak{p})$:

- if e_j appears for the first time, set $\mathfrak{c}_j = +$;
- if e_j appears for the second time, set $\mathfrak{c}_j = -$;
- otherwise, if $e_j = \tilde{e}_k$ for some $k \in [\ell(\mathfrak{p})]$, set $\mathfrak{c}_j = k$.

We want to count the number of possible codes that can be built with this procedure. First of all, by definition of $\ell(\mathfrak{p})$, notice that there can be at most $2m - 2\ell(\mathfrak{p})$ marks different from "$\pm$": their positions can be chosen in at most $\binom{2m}{2m-2\ell(\mathfrak{p})}$ ways and each of them can takes values in a set of $\ell(\mathfrak{p})$ elements. Moreover, notice that, for every $j \le 2m$

the number of marks "$-$" up to level j cannot exceed the number of marks "$+$" up to level j. In particular, writing ℓ for $\ell(\mathfrak{p})$ for simplicity, the number of such "$\pm$" sequences (which are called Dyck words) is given by the ℓ-th Catalan number

$$\mathcal{C}_\ell := \binom{2\ell}{\ell}\frac{1}{\ell+1} \leq 4^\ell.$$

As a consequence, the number of possible meaningful codes is at most

$$\mathcal{C}_\ell \binom{2m}{2m-2\ell} \ell^{2m-2\ell} \leq 4^\ell (2m\ell)^{2m-2\ell} \leq 4^m m^{4(m-\ell)}.$$

It is not difficult to see that, for each $l = 1, \dots, m$, the paths $\mathfrak{p}$ in $\mathcal{R}_{m,l,l+1}$ are in bijection with the corresponding couples $(\mathfrak{c}(\mathfrak{p}), V(\mathfrak{p}))$. Indeed, reading a code $\mathfrak{c}$ it is possible to completely reconstruct the structure of the path $\mathfrak{p}$, and the further knowledge of a sequence V with distinct vertices will allow to identify the labels of its vertices. This does not hold anymore for paths in $\mathcal{R}_{m,l,p}$ with $p < l+1$: in that case, the information contained in a couple $(\mathfrak{c}, V)$ is no more sufficient to determine the order of appearance for the repetitions of certain subsequences of directed edges. For instance, consider the couple $(\mathfrak{c}, V)$ where

$$\mathfrak{c} = (+,+,+,+,+,+,+,+,-,-,-,-,-,-,-,-), \qquad V = (1,2,3,4,1,5,6,7,1) \in [n]^9.$$

If we try to assign a path $\mathfrak{p} \in \mathcal{R}_{m,l,p}$ to $(\mathfrak{c}, V)$, the first 6 edges of the path are unequivocally determined, but the order of the remaining 6 edges (which will be repetitions of the first 6) can be chosen in 8 different ways. Two possibilities are, e.g., the sequence

$$((1,2),(3,2),(3,4),(1,4),(1,5),(6,5),(6,7),(1,7)),$$

and the sequence

$$((1,7),(6,7),(6,5),(1,5),(1,2),(3,2),(3,4),(1,4)).$$

However, we can bound the number of possible permutations of repeated vertices, by observing that the worst case is achieved when a vertex is repeated in $V(\mathfrak{p})$ a number of $l+1-p$ times. Taking into account 2 possible orientations for any meaningful sub–path (e.g. $((1,2),(3,2),(3,4),(1,4))$ or $((1,4),(3,4),(3,2),(1,2))$ in the previous example) we end up with an upper bound of

$$2^{l+1-p}(l+1-p)! \leq (2(l+1-p))^{l+1-p}$$

possible paths leading to a given couple $(\mathfrak{c}, V)$.

At this point, let us observe that, for every $k \geq 2$ and $x, y \in [n]$, it holds

$$\mathbb{E}[C_{xy}^k] = (1-p_{x,y})^k p_{x,y} + (-p_{x,y})^k (1-p_{x,y}) \leq p_{x,y}.$$

Indeed, for k even, we have $(1-p_{x,y})^{k-1}+(p_{x,y})^{k-1} < (1-p_{x,y}+p_{x,y})^{k-1} = 1$ which implies $(1-p_{x,y})^{k}+(p_{x,y})^{k-1}(1-p_{x,y}) < 1$, while for k odd we have $(1-p_{x,y})^{k-1} < 1+p_{x,y}^{k-1}$, implying $(1-p_{x,y})^{k} - p_{x,y}^{k-1}(1-p_{x,y}) < 1$.

As a consequence, the contribution of each path with l distinct edges will be at most given by $p_{\max}^l$, where $p_{\max} = \max_{x,y\in[n]} p_{x,y}$. Moreover, since the number of sequences V with $\#V = p$ can be bounded by $n^p l^{(l+1-p)}$ (p vertices chosen in $[n]$ and the remaining $l+1-p$ among the first $p \le l$), we can upper bound (2.7) as follows

$$\begin{aligned}\mathbb{E}[\|C_n\|^{2m}] &\le \sum_{l=1}^{m}\sum_{p=2}^{l+1}\sum_{\mathfrak{p}\in\mathcal{R}_{m,l,p}} p_{\max}^l \\ &\le \sum_{l=1}^{m} 4^m m^{4(m-l)} \sum_{p=2}^{l+1}(2(l+1-p))^{l+1-p} n^p l^{l+1-p} p_{\max}^l.\end{aligned} \tag{2.8}$$

Since

$$\begin{aligned}\sum_{p=2}^{l}(2(l+1-p))^{l+1-p} n^p l^{l+1-p} &\le \left(2l^2\right)^{l+1}\sum_{p=1}^{l}\left(\frac{n}{2l^2}\right)^p \\ &\le \left(2l^2\right)^{l+1} 2\left(\frac{n}{2l^2}\right)^{l+1} \le 2n^{l+1},\end{aligned}$$

we can bound the l.h.s. of (2.8) by $3\cdot 4^m \sum_{l=1}^m E_{m,l}$, where $E_{m,l} \coloneqq m^{4(m-l)} n^{l+1} p_{\max}^l$. Let us now consider the ratio

$$\frac{E_{m,m}}{E_{m,l}} = \frac{n^{m+1}p_{\max}^m}{m^{4(m-\ell(\mathfrak{p}))} n^{l+1} p_{\max}^l} = \left(\frac{np_{\max}}{m^4}\right)^{m-l}. \tag{2.9}$$

Since $p_{\max} \sim s_n/n$ by (2.1) and the boundedness of weights, and $1 \ll m \ll \sqrt[4]{s_n}$, the term in brackets diverges as n grows, we get

$$\mathbb{E}[\|C_n\|^{2m}] \le 3\cdot 4^m \sum_{l=1}^{m} E_{m,l} \le 6\cdot 4^m E_{m,m} = 6\cdot 4^m n^{m+1} p_{\max}^m \lesssim 6n(4s_n)^m, \tag{2.10}$$

which concludes the proof of Lemma 2.13. □

Remark 2.14 (Boundedness of weights)**.** *Notice that weights must be bounded for the proof of Lemma 2.13 to hold. In particular, if some some vertices have diverging weights, then the relevant contribution in the sum over $\mathcal{R}_{m,l,p}$ is given by paths involving those weights, and the uniform asymptotic bound in Eq. (2.10) does not work.*

Remark 2.15. *Assumption 2.4 is crucial in estimating the term in (2.9). If it does not hold, then the term $E_{m,m}$ does not dominate the others and the asymptotic bound in Eq. (2.10) does not work.*

2.2.3 Transition matrix

We adapt here the proof of Theorem 2.5 to the transition matrix case.

Proof of Theorem 2.6. Let us consider the transition matrix of the simple random walk on G_n, $T_n = D_n^{-1} A_n$. In this setting, we consider the matrix $\widetilde{C}_n = D_n^{-1}(A_n - \mathbb{E}[A_n])$, which is not centered, and we apply Lemma 2.11 with $H_n = \widetilde{C}_n$ and the choice of vectors

$$\boldsymbol{x} = \frac{s_n}{\sqrt{\boldsymbol{w}}}(D_1^{-1} w_1^+, \dots, D_n^{-1} w_n^+) \quad \text{and} \quad \boldsymbol{y} = v^- = \frac{1}{\sqrt{\boldsymbol{w}}}(w_1^-, \dots, w_n^-).$$

In order to conclude the proof, we need to provide a suitable bound for the radius

$$\tilde{\varepsilon}_n = 2\|\boldsymbol{x}\|^2 \|\boldsymbol{y}\|^2 (\boldsymbol{y}^T \boldsymbol{x})^{-2} \|\widetilde{C}_n\|.$$

Let us start by analyzing $\|\widetilde{C}_n\|$. Since $\widetilde{C}_n$ is not centered, we cannot directly apply the machinery developed in Subsection 2.2.2. We then define

$$\overline{C}_n := (\mathbb{E}[D_n])^{-1} (A_n - \mathbb{E}[A_n]) = (\mathbb{E}[D_n])^{-1} C_n.$$

This matrix is centered and by sub-multiplicativity, it holds

$$\|\widetilde{C}_n - \overline{C}_n\| = \|(D_n^{-1} - (\mathbb{E}[D_n])^{-1}) C_n\| \le \|D_n^{-1} - (\mathbb{E}[D_n])^{-1}\| \|C_n\|.$$

Thanks to the above analysis (Proposition 2.12) it holds $\|C_n\| = O_{v.h.\mathbb{P}}(\sqrt{s_n})$.
By Lemma 2.21, there exists $\eta > 4/3$ such that

$$\mathbb{P}\left(\max_{x \in V} \left| \frac{1}{D_x^+} - \frac{1}{w_x^+ s_n} \right| \ge 2 s_n^{-\frac{4}{3}} \right) \le \exp(-\log(n)^\eta).$$

Recalling that $\mathbb{E}[D_x^+] = w_x^+ s_n$, this implies that $\|D_n^{-1} - (\mathbb{E}[D_n])^{-1}\| = O_{v.h.\mathbb{P}}(s_n^{-4/3})$. Then

$$\|\widetilde{C}_n - \overline{C}_n\| = O_{v.h.\mathbb{P}}(s_n^{-5/6}) = O_{v.h.\mathbb{P}}(s_n^{-1/2}).$$

Then we can repeat the procedure of Subsection 2.2.2 to the centered matrix $\overline{C}_n$ and get that

$$\|\widetilde{C}_n\| \le \|\widetilde{C}_n - \overline{C}_n\| + \|\overline{C}_n\| = O_{v.h.\mathbb{P}}(s_n^{-1/2}).$$

It remains to bound the other terms appearing in the definition of $\tilde{\varepsilon}_n$. Notice that $\boldsymbol{x}$ is a random vector and hence the same holds for the unique non-zero eigenvalue of $S = \boldsymbol{y}(\boldsymbol{x})^T$, which is

$$\lambda_1(S) = \boldsymbol{y}^T \boldsymbol{x} = \sum_{x \in V} \frac{w_y^-}{\boldsymbol{w}} \frac{w_x^+ s_n}{D_x^+}.$$

However,

$$|\lambda_1(S) - 1| = \left| \sum_{x \in V} \frac{w_y^-}{\boldsymbol{w}} \left(\frac{w_x^+ s_n}{D_x^+} - 1 \right) \right| \le \sum_{x \in V} \frac{w_x^-}{\boldsymbol{w}} \left| \frac{w_x^+ s_n}{D_x^+} - 1 \right| \le \max_{x \in V} \left| \frac{w_x^+ s_n}{D_x^+} - 1 \right|,$$

where the first inequality is by convexity, and the last term is $O_{v.h.\mathbb{P}}(s_n^{-\frac{1}{3}})$ thanks to Lemma 2.21. Then

$$|\lambda_1(S)^2 - 1| \leq |\lambda_1(S) - 1| \cdot |\lambda_1(S) + 1| \leq |\lambda_1(S) - 1| \cdot (2 + |\lambda_1(S) - 1|),$$

and we conclude that $|\lambda_1(S)^2 - 1| = O_{v.h.\mathbb{P}}(s_n^{-\frac{1}{3}})$. Moreover

$$\begin{aligned} \|\boldsymbol{x}\|^2 = \sum_{x \in V} \frac{1}{\boldsymbol{w}} \left(\frac{w_x^+ s_n}{D_x^+} \right)^2 &\leq \frac{n}{\boldsymbol{w}} \max_{x \in V} \left(\frac{w_x^+ s_n}{D_x^+} \right)^2 \\ &\leq \frac{n}{\boldsymbol{w}} \left(1 + \max_{x \in V} \left| \frac{w_x^+ s_n}{D_x^+} - 1 \right| \right)^2, \end{aligned}$$

which yields $\|\boldsymbol{x}\|^2 = O_{v.h.\mathbb{P}}(1)$, again by Lemma 2.21. Then, it holds $\tilde{\varepsilon}_n = O_{v.h.\mathbb{P}}(s_n^{-\frac{1}{2}})$. Thus, w.v.h.p. , it holds $\mathcal{B}(0, \tilde{\varepsilon}_n) \cap \mathcal{B}(\lambda_1(S), \tilde{\varepsilon}_n) = \emptyset$, and, applying Lemma 2.11(ii), there exists a unique eigenvalue of P around 1, which is 1 itself; the other eigenvalues are contained in $\mathcal{B}(0, \tilde{\varepsilon}_n)$. This completes the proof of Theorem 2.6. □

2.2.4 Fluctuations around the mean

For notational convenience let $\lambda = \lambda_1(A_n)$ and let v denote a corresponding unit eigenvector. It holds $A_n v = C_n v + \mathbb{E}[A_n]v = \lambda v$, and pre-multiplying by v^T,

$$v^T C_n v + v^T \mathbb{E}[A_n] v = \lambda.$$

By Theorem 2.5, λ is of order s_n w.v.h.p. and, due to Proposition 2.12, $v^T C_n v$ has lower order (it holds $\|C_n\| = O_{v.h.\mathbb{P}}(\sqrt{s_n})$). We get that $v^T \mathbb{E}[A_n] v / s_n = v^T (v^+)(v^-)^T v$ does not vanish w.v.h.p., and so does $(v^-)^T v$. Moreover, there exists $\eta > 1$ and $K > 0$, such that the event $\{\|C_n\| \geq K\sqrt{s_n}\}$ has probability at most $\exp(-\log(n)^\eta)$, and thus the matrix $I_n - \frac{C_n}{\lambda}$ is w.v.h.p. invertible, so that w.v.h.p. the following display holds:

$$\begin{aligned} & (\lambda I_n - C_n) v = \mathbb{E}[A_n] v \\ \Longrightarrow \quad & \lambda v = \left(I_n - \frac{C_n}{\lambda} \right)^{-1} \mathbb{E}[A_n] v \\ \Longrightarrow \quad & \lambda v = \sum_{k=0}^{+\infty} \left(\frac{C_n}{\lambda} \right)^k \mathbb{E}[A_n] v \\ \Longrightarrow \quad & \lambda (v^-)^T v = \sum_{k=0}^{+\infty} (v^-)^T \left(\frac{C_n}{\lambda} \right)^k \mathbb{E}[A_n] v \\ \Longleftrightarrow \quad & \lambda (v^-)^T v = \sum_{k=0}^{+\infty} (v^-)^T \left(\frac{C_n}{\lambda} \right)^k s_n v^+ (v^-)^T v. \end{aligned} \tag{2.11}$$

Since $(v^-)^T v \neq 0$, w.v.h.p. we end up with

$$\lambda = s_n \sum_{k=0}^{+\infty} (v^-)^T \left(\frac{C_n}{\lambda}\right)^k v^+.$$

Let $L = \lceil \log(n) \rceil$. We have that, w.v.h.p.

$$\lambda = s_n (v^-)^T v^+ + s_n (v^-)^T \frac{C_n}{\lambda} v^+ + R^{(1)} + R^{(2)} + R^{(3)}, \tag{2.12}$$

where

$$R^{(1)} = s_n \sum_{k=L+1}^{+\infty} (v^-)^T \left(\frac{C_n}{\lambda}\right)^k v^+. \qquad \text{(high exponent)}$$

$$R^{(2)} = s_n \sum_{k=2}^{L} (v^-)^T \left(\frac{C_n - \mathbb{E}[C_n]}{\lambda}\right)^k v^+, \qquad \text{(centering)}$$

$$R^{(3)} = s_n \sum_{k=2}^{L} (v^-)^T \left(\frac{\mathbb{E}[C_n]}{\lambda}\right)^k v^+, \qquad \text{(main contribution)}$$

We are going to show that the only relevant term is the third one, being the other two negligible. For what concerns $R^{(1)}$, thanks to Proposition 2.12 and Theorem 2.5, w.v.h.p. it holds

$$|R^{(1)}| \leq \sum_{k=\log(n)+1}^{+\infty} s_n \frac{\|v^-\| \|v^+\| \|W\|^k}{\lambda^k} \leq \sum_{k=\log(n)+1}^{+\infty} \left(\frac{K_1 s_n^{1/2}}{K_0 s_n}\right)^k = O\left(s_n^{-\log(n)/3}\right).$$

To estimate $R^{(2)}$, applying Lemma 2.25 and Theorem 2.5, w.v.h.p.

$$|R^{(2)}| \leq s_n \sum_{k=2}^{L} \frac{\left(s_n^{1/2} \log(n)^{\xi/4}\right)^k}{n^{1/2}} \left(\frac{1}{K_0 s_n}\right)^k = \frac{s_n}{n^{1/2}} \sum_{k=2}^{L} \left(\frac{\log(n)^{\xi/4}}{K_0 s_n^{1/2}}\right)^k = O\left(\sqrt{\frac{\log(n)^\xi}{n}}\right).$$

Combining the above two estimates, we have

$$\lambda = (v^-)^T v^+ s_n + s_n \sum_{k=2}^{L} (v^-)^T \left(\frac{\mathbb{E}[C_n]}{\lambda}\right)^k v^+ + O_{v.h.\mathbb{P}}\left(\sqrt{\frac{\log(n)^\xi}{n}}\right) \tag{2.13}$$

It would be tempting to think that $R^{(3)}$ behaves as $R^{(1)}$ and $R^{(2)}$, but it turns out that the term $s_n (v^-)^T v^+$ alone does not provide an estimate of $\mathbb{E}[\lambda]$, which has to be given in terms of the entire sum over $k \in \{0, \dots, L\}$ as Lemma 2.16 and Lemma 2.17 will show. To get this, consider the fixed point equation

$$x = h(x) := s_n \sum_{k=0}^{\log(n)} \frac{(v^-)^T \mathbb{E}[W_n^k] v^+}{x^k} = s_n (v^-)^T v^+ + s_n \sum_{k=2}^{\log(n)} \frac{(v^-)^T \mathbb{E}[W_n^k] v^+}{x^k}. \tag{2.14}$$

For fixed n, $h : (0, +\infty) \to (0, +\infty)$ is decreasing as x grows, so that there exists a unique solution $\tilde{\lambda}$. Moreover choosing $x = ts_n$, for $t \in (0 + \infty)$, and using Lemma 2.24, we get that

$$\frac{h(ts_n)}{s_n} = (v^-)^T v^+ + \sum_{k=2}^{\log(n)} \frac{(v^-)^T \mathbb{E}[W_n^k] v^+}{(ts_n)^k} = (v^-)^T v^+ (1 + o(1)),$$

so that we conclude that $\tilde{\lambda} = (v^-)^T v^+ s_n (1 + o(1))$.

Lemma 2.16. *In the present setting, it holds*

$$\lambda - \tilde{\lambda} = s_n \frac{(v^-)^T C_n v^+}{\tilde{\lambda}} + o_{v.h.\mathbb{P}}\left(\sqrt{\frac{s_n}{n}}\right).$$

Proof. Combining (2.13) with the definition of $\tilde{\lambda}$ we get

$$\lambda - \tilde{\lambda} = s_n \frac{(v^-)^T C_n v^+}{\lambda} + s_n \sum_{k=0}^{\log(n)} \left(\frac{1}{\lambda^k} - \frac{1}{\tilde{\lambda}^k}\right) (v^-)^T \mathbb{E}[W_n^k] v^+ + O_{v.h.\mathbb{P}}\left(\sqrt{\frac{\log(n)^\xi}{n}}\right).$$

By Theorem 2.5 it holds

$$\frac{1}{\lambda^k} - \frac{1}{\tilde{\lambda}^k} = (\tilde{\lambda} - \lambda) \left(\frac{\sum_{j=1}^{k-1} \tilde{\lambda}^j \lambda^{k-j}}{\lambda^k \tilde{\lambda}^k}\right) = (\tilde{\lambda} - \lambda) O_{v.h.\mathbb{P}}\left(\frac{k}{s_n^{k+1}}\right), \tag{2.15}$$

so that, applying Lemma 2.24, it results

$$\begin{aligned} \left| s_n \sum_{k=0}^{\log(n)} \left(\frac{1}{\lambda^k} - \frac{1}{\tilde{\lambda}^k}\right) (v^-)^T \mathbb{E}[W_n^k] v^+ \right| &\le |\lambda - \tilde{\lambda}|\, O_{v.h.\mathbb{P}}\left(s_n \sum_{k=2}^{\log(n)} \frac{k}{s_n^{k+1}} (K_1 s_n)^{k/2}\right) \\ &= O_{v.h.\mathbb{P}}\left(\frac{|\lambda - \tilde{\lambda}|}{s_n}\right), \end{aligned} \tag{2.16}$$

where we used that $\sum_{k=2}^{\log(n)} k / s_n^{k/2} = O(1/s_n)$. As a consequence,

$$\lambda - \tilde{\lambda} = s_n \frac{(v^-)^T C_n v^+}{\lambda} + O_{v.h.\mathbb{P}}\left(\frac{|\lambda - \tilde{\lambda}|}{s_n}\right) + O_{v.h.\mathbb{P}}\left(\sqrt{\frac{\log(n)^\xi}{n}}\right). \tag{2.17}$$

By Lemma 2.23 and Theorem 2.5 it holds $s_n \frac{(v^-)^T C_n v^+}{\lambda} = O_{v.h.\mathbb{P}}(\sqrt{\frac{s_n}{n}})$. Then, Eq. (2.17) implies

$$|\lambda - \tilde{\lambda}| = O_{v.h.\mathbb{P}}\left(\sqrt{\frac{s_n}{n}}\right). \tag{2.18}$$

Consequently, we can omit the second addend in the r.h.s. of Eq. (2.17) and get the more precise estimate

$$\lambda - \tilde{\lambda} = s_n \frac{(v^-)^T C_n v^+}{\lambda} + o_{v.h.\mathbb{P}}\left(\sqrt{\frac{s_n}{n}}\right). \tag{2.19}$$

Reasoning as in (2.15), and thanks to (2.18), we also have

$$\left|\left(\frac{1}{\lambda} - \frac{1}{\tilde{\lambda}}\right)(v^-)^T C_n v^+\right| = o_{v.h.\mathbb{P}}(|\lambda - \tilde{\lambda}|) = o_{v.h.\mathbb{P}}\left(\sqrt{\frac{s_n}{n}}\right).$$

Then, we can change the λ in the denominator of (2.19) to $\tilde{\lambda}$. This proves Lemma 2.16. □

Lemma 2.17. *It holds* $\mathbb{E}[\lambda] - \tilde{\lambda} = o\left(\sqrt{\frac{s_n}{n}}\right)$.

Proof. Let $R := \lambda - \tilde{\lambda} - s_n (v^-)^T \frac{C_n}{\tilde{\lambda}} v^+$. By equation (2.19) there exists $\eta > 1$ such that, for any $\delta > 0$, it holds

$$\mathbb{E}[|R|] < \delta\sqrt{\frac{s_n}{n}} + \left(\mathbb{E}\left[\left(\lambda - \tilde{\lambda} - s_n \frac{v^- C_n v^+}{\tilde{\lambda}}\right)^2\right]\right)^{\frac{1}{2}} \exp\left(-\frac{(\log(n))^\eta}{2}\right) = o\left(\sqrt{\frac{s_n}{n}}\right).$$

Since $|\mathbb{E}[\lambda] - \tilde{\lambda}| = |\mathbb{E}[R]| \leq \mathbb{E}[|R|]$, we conclude. □

We are ready to prove the main theorem.

Proof of Theorem 2.8. Thanks to Lemma 2.16 and Lemma 2.17, it holds

$$\sqrt{\frac{n}{s_n}}\,(\lambda_1(A_n) - \mathbb{E}[\lambda_1(A_n)]) = \sqrt{n s_n}\frac{(v^-)^T C_n v^+}{\tilde{\lambda}} + o_{v.h.\mathbb{P}}(1).$$

The first term of the r.h.s. is a sum of independent random variables satisfying the hypotheses of Lindeberg CLT. To identify the variance we just need to compute

$$\begin{aligned}\mathrm{Var}\left(\sqrt{n s_n}\frac{(v^-)^T C_n v^+}{\tilde{\lambda}}\right) &= \frac{n s_n}{(\sum_{x\in[n]} v_x^+ v_x^- s_n)^2 (1+o(1))} \sum_{x,y\in[n]} (v_x^-)^2 (v_y^+)^2 v_x^+ v_y^- s_n (1 - p_{x,y}), \\ &\sim \frac{\left(\frac{1}{n}\sum_{x\in[n]}(\sqrt{\frac{n}{\boldsymbol{w}}}w_x^-)^2(\sqrt{\frac{n}{\boldsymbol{w}}}w_x^+)\right)\left(\frac{1}{n}\sum_{y\in[n]}(\sqrt{\frac{n}{\boldsymbol{w}}}w_y^-)(\sqrt{\frac{n}{\boldsymbol{w}}}w_y^+)^2\right)}{\left(\frac{1}{n}\sum_{x\in[n]}(\sqrt{\frac{n}{\boldsymbol{w}}}w_x^+)(\sqrt{\frac{n}{\boldsymbol{w}}}w_x^-)\right)^2}.\end{aligned} \tag{2.20}$$

Using the hypothesis, the sums converge to integrals and we get (2.4). □

2.3 Analysis of outliers, higher-rank case

2.3.1 Existence of outliers

For the model defined in (2.2), Proposition 2.12 and Lemma 2.13 still hold, with the according definition of C_n. To establish Theorem 2.7, we just need to adapt the Bauer–Fike step, by employing directly Theorem 2.10, with the choice $S_n = \mathbb{E}[A_n]$. We need to show that

$$\varepsilon_n = \|P_n\|\|P_n^{-1}\|\|C_n\| = O_{v.h.\mathbb{P}}(\sqrt{s_n}),$$

where P_n is a diagonalizing change of basis for $\mathbb{E}[A_n]$. In this setting, the change of basis P can be chosen to be

$$P_n = (v_1^+, \dots, v_r^+, e_{r+1}, \dots, e_n),$$

where $(e_l)_{r+1\le l\le n}$ is an orthonormal basis of $\mathrm{Span}(v_1^-, \dots, v_r^-)^\perp$. P_n is not an orthogonal matrix, but it holds $\|P_n\|_F = \sqrt{n}$. Moreover, considering the matrix $X_n = (v_1^-, \dots, v_r^-, e_{r+1}, \dots, e_n)$, we have that $X_n^* P_n$ is lower triangular with unit determinant, so that $\det(P_n) = \det(X_n)^{-1}$. It holds

$$\max_{i,j\le r} \mathrm{dist}(v_i^-, v_j^-)^2 = \frac{1}{n} \cdot \max_{i,j\le r} \sum_{x\in[n]} \left(\sqrt{n}\, v_i^-(x) - \sqrt{n}\, v_j^-(x)\right)^2,$$

and the l.h.s. is uniformly bounded, thanks to Assumption 2.2 (3). Then

$$\det(X_n) = \prod_{l=0}^{r-1} \mathrm{dist}(v_{l+1}^-, V_l^-) = O(1).$$

We have the bound (which is the main contribution in Guggenheimer et al. [1995])

$$\|P_n\|\|P_n^{-1}\| \le \frac{2}{|\det(P_n)|}\left(\frac{\|P_n\|_F}{\sqrt{n}}\right)^n = 2|\det(X_n)|.$$

Then, by Proposition 2.12 we conclude $\varepsilon_n = O_{v.h.\mathbb{P}}(\sqrt{s_n})$.

2.3.2 Fluctuations around the mean

Let us fix $l \in \{1, \dots, r\}$. To simplify the notation, let $\lambda_l = \lambda_l(A_n)$. We also consider the $r \times r$ matrix with entries

$$V_n(i,j) := s_n \sqrt{\theta_i\theta_j} (v_i^-)^T \left(I_n - \frac{C_n}{\lambda_l}\right)^{-1} v_j^+ \mathbf{1}_{\{\|W\|<\lambda_l\}} \qquad i,j = 1, \dots, r. \tag{2.21}$$

Notice that, thanks to Theorem 2.9 and by the conditions on $(v_i^\pm)_{i\le r}$, it holds that

$$V_n = s_n \,\mathrm{Diag}(\theta_1, \dots, \theta_r)\left(1 + o_{v.h.\mathbb{P}}\left(\frac{\|C_n\|}{\lambda_l}\right)\right),$$

that is, V_n is a perturbation of a diagonal matrix and it is diagonalizable (say, because it is with high probability strictly dominant). Precisely, the outliers of A_n provide all the eigenvalues of V_n, as the following lemma states.

Lemma 2.18. *With very high probability it holds $\lambda_l(A_n) = \lambda_l(V_n)$.*

Proof. Let v be a right unit eigenvector of A_n corresponding to the eigenvalue $\lambda_l = \lambda_l(A_n)$,

$$\lambda_l v = C_n v + s_n \sum_{j=1}^{r} \theta_j v_j^+ (v_j^-)^T v. \tag{2.22}$$

Reasoning as in (2.11), w.v.h.p. it holds

$$\lambda_l v = \left(I_n - \frac{C_n}{\lambda_l} \right)^{-1} s_n \sum_{j=1}^{r} \theta_j v_j^+ (v_j^-)^T v,$$

so that, pre-multiplying by $\sqrt{\theta_i}(v_i^-)^T$, for $i = 1, \dots, r$, and recalling the definition (2.21), we have

$$\lambda_l \sqrt{\theta_i}(v_i^-)^T v = \sum_{j=1}^{r} V_n(i,j) \sqrt{\theta_j}(v_j^-)^T v, \qquad i = 1, \dots, r.$$

Calling $u^\pm = (\sqrt{\theta_1}(v_1^\pm)^T v, \dots, \sqrt{\theta_r}(v_r^\pm)^T v)^T$, we have u^- that is a candidate eigenvector of eigenvalues λ_l for V_n. We just have to show that it is not the null vector. Pre-multiplying (2.22) by v^T it holds

$$\lambda_l = v^T C_n v + s_n u^+ u^-.$$

Since λ_l is of order s_n w.v.h.p. and $v^T C_n v$ has lower order (thanks to Proposition 2.12 it holds $\|C_n\| = O_{v.h.\mathbb{P}}(\sqrt{s_n})$), we deduce that u^- has at least one non-vanishing entry. This shows that $\lambda_l(A_n) \in Sp(V_n)$. To have the thesis, it suffices to employ Gerschgorin Theorem ([Varga, 2004, Theorem 1.6] or [Chakrabarty et al., 2020, Fact 5.1]) as in [Chakrabarty et al., 2020, Lemma 5.2], after having noticed that A_n does not need to be symmetric. ☐

Let us rewrite V_n as the following sum

$$V_n = \sum_{k=0}^{+\infty} V_{k,n},$$

where for every $k \in \mathbb{N}$, $V_{k,n}$ is the matrix with entries

$$V_{k,n}(i,j) := s_n \sqrt{\theta_i \theta_j} (v_i^-)^T (C_n)^k v_j^+ \qquad i,j = 1, \dots, r.$$

The decomposition (2.12) needs to be adapted to the r-dimensional context. This will be the aim of the next Lemmata. Let us consider the following fixed point equation

$$x = h_l(x) := \lambda_l \left(\sum_{k=0}^{L} \frac{\mathbb{E}[V_{k,n}]}{x^k} \right),$$

which generalizes the one in (2.14). Letting $x = ts_n$ for $t \in (0, +\infty)$, by Lemma 2.24 it holds

$$\left\| \sum_{k=2}^{L} \frac{\mathbb{E}[V_{k,n}]}{(ts_n)^k} \right\| \leq \sum_{k=2}^{+\infty} (ts_n)^{-k} (K_1 s_n)^{k/2+1} = \left(\frac{K_1}{t}\right)^2 (1 + O(s^{-1/2})).$$

As a consequence, by definition of $V_{0,n}$ and the properties of $(v_i^{\pm})_{i \leq r}$,

$$s_n^{-1} \sum_{k=0}^{L} (ts_n)^{-k} \mathbb{E}[V_{k,n}] = \mathrm{Diag}(\theta_1, \dots, \theta_r)(1 + o(1)).$$

In particular $h_l(ts_n) = \theta_l s_n(1 + o(1))$. From this follows that, for $t < \theta_l$ and large n it holds $ts_n < h_l(ts_n)$, and the converse for $t > \theta_l$. Thus, the equation must have a solution λ_l living at the scale s_n.

Lemma 2.19. *In the present setting, it holds*

$$\lambda_l - \tilde{\lambda}_l = O_{v.h.\mathbb{P}}\left(\frac{\|V_{1,n}\|}{s_n} + \sqrt{\frac{s_n}{n}}\right),$$

Proof. Let $S_n^{(0)} = V_n$. Thanks to (2.21), this matrix is is with high probability diagonalizable and the entries of its eigenvectors turn to be approximated, up to a multiplicative error $1 + o_{v.h.\mathbb{P}}(\|C_n\|/\lambda)$ of the ones of $\mathrm{Diag}(\theta_1, \dots, \theta_r)$, which are given by the canonical basis. as a consequence eigenvectors of $S_n^{(0)}$ are approximately orthogonal. Let now $L = \lceil \log(n) \rceil$ and consider the following $r \times r$ matrices:

$$S_n^{(1)} = \sum_{k=0}^{L} \frac{V_{k,n}}{\lambda_l^k},$$

$$S_n^{(2)} = V_{0,n} + \frac{V_{1,n}}{\lambda_l} + \sum_{k=2}^{L} \frac{\mathbb{E}[V_{k,n}]}{\lambda_l^k},$$

$$S_n^{(3)} = \sum_{k=0}^{L} \frac{\mathbb{E}[V_{k,n}]}{\tilde{\lambda}_l^k}.$$

It is not difficult to see that for $\ell = 1, 2, 3$, the same diagonal approximation holds and $S_n^{(\ell)}$ is a random perturbation of the matrix $S_n^{(\ell-1)}$. Then it is possible to apply sequentially Theorem 2.10 with the choices $H_n^{(\ell)} = S_n^{(\ell)} - S_n^{(\ell-1)}$ and get that with high probability

$$|\lambda_l(S_n^{(\ell)}) - \lambda_l(S_n^{(\ell-1)})| \leq \|P_n^{(\ell)}\| \|(P_n^{(\ell)})^{-1}\| \|H_n^{(\ell)}\|,$$

where $P_n^{(\ell)}$ has as columns the (unit) eigenvectors of $S_n^{(\ell-1)}$. Because of the bound

$$\|P_n^{(\ell)}\| \|(P_n^{(\ell)})^{-1}\| \leq \frac{2}{|\det(P_n^{(\ell)})|} \left(\frac{\|P_n^{(\ell)}\|_F}{\sqrt{n}}\right)^n \lesssim 2,$$

(which comes from Guggenheimer et al. [1995]) we get that

$$|\lambda_l - \tilde{\lambda}_l| = |\lambda_l(S_n^{(0)}) - \lambda_l(S_n^{(3)})| \leq 2(\|H_n^{(1)}\| + \|H_n^{(2)}\| + \|H_n^{(3)}\|).$$

Hence, it is sufficient to bound the l.h.s. to prove the thesis. $\|H_n^{(1)}\|$ is bounded in the same way as M_1 was bounded in the rank-one case. To bound $\|H_n^{(2)}\|$, it is sufficient to observe that

$$\|H_n^{(2)}\| \leq \sum_{k=2}^{L} \|V_{k,n} - \mathbb{E}[V_{k,n}]\| \leq K_5 \max_{i,j\leq r} \sum_{k=2}^{L} |V_{k,n}(i,j) - \mathbb{E}[V_{k,n}(i,j)]|,$$

and then employ Lemma 2.25 to bound the terms in the r.h.s. of the above display, uniformly in i, j. Finally, to bound $\|H_n^{(3)}\|$, notice that

$$\|H_n^{(3)}\| = \left\| \frac{V_{1,n}}{\lambda_l} + \sum_{k=2}^{L} \mathbb{E}[V_{k,n}] \left(\frac{1}{\lambda_l^k} - \frac{1}{\tilde{\lambda}_l^k} \right) \right\|. \tag{2.23}$$

Reasoning as in the proof of Lemma 2.16, we can bound the r.h.s. of (2.23) by

$$\left\|\lambda_l^{-1} V_{1,n}\right\| + |\lambda_l - \tilde{\lambda}_l| \sum_{k=2}^{L} \|\mathbb{E}[V_{k,n}]\| \, \frac{\sum_{j=1}^{k-1} \tilde{\lambda}_l^j \lambda^{k-j}}{\lambda^k \tilde{\lambda}_l^k},$$

which in the fashion of (2.16) and thanks to Theorem 2.7, needed to estimate the first term, implies that

$$\|H_n^{(3)}\| = O_{v.h.\mathbb{P}}\left(\frac{\|V_{1,n}\|}{s_n} \right) + |\lambda_l - \tilde{\lambda}_l| O_{v.h.\mathbb{P}}({s_n}^{-1}).$$

Putting all estimates together we get,

$$|\lambda_l - \tilde{\lambda}_l|(1 + O_{v.h.\mathbb{P}}(s_n^{-1})) = O_{v.h.\mathbb{P}}\left(\frac{\|V_{1,n}\|}{s_n} + \sqrt{\frac{s_n}{n}} \right),$$

which concludes the proof of Lemma 2.19. □

Finally, we can refine the previous result to the following one, which is analogous to Lemma 2.16.

Lemma 2.20. *In the present setting, it holds*

$$\lambda_l - \tilde{\lambda}_l = s_n \theta_l \frac{(v_l^-)^T C_n v_l^+}{\tilde{\lambda}_l} + o_{v.h.\mathbb{P}}\left(\frac{\|V_{1,n}\|}{s_n} + \sqrt{\frac{s_n}{n}} \right).$$

Proof. Let us apply again the Bauer-Fike approach with the choice, for the third step,

$$\widetilde{S}_n^{(3)} = \frac{V_{1,n}}{\lambda_l} + \sum_{k=0}^{L} \frac{\mathbb{E}[V_{k,n}]}{\tilde{\lambda}_l^k} = \frac{V_{1,n}}{\lambda_l} + S_n^{(3)},$$

where $S_n^{(3)}$ as in the previous proof. We get that

$$\left|\lambda_l - \lambda_l\left(\frac{V_{1,n}}{\lambda_l} + \sum_{k=0}^{L}\frac{\mathbb{E}[V_{k,n}]}{\tilde{\lambda}_l^k}\right)\right| = \left|\lambda_l(S_n^{(0)}) - \lambda_l(\widetilde{S}_n^{(3)})\right|$$
$$= o_{v.h.\mathbb{P}}(|\lambda_l - \tilde{\lambda}_l|) = o_{v.h.\mathbb{P}}\left(\frac{\|V_{1,n}\|}{s_n} + \sqrt{\frac{s_n}{n}}\right),$$

where the first asymptotic estimate can be obtained reasoning as in (2.16), and the second one follows from Lemma 2.19. Let us now consider the matrices

$$\tilde{H} := \widetilde{S}_n^{(3)} - \widetilde{S}_n^{(3)}(l,l)I_n,$$
$$\tilde{M} := \widetilde{S}_n^{(3)} - \frac{V_{1,n}}{\lambda_l} - \left(\widetilde{S}_n^{(3)}(l,l) - \frac{V_{1,n}(l,l)}{\lambda_l}\right)I_n,$$

obtained adding and subtracting to $\widetilde{S}_n^{(3)}$ and $\widetilde{S}_n^{(3)} - \frac{V_{1,n}}{\lambda_l}$ multiples of the identity (we highlight that $V_{1,n}(l,l) = s_n\theta_l(v_l^-)^T C_n v_l^+$ is the (l,l) entry of $V_{1,n}$). Since this just translate eigenvalues, it follows

$$\lambda_l\big(\widetilde{S}_n^{(3)}\big) = \lambda_l(\tilde{H}) + \widetilde{S}_n^{(3)}(l,l) = \lambda_l(\tilde{H}) + \frac{V_{1,n}(l,l)}{\lambda_l} + \lambda_l\left(\widetilde{S}_n^{(3)} - \frac{V_{1,n}}{\lambda_l}\right) - \lambda_l(\tilde{M}),$$

which means, recalling that $\widetilde{S}_n^{(3)} - \frac{V_{1,n}}{\lambda_l} = S_n^{(3)}$ and $\lambda_l\big(S_n^{(3)}\big) = \tilde{\lambda}_l$,

$$\lambda_l\left(\frac{V_{1,n}}{\lambda_l} + \sum_{k=0}^{L}\frac{\mathbb{E}[V_{k,n}]}{\tilde{\lambda}_l^k}\right) = \tilde{\lambda}_l + \frac{V_{1,n}(l,l)}{\lambda_l} + \lambda_l(\tilde{H}) - \lambda_l(\tilde{M}).$$

To conclude the proof of the lemma, we need to show

$$|\lambda_l(\tilde{H}) - \lambda_l(\tilde{M})| = o_{v.h.\mathbb{P}}\left(\frac{\|V_{1,n}\|}{s_n}\right).$$

This follows reasoning as in the proof of [Chakrabarty et al., 2020, Lemma 5.8], where the same kind of estimate is shown, after having noticed that symmetry is not used. □

The obvious analogous of Lemma 2.17, and some computations as in (2.20), provide what remains to prove Theorem 2.9.

2.4 Inequalities and lemmata

2.4.1 Concentration results

We recall that, being D_x^+ a sum of Bernoulli random variables, the following inequalities hold (see for example [van der Hofstad, 2016, Prop. 2.21]):

$$\mathbb{P}(D_x^+ \geq \mathbb{E}[D_x^+] + t) \leq \exp\left(-\frac{t^2}{2(\mathbb{E}[D_x^+] + t/3)}\right), \tag{2.24}$$

$$\mathbb{P}(D_x^+ \leq \mathbb{E}[D_x^+] - t) \leq \exp\left(-\frac{t^2}{2\mathbb{E}[D_x^+]}\right). \tag{2.25}$$

Choosing $t = s_n^{\frac{2}{3}}$ in (2.24) and (2.25), thanks to due to Assumption 2.4, we get for example

$$\mathbb{P}(\max_{x\in V} |D_x^+ - w_x^+ s_n| \geq s_n^{\frac{2}{3}}) \leq 2n \exp\left(-s_n^{\frac{1}{3}}\right) = o\left(e^{-\log(n)^\eta}\right),$$

for some $\eta > \frac{4}{3}$. Choosing $t = m\mathbb{E}[D_x^+]$ in (2.24) and (2.25), we get

$$\mathbb{P}\left(\frac{1}{D_x^+} \geq \frac{1}{\mathbb{E}[D_x^+](1-m)}\right) \leq \exp\left(-\frac{m^2\mathbb{E}[D_x^+]}{2}\right),$$
$$\mathbb{P}\left(\frac{1}{D_x^+} \leq \frac{1}{\mathbb{E}[D_x^+](1+m)}\right) \leq \exp\left(-\frac{m^2\mathbb{E}[D_x^+]}{2(1+m/3)}\right).$$

Taking $m = s_n^{-\frac{1}{3}}$, since $(1 \pm m)^{-1} = 1 \mp m + o(m)$, the following lemma holds.

Lemma 2.21. *There exists $\eta > \frac{4}{3}$ such that,*

$$\mathbb{P}\left(\max_{x\in V}\left|\frac{w_x^+ s_n}{D_x^+} - 1\right| \geq 2s_n^{-\frac{1}{3}}\right) \leq 2n\exp\left(-\frac{cs_n^{\frac{1}{3}}}{3}\right) \leq \exp(-\log(n)^\eta).$$

2.4.2 Useful lemmata

Lemma 2.22. $\mathbb{V}\mathrm{ar}((v^-)^T C_n v^+) = O(s_n/n)$.

Proof. By direct computation

$$\mathbb{V}\mathrm{ar}((v^-)^T C_n v^+) = \mathbb{V}\mathrm{ar}((v^-)^T A v^+) = \sum_{x,y} \frac{(w_x^-)^2(w_y^+)^2}{\boldsymbol{w}^2} p_{x,y}(1-p_{x,y}) = O\left(s_n \frac{n^2}{\boldsymbol{w}^3}\right).$$

□

Lemma 2.23. $|(v^-)^T C_n v^+| = O_{\mathbb{P}}\left(\sqrt{\frac{s_n}{n}}\right)$.

Proof. It holds

$$\mathbb{E}[|(v^-)^T C_n v^+|] \leq \sqrt{\mathbb{V}\mathrm{ar}((v^-)^T C_n v^+)} = O\left(\sqrt{\frac{s_n}{n}}\right).$$

□

Lemma 2.24. *There exists a constant $K_1 < +\infty$ such that, for $2 \leq k \leq L$,*

$$\left|\mathbb{E}\left[(v^-)^T C_n^k v^+\right]\right| \leq (K_1 s_n)^{k/2}.$$

Moreover,

$$(v^-)^T C_n v^+ = o_{v.h.\mathbb{P}}(s_n). \tag{2.26}$$

Proof. By Proposition 2.12, there exists $K > 0$ and $\eta > 1$ such that for the event $\mathcal{A} := \{\|C_n\| \le C\sqrt{s_n}\}$ it holds $\mathbb{P}(\mathcal{A}^\complement) \le e^{-(\log(n))^\eta}$. Then

$$\left|\mathbb{E}\left[(v^-)^T C_n^k v^+\right]\right| \le \left|\mathbb{E}\left[(v^-)^T C_n^k v^+ \mathbf{1}_{\mathcal{A}}\right]\right| + \left|\mathbb{E}\left[(v^-)^T C_n^k v^+ \mathbf{1}_{\mathcal{A}^\complement}\right]\right|.$$

The first addend can be estimated by $|\mathbb{E}\|v^-\|\|v^+\|\|C_n\|| = K_1 s_n^{k/2}$. For the second addend, noticing that $\left((v^-)^T C_n^k v^+\right)^2$ cannot exceed a power of n, say $n^{kC'}$, for $C' > 0$, by Cauchy–Schwartz,

$$\begin{aligned}\left|\mathbb{E}\left[(v^-)^T C_n^k v^+ \mathbf{1}_{\mathcal{A}^\complement}\right]\right| &\le \left(\mathbb{E}\left[\left((v^-)^T C_n^k v^+\right)^2\right]\right)^{\frac{1}{2}} \mathbb{P}\left(\mathcal{A}^\complement\right)^{\frac{1}{2}} \\ &\le n^{kC'/2} e^{-(\log(n))^\eta/2} = o(1).\end{aligned}$$

To prove (2.26), recalling that $C_n = A_n - \mathbb{E}[A_n]$, Hoeffding inequality can be employed to have that, for every $\varepsilon > 0$, it holds

$$\begin{aligned}\mathbb{P}\left(\left|\sum_{x,y\in[n]} v_x^- A_{xy} v_y^+ - \sum_{x,y\in[n]} \mathbb{E}\left[v_x^- A_{xy} v_y^+\right]\right| > \varepsilon s_n\right) \\ \le 2\exp\left(-\frac{2\varepsilon^2 s_n^2}{n^2((\max_x w_x^\pm - \min_x w_x^\pm)/\boldsymbol{w})^2}\right).\end{aligned}$$

Since weights are bounded, the r.h.s. is at most $O(\exp(-2\varepsilon^2(\log(n))^{2\xi}))$. □

Lemma 2.25. *There exists $\eta > 1$ such that*

$$\max_{2\le k\le L} \mathbb{P}\left(\left|(v^-)^T C_n^k v^+ - \mathbb{E}[(v^-)^T C_n^k v^+]\right| > s_n^{k/2} n^{-1/2} \log(n)^{k\xi/4}\right) = O\left(e^{-(\log(n))^\eta}\right).$$

Proof. The proof of this result comes by showing that it holds

$$\mathbb{E}[\left|(v^-)^T\left(C_n^k - \mathbb{E}[C_n^k]\right)v^+\right|^p] < (K_3 kp)^{kp} s_n^{\frac{kp}{2}},$$

where $p := \frac{\log(n)^\eta}{K_3 k}$. This high moment estimate is obtained adapting [Erdős et al., 2013, Lemma 6.5] to the inhomogeneous setting as in [Chakrabarty et al., 2020, Lemma 4.3] and observing that in our non–reversible setting the entries of the matrix are truly independent, so that there is no need to decompose C_n in a sum of a upper and lower triangular matrix. □

2.5 A conjecture on the bulk

In this section, we present a non-rigorous discussion on the spectrum of the Chung–Lu digraph. In particular, we present and motivate a conjecture regarding the limiting empirical spectral distribution (ESD) of its adjacency matrix. We recall that the ESD of a $n \times n$ matrix M_n is defined as

$$\mu_{M_n} = \frac{1}{n}\sum_{i=1}^n \delta_{\lambda_i(M_n)}.$$

The circular law. As mentioned, the eigenvalues of non-Hermitian (not symmetric) matrices spread out in the complex plane. In the 1950s, matrices of size n and with i.i.d. centered entries with variance $1/n$, have been conjectured to satisfy a limiting circular law, meaning that the eigenvalues become uniformly distributed in the unit disk, as n grows (see the survey Bordenave and Chafaï [2012]). After many other contributions, in Tao and Vu [2010], the circular law was established for $\frac{1}{\sqrt{n}}B_n$, where B_n is a dense $n \times n$ matrix with i.i.d. entries and bounded second moments.

The argument of Tao and Vu [2010] is articulated as follows. The first ingredient is a general replacement principle that ensures that the limiting ESD does not depend on the specific law of matrix entries. Since the circular law for Gaussian matrices was proved in Mehta [1967], this leads to a universal statement. The replacement principle can be applied provided that the function $\log(\cdot)$ is integrable w.r.t. the empirical singular value distribution of the shifted matrices $\frac{1}{\sqrt{n}}B_n - zI_n$, for $z \in \mathbb{C}$. To prove it, a bound on the smallest singular value of the matrix is needed, together with a control on the cumulative contribution of small singular values. Finally, the limit for the empirical singular value distribution of $\frac{1}{\sqrt{n}}B_n - zI_n$ is studied.

In a sequence of works, the analysis has also been extended to sparser matrices with entries of the form $\delta_{xy}X_{xy}$, where X_{xy} are i.i.d. centered random variables with unit variance, and δ_{xy} are i.i.d. Bernoulli random variables with parameter $p_n \ll 1$ such that $np_n \gg 1$ Basak and Rudelson [2019], Götze and Tikhomirov [2010], Rudelson and Tikhomirov [2019], Tao and Vu [2008]. The sparse case, with p_n of order $1/n$, has been recently considered, and it has been shown that, in that case, an atom at the origin arises Sah et al. [2025].

Beyond the i.i.d. case, in Tao and Vu [2008] and Tao and Vu [2010], matrices whose entries are independent but not identically distributed were also considered, extending part of the estimates of least singular values (and the consequent universal results) to an inhomogeneous setting. Moreover, matrices with various dependency structures or variance profiles often still exhibit a circular-type limiting distribution. For example, when row sums are constrained (random stochastic matrices), the eigenvalue cloud remains circular under suitable conditions, as shown in Bordenave et al. [2012]. In Bordenave et al. [2014] the limiting ESD of the random walk generator of a sparse directed random graph is characterized. In Litvak et al. [2021] the circular law is proved for the adjacency matrix of a sparse regular random directed graph.

The framework that we considered in the previous sections corresponds to a non-centered, inhomogeneous and (weakly) sparse setting where average degrees scale as $\log(n)^4 \ll s_n = p_n n \ll n$. Here the determination of the limiting behavior of the ESD for some rescaling of the matrix is a difficult and technical problem. Surprisingly, in the aforementioned work Basak and Rudelson [2019], the techniques from Tao and Vu [2010] could be employed and extended to a matrix with non i.i.d. and non-centered entries. In particular, the circular law has been shown to hold for the adjacency matrix of a sparse directed Erdős–Rényi graph with connection probability $p_n \gg \log(n)^2/n$, as the following theorem states.

Theorem 2.26 ([Basak and Rudelson, 2019, Thm. 1.7])**.** *Let A_n be the adjacency matrix of a directed Erdős–Rényi graph, with connection probability $p_n \in (0,1)$. It $\bar{p}_n = \min\{p_n, 1-p_n\}$ and $n\bar{p}_n \gg \log(n)^2$ then, as $n \to +\infty$, the ESD of the rescaled matrix $(np_n(1-p_n))^{-\frac{1}{2}} A_n$ converges weakly in probability to the circular distribution. Moreover, there exists $c > 0$ such that, if $n\bar{p}_n > \exp(c\sqrt{\log(n)})$ the convergence is almost sure.*

Remark 2.27. *The assumption $np_n \geq \log(n)^2$, implied by Assumption 2.4 has been shown to be technical, and in Rudelson and Tikhomirov [2019] the circular law has been proved to hold whenever np_n diverges. In this more general setting, the matrix has many zero row and columns with constant (non-zero) probability. This requires deeper analysis than the one in Basak and Rudelson [2019].*

Notice that their construction of the Erdős–Rényi digraph is slightly different, since it does not allow self-loops. We are convinced that the approach developed in Tao and Vu [2010] and Basak and Rudelson [2019] to study small singular values can be extended to our random matrices. For this reason, following Theorem 2.26, we conjecture that the limiting ESD of $(s_n)^{-1/2} A_n$ is equal to the one of the random matrix $n^{-1/2} \bar{D}_n^+ Y_n \bar{D}_n^-$, where Y_n is a $n \times n$ i.i.d. standard Gaussian array and $\bar{D}_n^{\pm}$ are the $n \times n$ diagonal matrices containing the vectors $(\frac{n}{\boldsymbol{w}})^{-1/4} (w^{\pm})^{-1/2}$. Notice that this choice centers the entries and modifies their variance from $p_{x,y}(1-p_{x,y})$ to the asymptotically equivalent $p_{x,y}$.

Free probability. Before stating our precise conjecture, we introduce the notion of non-commutative probability space and free independence, which will provide the necessary notation. See Mingo and Speicher [2017] for an introduction to the topic.

Definition 2.28 (Non-commutative probability space)**.** *A pair $(\mathcal{A}, \varphi)$ consisting of a unital algebra $\mathcal{A}$ and a linear functional $\varphi : \mathcal{A} \to \mathbb{C}$ with $\varphi(1) = 1$ is said to be a non-commutative probability space. An element of $a \in \mathcal{A}$ is called random variable, and a probability distribution μ_a is said to be the law of a if it is the unique probability distribution μ_a on $\mathbb{R}$ such that, for any $k \in \mathbb{N}$,*

$$\int_{\mathbb{R}} t^k \, d\mu_a(t) = \varphi(a^k).$$

Definition 2.29 (Free and asymptotically free independence)**.** *Let $(\mathcal{A}, \varphi)$ be a non-commutative probability space. Given a family $(a_i)_{i\in I} \subseteq \mathcal{A}$, its elements are said to be free (or freely independent) if it holds*

$$\varphi\big((a_{i_1} - \varphi(a_{i_1})1)(a_{i_2} - \varphi(a_{i_2})1) \cdots (a_{i_k} - \varphi(a_{i_k})1)\big) = 0, , \qquad \forall k \in \mathbb{N},\ \forall i_1, i_2, \ldots, i_k \in I.$$

Given a family of $n \times n$ random matrices $(M_{n,i})_{i\in I}$, its elements are said to be asymptotically free if there exists a non-commutative probability space $(\mathcal{A}, \varphi)$ and a family of free operators $(a_i)_{i\in I} \subseteq \mathcal{A}$ such that

$$\lim_{n\to+\infty} \frac{1}{n} \operatorname{Tr}(M_{n,i_1} M_{n,i_2} \cdots M_{n,i_k}) = \varphi(a_{i_1} a_{i_2} \cdots a_{i_k}), \qquad \forall k \in \mathbb{N},\ \forall i_1, i_2, \ldots, i_k \in I.$$

Definition 2.30 (Free sums and products). *If $M_{n,1}$ and $M_{n,2}$ are asymptotically free random matrices with limiting ESD μ_1 and μ_2 respectively, and a_1 and a_2 denote the corresponding free operators in a non-commutative probability space $(\mathcal{A}, \varphi)$, the free additive convolution $\mu_1 \boxplus \mu_2$ of μ_1 and μ_2 is defined as the law of $a_1 + a_2$. Analogously, the free multiplicative convolution $\mu_1 \boxtimes \mu_2$, is defined as the law of $a_1 a_2$.*

The conjecture. Given a random matrix B_n, consider, for $z \in \mathbb{C}$, the symmetrized matrix

$$\begin{pmatrix} 0 & B_n - zI_n \\ B_n^* - \bar{z}I_n & 0 \end{pmatrix}, \tag{2.27}$$

and call $\theta_{B_n,z}$ its ESD. It has been shown (see [Bordenave and Chafaï, 2012, Lemma 4.3] and [Kösters and Tikhomirov, 2018, Theorem 2.1]) that, under suitable hypotheses, the determination of a limit θ_z for the $\theta_{B_n,z}$, for every $z \in \mathbb{C}$, is sufficient to reconstruct the limit of μ_{B_n}. It will be indeed given by the unique probability distribution μ such that, for every $z \in \mathbb{C}$,

$$\int \log|\zeta - z| \, d\mu(\zeta) = \int \log|x| d\theta_z(x).$$

We can prove the following result, which can be used to determine θ_z in our setting.

Proposition 2.31. *Let $\rho_{\sqrt{w^\pm}}$ denote the weak limits of the empirical distributions*

$$\frac{1}{n} \sum_{x \in [n]} \delta_{\sqrt[4]{\frac{n}{\boldsymbol{w}}} \sqrt{w_x^\pm}}.$$

If $\rho_{\sqrt{w^+}} = \rho_{\sqrt{w^-}} = \rho_{\sqrt{w}}$, then, for every $z \in \mathbb{C}$, the matrices

$$\begin{pmatrix} \bar{D}_n^+ & 0 \\ 0 & \bar{D}_n^- \end{pmatrix}, \qquad \begin{pmatrix} 0 & \frac{1}{\sqrt{n}} Y_n \\ \frac{1}{\sqrt{n}} Y_n^* & 0 \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} 0 & -zI_n \\ -\bar{z}I_n & 0 \end{pmatrix},$$

are asymptotically free.

For the sake of brevity, we omit the proof of this proposition, which is based on the bi-unitary invariance of Gaussian matrices, and we refer to [Götze et al., 2015, Proposition 5.8] for further details. We only recall that a random matrix M_n is bi-unitary invariant if the joint distribution of the entries of M_n and those of $U_n M_n V_n$ are the same for every choice of unitary matrices U_n and V_n.
The symmetrized version (2.27) of $n^{-1/2} \bar{D}_n^+ Y_n \bar{D}_n^-$ is given by

$$\begin{pmatrix} \bar{D}_n^+ & 0 \\ 0 & \bar{D}_n^- \end{pmatrix} \begin{pmatrix} 0 & n^{-1/2} Y_n \\ n^{-1/2} Y_n^* & 0 \end{pmatrix} \begin{pmatrix} \bar{D}_n^+ & 0 \\ 0 & \bar{D}_n^- \end{pmatrix} + \begin{pmatrix} 0 & -zI_n \\ -\bar{z}I_n & 0 \end{pmatrix}.$$

Moreover, the symmetrized version of $n^{-\frac{1}{2}} Y_n$ has a semicircular limit (by standard facts on Wishart matrices, see again Mingo and Speicher [2017]), and the one of $-zI_n$ has eigenvalues $\pm|z|$. Then, we are finally ready to state the conjecture, which corresponds to a non-Hermitian counterpart of Proposition 5.2 in Chakrabarty et al. [2021].

Conjecture 2.32. *Let $\rho_{\sqrt{w^\pm}}$ denote the weak limits of the empirical distributions*

$$\frac{1}{n}\sum_{x\in[n]}\delta_{\sqrt[4]{\frac{n}{\boldsymbol{w}}}\sqrt{w_x^\pm}}.$$

If $\rho_{\sqrt{w^+}} = \rho_{\sqrt{w^-}} = \rho_{\sqrt{w}}$, then $\mu_{\frac{1}{\sqrt{s_n}}A_n}$ converges weakly in probability to a limit μ, which is the unique probability measure on $\mathbb{C}$ such that, for all $z \in \mathbb{C}$,

$$\int_{\mathbb{C}} \log|\zeta - z|\, d\mu(\zeta) = \int_{\mathbb{R}} \log|x|\; d\Big[\rho_{\sqrt{w}} \boxtimes s \boxtimes \rho_{\sqrt{w}} \boxplus \tfrac{1}{2}\left(\delta_{-|z|} + \delta_{|z|}\right)\Big](x)\,,$$

where s is the semicircular distribution with density $ds(x) = \frac{1}{2\pi}\sqrt{4-x^2}\mathbf{1}_{[-2,2]}(x)dx$, while $\boxplus$ and $\boxtimes$ respectively denote free additive and multiplicative convolutions of probability distributions (see Def. 2.30).

Remark 2.33. *The probability distributions $\rho_{\sqrt{w^+}}$ and $\rho_{\sqrt{w^-}}$ are well defined. Indeed, given the convergence in Assumption 2.1, it is possible to deduce the convergence of other empirical distributions related to the bi-weight distribution $(w_x^+, w_x^-)_{x\in[n]}$. For instance, projecting onto the first or second component, we obtain the two marginals, and taking square roots, by the continuous mapping theorem, we get that the distributions $\rho_{\sqrt{w^\pm}}$ can be expressed as a push-forward of ρ. In the homogeneous case, we recover the circular law in Theorem 2.26.*

Part II

Mixing time of random walks on random directed graphs

Introduction to Part II

The study of stochastic processes evolving on a random structure is fundamental for the understanding of many real-world systems that are characterized by intrinsic randomness. The latter can be due to the presence of many microscopical forces, noisy channels, impurity of materials and media, or the unpredictability of human behavior. Random graphs endow with a source of randomness the interaction networks where agents can flow, exchange information and update according to probabilistic rules, so that randomness is encoded in two steps. The first step of randomness is given by the environment, whose associated probability measure is denoted by $\mathbb{P}$. The second step is given by the dynamics, which depends on the specific instance of G. We will denote by $\mathbf{P}^G$ the quenched distribution associated with the dynamics. Given the discrete nature of the state space, the simplest and probably most effective way of describing some random dynamics is a Markov chain. Let us briefly mention that a different but related line of research is related to random walks in random environment, meaning random walks on a deterministic lattice, which evolve according to random transition probabilities. These are in turn given by weights, which form the environment. See Zeitouni [2004].

Markov chains on random graphs

The state space for the Markov dynamics can be chosen to be the vertex set of the graph or something more complex, such as the set of binary spin configurations supported on the graph. This choice has been particularly fruitful in the context of mathematical statistical physics (see Durrett [2007], van der Hofstad [2025]). Classical spin systems are defined on lattices and the most studied models in this class, introduced, e.g., in Liggett [1985, 1999], are the stochastic Ising model, the contact process and the voter model. The dynamics is usually given by exponential clocks attached to each particle (or spin), and the state of the spin is modified, according to some random rule, as the clock rings. The main questions are concerned with the characterization of the evolution towards the equilibrium, the determination of metastable timescales, phase transitions and, for the finite random graph formulation, understanding the link with the model on infinite and deterministic trees. We will now mention some contributions in this area, with no aim of being exhaustive, since the literature on this topic is extremely vast.

For the Ising model, introduced to study ferro-magnetic interactions, the update rule is given through an acceptance-rejection scheme, depending on the energy gain of the transition. Besides contributions in the analysis of the static model, such as Dembo and Montanari [2010], we mention Dommers [2017], Dommers et al. [2017], Bovier et al. [2021] and den Hollander and Jovanovski [2021] for achievements on sparse graphs.

In the contact process, which describes the evolution of an infectious disease, active vertices spread an infection and recover from it at certain rates. We mention Mourrat and Valesin [2016] and Bhamidi et al. [2021], where phase transitions for the model on

sparse random graphs are determined, and Chatterjee and Durrett [2009], Mountford et al. [2013], for achievements in a power-law setting.
In the voter model, which describes opinion dynamics, the selected vertex updates its state adopting the one of a random neighbor. The quantity of interest is the consensus time at which all spins are aligned. In Cooper et al. [2009/10] a sparse regular graph is studied, while in Fernley and Ortgiese [2023] and Hermon et al. [2022] more general models are considered. See Avena et al. [2024] and Capannoli [2025] for recent achievements in the directed setting, which is of our interest.

Random walks represent a simple yet interesting model for diffusion phenomena and real-world applications (e.g., the Web-indexing PageRank algorithm Page et al. [1999]). Moreover, they constitute a fundamental brick before passing to more complicated dynamics. In this regard, it is worth mentioning that, under suitable hypotheses, the coalescing time of a system of independent random walks can be used to determine a first-order asymptotics for the consensus time in the voter model, as established in Oliveira [2012]. Indeed, the graphical representation of the model reveals that its dual is given by a system of coalescing walkers. Quantities of interest for random walks are, e.g., cover, meeting and mixing times. Part II will be devoted to the mixing time, which we are going to introduce in the following paragraph. As it will be soon clear, its asymptotic behavior can undergo a phase transition, called *cutoff*.

Mixing times of Markov chains

Given a finite state space Markov chain it is classical that, under irreducibility conditions, there exists a unique equilibrium distribution π, independent of the initial state, towards which the dynamics evolves. In this framework a fundamental question concerns the time needed to converge to the equilibrium, namely the mixing time. Given a discrete time Markov chain with t-step transition kernel $P^t(\cdot,\cdot)$, and a distance $d(\cdot,\cdot)$ for probability distributions, for $\varepsilon > 0$, the ε-mixing time is defined as

$$\boldsymbol{t}_{\mathrm{mix}}(\varepsilon) \coloneqq \inf\{t > 0 : \max_{x\in[n]} d(P^t(x,\cdot),\pi) \leq \varepsilon\},$$

that is the minimal time at which the distribution of the chain is ε-near to the equilibrium. If the distance is convex, the maximum can be taken over all initial distributions. An interesting case is given by the total variation distance $\|\cdot\|_{\mathrm{TV}}$, which corresponds to a L^1 distance. Besides a theoretical interest, providing estimates for the mixing time can be very useful from an applicative perspective, motivated for instance, by Monte Carlo simulations.

Estimates on the mixing time can be derived in several ways. The relaxation time of the chain, given by the inverse of the second least eigenvalue of the infinitesimal generator of the chain can be used to derive exponential contraction of the total variation distance and provide an upper bound on the mixing time. In discrete time, the generator is given by $I - P$, where P denotes the transition matrix. In this setting, the quantity of interest is instead the least eigenvalue modulus, up to pass to the lazy version the chain with transition matrix $(P + I)/2$. Other approaches include the estimate Cheeger constants and logarithmic Sobolev inequalities. See Levin and Peres [2017] for an introduction on the topic. All these techniques work well in the reversible setting, where the state space of the process can be endowed with a Hilbert space structure. This is not the case for non-reversible Chains, even though recent advances show that it is possible to define a relaxation time in such a way that certain bounds are preserved Chatterjee [2025].

There is a strong interest in studying how the distance to equilibrium decays in time and, in particular, in characterizing the precise shape of its graph as the size of the system grows. I,f under proper rescaling of time, a limit object exists, it is referred to as a limit profile. A remarkable limit profile is achieved when the process has a fast mixing. A process is said to exhibit cutoff if for every $\varepsilon > 0$,

$$\lim_{n\to+\infty} \frac{\boldsymbol{t}_{\mathrm{mix}}(\varepsilon)}{\boldsymbol{t}_{\mathrm{mix}}(1-\varepsilon)} = 1.$$

In a few words, this means that the decay of its distance to equilibrium takes place in an abrupt manner: there exists some t (which depends on the size of the system) such that, for any $\varepsilon > 0$, at time $(1-\varepsilon)t$ its distribution is arbitrarily far from the equilibrium one, while at time $(1+\varepsilon)t$ it is arbitrarily near. In this sense, the "parameter" inducing the phase transition here is *time*, and for this reason it is often said that this is an example of a *dynamical* phase transition. This limit behavior was studied in the context of random transpositions by Diaconis and Shahshahani [1981] and for random walks on the hypercube by Aldous [1983], and later again by Aldous and Diaconis [1986], Diaconis [1996], and Diaconis and Saloff-Coste [1996] in the context of card shuffling.

More recently, this topic has received a renovated interest and the occurrence of the cutoff has been proved for several models. For what concerns random walks on random graphs, to which the following sections will be devoted, we first refer to the seminal work Lubetzky and Sly [2010]. We also mention birth and death chains Ding et al. [2010], the Ising model on the lattice Lubetzky and Sly [2013], the exclusion process on segment and circles Lacoin [2016a,b], and the averaging process on complete and bipartite graphs Chatterjee et al. [2022], Caputo et al. [2023]. Despite an increasing amount of work on the subject, the cutoff phenomenon is still far from being completely understood, and the research of simple conditions (i.e., easy-to-check and model independent) guaranteeing the presence of a cutoff is very active. Remarkable advances have been obtained for non-negatively curved Markov Chains employing functional inequalities and entropy criteria. See Salez [2024a,b], Pedrotti and Salez [2025] and references therein.

In the second part of the dissertation we turn to the analysis of dynamics on a random graph. Specifically in Chapters 3 and 4 we will study the mixing time of the simple random walk on inhomogeneous random directed graphs.

Cutoff for random walks on random graphs

As mentioned, random walks on random graph nowadays represent a prototypical example of Markov chains that exhibit the cutoff phenomenon. In recent years, random walks on random graphs have been extensively studied on various random graph models.

Many attempts have been done to characterize the mixing time of this dynamics. See, e.g., Benjamini et al. [2014] (initially submitted in 2006), Fountoulakis and Reed [2008], and references therein. Notice that, in this setting the mixing time and the total variation profile are random objects, being observables depending on the graph, and it is crucial to understand the relation between environment and dynamics. We again mention the work Lubetzky and Sly [2010], which constituted a breakthrough in the field. Other notable contributions include the establishment of the cutoff for random walks on the giant component of the Erdös-Rényi graph Berestycki et al. [2018], on the configuration model Ben-Hamou and Salez [2017], Ben-Hamou et al. [2019], on all Ramanujan graphs Lubetzky and Peres [2016], and on random lifts Bordenave and Lacoin [2022], Conchon-Kerjan [2022].

Part of these investigations has also focused on the directed setting, which, as mentioned, is particularly challenging due to the non-reversibility of the dynamics and the poor knowledge of the stationary distribution. In this setting, the stationary distribution of the Markov chain is not explicit and its characterization represents itself an and important theoretical challenge (see, e.g., Caputo and Quattropani [2020], Chen et al. [2014, 2017], Garavaglia et al. [2020]). This framework was explored in Bordenave et al. [2018, 2019], where the cutoff was established for random walks on the directed configuration model, and later extended to PageRank dynamics Caputo and Quattropani [2021a] and to the case of heavy-tailed degrees Cai et al. [2023]. Other results in the same spirit have also been obtained in Dubail [2024a,b]. A common thread in all these works is the characterization of the mixing time in terms of the entropy production rate (or simply, the *entropy*) of the random walk on its *local weak limit*. The cutoff, and hence mixing, time is then shown to given by logarithm of the size of the system, normalized by the entropy. This quantity is called entropic time.

Overview of chapter 3: Cutoff for the SRW on the directed Chung–Lu graph

Chapter 3 contains the result of Bianchi and Passuello [2025].
Here we analyze the motion of a random walk on the Chung–Lu directed graph. In Section 3.1 we introduce the model and state the results. Subsection 3.1.1 is dedicated to the definition of the Chung–Lu directed graph in terms of in- and out-weights. Then we introduce the discrete time simple random walk, namely the Markov chain $(X_t)_{t\in\mathbb{N}}$, with transition matrix

$$P(x,y) := \begin{cases} \frac{1}{D_x^+} & \text{if } x \to y \\ 0 & \text{otherwise} \end{cases}, \qquad \forall x, y \in [n],$$

where D_x^+ denotes the out-degree of the vertex $x \in [n]$, and the notation $x \to y$ means that the oriented edge (x, y) is in the graph, and we provide some comments on the existence of its invariant distribution π.

As mentioned above, an important quantity to describe the mixing time of this dynamics is given by the entropy, which we denote by H. In Subsection 3.1.3 we define it properly in terms of the expectation of a random variable (see (3.10)), with variance σ^2 (see (3.11)), and we provide the main result, Theorem 3.3 which can be summarized as follows: for $\beta > 0$ and $\beta \neq 1$, it holds

$$\max_{x\in[n]} \left| \|\mathbf{P}_x^G(X_{\beta \boldsymbol{t}_{\text{ent}}} \in \cdot) - \pi\|_{\text{TV}} - \mathbf{1}_{\{\beta<1\}} \right| \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

This means that, rescaling time by $\boldsymbol{t}_{\text{ent}}$, the total variation profile approximates a step function, so that cutoff displays. A refinement of this statement holds when adding to the entropic time $\boldsymbol{t}_{\text{ent}}$ a term lying at the scale of the lower order term $\mathbf{w}_{\text{ent}} = \frac{\sigma}{\text{H}}\sqrt{\boldsymbol{t}_{\text{ent}}}$. This is our second result, Theorem 3.5, which states that if the variance σ^2 satisfies a proper non-degeneracy condition (see (3.15)), then, for $\lambda \in \mathbb{R}$ fixed, it holds

$$\max_{x\in[n]} \left| \|\mathbf{P}_x^G(X_{\boldsymbol{t}_{\text{ent}}+\lambda\mathbf{w}_{\text{ent}}} \in \cdot) - \pi\|_{\text{TV}} - \frac{1}{\sqrt{2\pi}} \int_\lambda^{+\infty} e^{-\frac{u^2}{2}}\, du \right| \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

It means that inside a window of size $\mathbf{w}_{\text{ent}}$ the total variation profile takes a smooth Gaussian universal shape, independent of the parameters.

Section 3.2 is devoted to outline the proof. The main idea is to identify some properties, valid for random walk paths of asymptotic length $\boldsymbol{t}_{\text{ent}}$ that hold with high probability and which allow for a fast mixing. Paths satisfying them are called *nice* paths (see Definition 3.7). The main one is related to the mass of a path $\mathfrak{p} = (x_0, \dots, x_t) \in [n]^t$, which is defined as $\mathbf{m}(\mathfrak{p}) = \prod_{s=0}^{t-1} P(x_s, x_{s+1})$.

In Section 3.3 the main tools of the analysis are designed. In Subsection 3.3.1 we describe the distribution of the annealed random walk, the process with law given in terms of the averaged measure $\mathbb{P}^{\text{an}}(\cdot) = \mathbb{E}[\mathbf{P}^G(\cdot)]$. This non-Markovian evolution on the averaged graph will turn to provide a good approximation on the Markovian dynamics, especially for short times $t = o(\sqrt{n})$. The main result of the section is Lemma 3.10, which shows that, on the event that the trajectory visits a new fresh vertex, the distribution of the annealed walk is well described by the a distribution to in-weights. In Subsection 3.3.2 we consider some properties of the graph: degree concentration, which allows to provide the first asymptotics of the entropy (Proposition 3.2), the size of in-neighborhoods (Lemma 3.18), and the shape of out-neighborhoods (Lemma 3.19). A common denominator is that the graph exhibits a locally tree-like structure.

Subsection 3.4 contains the core the analysis, Theorem 3.21. The latter is a concentration result for the mass of random walk paths of length $t = \Theta(\boldsymbol{t}_{\text{ent}})$, whose statement, which reminds the shape of the main theorem, goes as follows. If $\theta \in (0, 1)$ is such that there exists $\rho > 0$, $\rho \neq 1$, satisfying $\log\theta = \rho \text{H} t(1 + o(1))$, then

$$\max_{x\in[n]} \left| \mathbf{P}_x^G\left(\mathbf{m}(X_0, X_1, \dots, X_t) > \theta\right) - \mathbf{1}_{\{\rho>1\}} \right| \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

This result is referred to as quenched law of large numbers, and it is refined to the quenched CLT 3.22 for the window analysis.

Everything is wrapped up in Section 3.5. In Subsection 3.4.2 we prove the previous results to show that random walk paths are nice with high probability. In Subsection 3.5.1, the upper bound on the mixing time is provided thanks to Proposition 3.26, which exploits the properties of nice paths and makes use of a suitable concentration result for independent variables. The lower bound is proved in Subsection 3.5.2, applying the Quenched LLN 3.21. Finally in Subsection 3.5.3, the asymptotics for the cutoff window is proved. The further result which is needed here is given by Proposition 3.29, a bound on the L^2 of π, which allows to complete the proof of the lower bound.

Perturbed Markov chains

An interesting line of further investigations is related to parametric perturbations of Markov chains that exhibit a cutoff. More explicitly, focusing the analysis on a particular choice of chain and a way of implementing the perturbation, it is interesting to quantify the robustness of the cutoff phenomenon with respect to the strength of the perturbation.

In this setup, the mixing time of the dynamics may exhibit an additional phase transition, governed by the strength of the perturbation and referred to as a *mixing trichotomy*. This involves the identification of a *subcritical* regime, where the perturbation is sufficiently weak to preserve the mixing behavior of the original chain, contrasted with a *supercritical* regime, where the "cutoff picture" is disrupted, leading to smooth rather than abrupt mixing. As in the broader framework of statistical physics, the description of such a phase transition is typically accompanied by the identification of a *critical* regime, at the interphase between the other two, where the system exhibits some sort of intermediate behavior.

Among the models for which such a phenomenon has been rigorously proved, we recall: Glauber dynamics for the Ising model Lubetzky and Sly [2013, 2014], random walks on dynamic graphs Avena et al. [2019, 2022], Caputo and Quattropani [2021b], random walks with reset (and the so-called PageRank dynamics) Caputo and Quattropani [2021a], Vial and Subramanian [2025], Ehrenfest urns with multi-type particles Quattropani [2024], mass redistribution models Caputo et al. [2024]. In particular, although the models in Avena et al. [2019], Caputo and Quattropani [2021a,b] are very different from each other, they all share a common feature: in the supercritical regime, the total variation distance over time—properly rescaled—converges to an exponential function. Moreover, in all these examples, such an exponential decay of the distance to equilibrium can be read from the point of view of the trajectory of the process: the arrival at equilibrium is due to the occurrence of a certain event, and the time of the first occurrence of the latter event is (asymptotically) exponentially distributed.

From a high-level perspective, this is the cartoon underlying another classical phenomenon in statistical physics, known as *metastability*: the system is trapped in a local equilibrium up to some (large) exponential time in which the global equilibrium is eventually reached. In other words, looking through the lens of the mixing trichotomy approach, one might be tempted to see metastability and cutoff as "opposed" phenomena. To make this idea more convincing, it is worth recalling that the emergence of a cutoff is often explained in terms of a concentration phenomenon. Conversely, the memorylessness of the exponential distribution, which characterizes metastable behaviors, can be seen as the complete opposite of concentration.

In the landscape of the previous paragraph, a quite natural choice is given by altering the random graph structure, in order to alter the mixing behavior of the random walk. This has been done in Ben-Hamou [2020] for the non-backtracking random walk on the configuration model. Here, a two community structure is considered, whose strength is governed by a parameter α. This creates a bottleneck which disrupts the cutoff, if and only if α asymptotically dominates the inverse of the entropic time. This was extended later in Hermon et al. [2025] for the simple random walk. Here a m-community structure is considered and the same kind of dichotomy is proved. In Chapter 4 we extend their results to the simple random walk on a directed Erdős–Rényi graph. We consider a m-community structure, depending on a parameter α, for which we achieve a complete *trichotomy picture*. We provide sharp results on the total variation profile, where an exponential profile appears, when α touches the inverse of the entropic time.

Overview of chapter 4: Mixing trichotomy for the SRW on directed block models

In this Chapter we discuss the results in Bianchi et al. [2025].
We consider a block model with $m > 1$ communities, constructed from m independent directed Erdős–Rényi random graphs on the sets $(V_i)_{1\leq m}$ corresponding to m distinct copies of $[n]$ (the *communities*) and a random rewiring procedure governed by a parameter $\alpha \equiv \alpha_n \in [0,1]$, which introduces connections between these communities. The smaller α, the sparser the inter-community connectivity. We then analyze the convergence to the equilibrium of a random walk on the resulting graph, letting the interaction parameter α, or rather the sequence α_n, $n \in \mathbb{N}$, vary on $[0,1]$.

In Section 4.1 we present the model and the main result. Subsection 4.1.1 is dedicated to the formal definition of the graph model. Specifically, we work in a *weakly sparse* regime, setting the connection probability of the Erdős–Rényi communities to $p \equiv p_n = \lambda \log(n)/n$, for a constant $\lambda > 1$, and assuming that each edge is *rewired* so as to point to a different community with probability $\alpha \gg (\lambda n \log n)^{-1}$. The choice $\lambda > 1$ ensures that, before the rewiring, each community is strongly connected with high probability. The requirement $\alpha \gg (n \log n)^{-1}$ is needed for the communities to be typically connected, even though such connections may be sparse.

After having recalled some preliminary material from Chapter 3 in Subsection 4.1.2, we employ Subsection 4.1.3 to state the trichotomy, Theorem 4.3, which can be described as follows. For high values of α, say, $\alpha \gg \boldsymbol{t}_{\text{ent}}^{-1}$, the system is in a subcritical regime with cutoff at time $\boldsymbol{t}_{\text{ent}}$, as described in Eq. (4.3). For lower values of α, say, $\alpha \ll \boldsymbol{t}_{\text{ent}}^{-1}$, the system enters a supercritical regime where the cutoff behavior is disrupted. In this regime, we identify two relevant timescales.

(i) The first corresponds to the attainment of a local equilibrium within the starting community, which causes a sharp drop in the total variation distance from 1 to $(m-1)/m$ at time $\boldsymbol{t}_{\rm ent}$ (see Eq. (4.5)).

(ii) The second timescale is of order α^{-1}, and governs the convergence to the global equilibrium through a smooth exponential decay of the total variation distance: for $t = \beta\alpha^{-1}$ with $\beta > 0$, the distance converges to $\frac{m-1}{m}{\rm e}^{-\frac{\beta m}{m-1}}$ (see Eq. (4.6)).

At criticality, for $\alpha \asymp \boldsymbol{t}_{\rm ent}^{-1}$, where the two timescales are of the same order, the dynamics exhibits an intermediate behavior, where (i) and (ii) are visible at the same scale, as illustrated on the rightmost side of Figure 4.1 (see Eq. (4.4)). Comments on possible generalization follow in Subsection 4.1.4.

Section 4.2 contains some auxiliary material, that follows from Chapter 3, on which the forthcoming analysis is based: the characterization of the stationary distribution of the simple random walk on a single community (Subsection 4.2.1); the fact local neighborhoods are tree–like (Subsection 4.2.2); asymptotics on the annealed random walk (Proposition 4.10 in Subsection 4.2.3) are proved in the spirit of the proof of the quenched LLN 3.21.

Section 4.3 is devoted to the analysis of the weakly supercritical regime. Here, the annealed approximation is successful and allows to prove asymptotics on the first time at which the random walk changes community (Subsection 4.3.1), on the law of the random walk at the time scale α^{-1}, corresponding to the local equilibrium, (see Theorem 4.13 in Subsection 4.3.2), and on the law of the the random walk at the time scale $\boldsymbol{t}_{\rm ent}$ (see Proposition 4.14 in Subsection 4.3.3).

Section 4.4 is devoted to the analysis of the strongly supercritical regime. This section contains the real technical core of the work. In this context, the annealed approximation is no more valid. Nevertheless, we can substitute Proposition 4.10 and Theorem 4.14 with equivalent statement, Proposition 4.15 and Theorem 4.16. To do this we construct an iterative coupling of the original process with a new one, for which explicit and exact computations are feasible. The entire section is devoted to the construction of the random coupling. In Subsection 4.4.1 we study, for each community V_i, the set $\mathcal{G}_i$ of vertices that provide a notion of out-boundary, which we call *gates* (see Eq. (4.21)). The computation of its hitting times $\tau_{\mathcal{G}_i}$ is necessary to study mixing. In Subsection 4.4.2 we introduce the notion of quasi stationary distribution. If $(X_t)_{t\in\mathbb{N}}$ denotes the random walk initialized on V_i, the quasi stationary distribution $\mu_i^\star$ corresponding to $\mathcal{G}_i$ is the distribution on V_i such that

$$\mathbf{P}^G_{\mu_i^\star}(X_t = x|\, \tau_{\mathcal{G}_i} > t) = \mu_\star(x), \qquad \forall x \in V_i \setminus \mathcal{G}_i.$$

This implies exact geometric distribution for $\tau_{\mathcal{G}_i}$ and for this reason we use $\mu_i^\star$ for the initial distribution for our new dynamics. In Subsection 4.4.3 the iterative coupling is formally constructed (see Definition 4.24): the new process $(Y_t)_{t\in\mathbb{N}}$ is a random walk whose initial distribution is the one-step evolution of $\mu_i^\star$ and which is reset each time that fails to change community, after having reached $\mathcal{G}_i$. When the transition is successful, the procedure is repeated on the new community. For the new process, in Proposition 4.27, it is proved that the time needed to change community, rescaled by α^{-1} is asymptotically distributed as an exponential variable of parameter 1. To do this, homogenization properties of the graph are considered, together with the results on the typical points of $\mathcal{G}_i$. The two give that the hitting distribution of $\mathcal{G}$ is asymptotically uniform and that the number of failures before changing community is asymptotically geometric. Later, the coupling is proved to be successful with high probability, as stated in Proposition 4.25. Subsection 4.4.4 concludes the proof of Proposition 4.15 and Theorem 4.16.

Section 4.5 is devoted to the analysis of the subcritical and critical regimes, which correspond to the adaptation to the m community case of the entropic method.

In Section 4.6 we formally prove the trichotomy, wrapping up all previous results and completing with Lemma 4.35 the estimate for the critical case.

Chapter 3

Cutoff for the simple random walk on the directed Chung–Lu graph

In the present chapter, which contains the result of Bianchi and Passuello [2025], we analyze the motion of a random walk on a *Chung–Lu digraph*. This is an inhomogeneous random network obtained by sampling edges independently via vertex weights, which represent fixed average degrees. This setting clearly includes the directed homogeneous *Erdős–Rényi graph*. To ensure that the random graph is strongly connected, and hence to guarantee the uniqueness of the equilibrium measure, we will work on a *weakly sparse regime*, where the average vertex degrees grow as $\log n$, n being the size of the graph.

Our study will mainly refer to the techniques introduced in Bordenave et al. [2018, 2019] to deal with the dynamics on the directed configuration model in the sparse regime. As highlighted in these papers (see also Avena et al. [2024], Cai et al. [2023], Caputo and Quattropani [2020, 2021a] for further developments), two fundamental statistics for the characterization of the mixing time are the *in-degree distribution*, which provides an easily computable approximation of the reversible measure, and the *entropy* of the graph, which measures the spread of the random walk among the network. However, a main hurdle in implementing theses ideas in our framework is that vertex degrees are random, as well as the corresponding in-degree distribution. To overcome this difficulty we shall introduce an approximated, but deterministic, in-degree distribution (see (3.4)), and then leverage on some *concentration results* on the vertex degrees in order to control this approximation error along the dynamics and to characterize asymptotically the entropy (see (3.10) and Proposition 3.2). By implementing this entropic method, devised in Bordenave et al. [2018, 2019], we will prove that under suitable assumptions the dynamics exhibits a cutoff phenomenon at a time of order $\log n/\log\log n$. Moreover, we will show that, in an appropriate time window, the cutoff profile approaches a Gaussian tail function. This work can be seen as a generalization of the cutoff results achieved in Bordenave et al. [2018], Cai et al. [2023], where hard constraints on vertex degrees are replaced with a softer randomized version.

3.1 Setup and results

3.1.1 Model

Let $[n] := \{1, \dots, n\}$ represent a set of vertices of size $n \in \mathbb{N}$, and consider two sequences $(w_x^-)_{x\in[n]}$ and $(w_x^+)_{x\in[n]}$ of positive numbers, called weights, such that

$$\sum_{x\in[n]} w_x^+ = \sum_{x\in[n]} w_x^- =: \boldsymbol{w}(n) = \boldsymbol{w}\,.$$

We consider a directed version of the Chung–Lu model, where two distinct vertices $x, y \in [n]$ are connected by an oriented edge from x to y, in short $x \to y$, independently and with probability

$$p_{x,y} = w_x^+ w_y^- \frac{\log n}{n} \wedge 1\,, \quad \forall x, y \in [n]\,, x \neq y\,. \tag{3.1}$$

We will denote by $\mathbb{P} = \mathbb{P}_n^{w^\pm}$ the measure associated to this Chung–Lu random graph and by $\mathbb{E}$ the corresponding average, and write G for a given realization of the graph. The next remark shows that this formulation of the model is equivalent to the one given in Chapter 2.

Remark 3.1. *The standard non-oriented Chung–Lu model, introduced in Chung and Lu [2002a], is defined through a sequence of positive weights $(\widetilde{w}_x)_{x\in[n]}$ and connection probabilities*

$$p_{xy}^{(\mathrm{CL})} := \frac{\widetilde{w}_x \widetilde{w}_y}{\ell_n} \wedge 1, \quad \forall x \neq y \in [n], \qquad \text{where} \qquad \ell_n := \sum_{x\in[n]} \widetilde{w}_x.$$

This can be easily adapted to the above directed framework taking two sequences $(\widetilde{w}_x^\pm)_{x\in[n]}$ with equal sum, and setting

$$p_{xy}^{(\mathrm{DCL})} := \frac{\widetilde{w}_x^+ \widetilde{w}_y^-}{\ell_n} \wedge 1, \quad \forall x \neq y \in [n], \qquad \text{where} \qquad \ell_n := \sum_{x\in[n]} \widetilde{w}_x^+ = \sum_{x\in[n]} \widetilde{w}_x^-\,. \tag{3.2}$$

Choosing $\widetilde{w}_x^\pm = w_x^\pm \boldsymbol{w} \frac{\log n}{n}$ and plugging this value in (3.2), we get that $\ell_n = \sum_{x\in[n]} \widetilde{w}_x = \boldsymbol{w}^2 \frac{\log n}{n}$, and we recover our model.

As main observables on this random structure, we introduce the random out-degree of a vertex $x \in [n]$, denoted by D_x^+, and set

$$\delta_+ := \min_{x\in[n]} D_x^+ \qquad \text{and} \qquad \Delta_+ := \max_{x\in[n]} D_x^+\,,$$

which are, respectively, the minimum and maximum out-degree of the random graph. With obvious notation, we introduce also the corresponding in-degrees random variables $(D_x^-)_{x\in[n]}$, δ_- and Δ_-.

By assumption, the out- and in-degrees of each vertex $x \in [n]$ are distributed as a sum of independent Bernoulli random variables of parameters $p_{x,y}$ and $p_{y,x}$ respectively, for $y \in [n] \setminus \{x\}$. In particular, their averages are easily given by

$$\mathbb{E}[D_x^\pm] = \sum_{y \in [n] \setminus \{x\}} w_x^\pm w_y^\mp \frac{\log n}{n}.$$

Along the paper, we will use usual Landau asymptotic notation (cfr. p. vii). In Subsection 3.1.2, we will set assumptions on $(w_x^\pm)_{x\in[n]}$ which imply, in the above notation, that $\mathbb{E}[D_x^+] = \Theta(\log n)$, for $x \in [n]$. The corresponding random graph will be then in a *weakly sparse regime*.
At last, note that the Erdős–Rényi digraph with connection probability $p = \lambda \log n/n$, for $\lambda > 0$, corresponds to a homogeneous Chung–Lu digraph with constant weights $\sqrt{\lambda}$.

The Chung–Lu model that we have just portrayed offers a good framework to study random dynamics. We consider the discrete time simple random walk, $(X_t)_{t\in\mathbb{N}}$, whose transition matrix is

$$P(x,y) := \begin{cases} \frac{1}{D_x^+} & \text{if } x \to y \\ 0 & \text{otherwise} \end{cases}, \qquad \forall x, y \in [n].$$

For every time $t > 0$ (for the sake of simplicity t has to be understood as an integer, or its integer part), we denote with $P^t(\cdot,\cdot)$ its related t-step transition kernel, while for an oriented path $\mathfrak{p} = (x_0, \dots, x_t)$ in the graph, we define the probability mass of $\mathfrak{p}$ as

$$\mathbf{m}(\mathfrak{p}) := \prod_{i=0}^{t-1} P(x_i, x_{i+1}), \tag{3.3}$$

which corresponds to the probability that a random walk starting at x_0 follows the trajectory $\mathfrak{p}$. We point out that $\mathbf{m}(\cdot)$, as the transition kernel $P(\cdot,\cdot)$, is a random object whose dependence on the random graph is implicit in the notation.

For any given realization G of the random graph, we can consider the probability measure $\mathbf{P}_\mu^G$ associated to the simple random walk, when the initial position of the walk has distribution μ on $[n]$. For a probability distribution μ on $[n]$, we will denote with $\mathbf{P}_\mu^G(X_t \in \cdot)$ the quenched law of the random walk at time t with initial distribution μ. When $\mu = \delta_x$ for some $x \in [n]$, this can also be denoted by $P^t(x,\cdot)$. Averaging over all graph realizations, we obtain the corresponding annealed measure $\mathbb{P}_\mu^{\mathrm{an}}$, which is defined by $\mathbb{P}_\mu^{\mathrm{an}}(A) := \mathbb{E}[\mathbf{P}_\mu^G(A)]$ for every measurable set A of trajectories of the random walk. In our framework the random structure is fixed once forever. We refer to Avena et al. [2019, 2022], Caputo and Quattropani [2021b], Sousi and Thomas [2020] for the analysis of dynamic networks.

Uniqueness of the invariant distribution. As long as a realization G of the Chung–Lu digraph is strongly connected, i.e. there exists a directed path among every couple of vertices $x, y \in [n]$, the irreducibility condition of simple random walks is satisfied and this guarantees that there exists a unique invariant measure π on $[n]$ such that $\pi P = \pi$.

Then, we are at first interested in finding sufficient conditions which ensure strong connectivity with probability tending to 1 as $n \to \infty$ (in short *with high probability* or simply *w.h.p.*). It was proved in Cooper and Frieze [2012] that the Erdős–Rényi digraph with parameter $\lambda \log n/n$, where $\lambda > 1$, is w.h.p. strongly connected. To be more precise this remains true when $(\lambda - 1)\log(n) \to +\infty$ as $n \to +\infty$. Provided that there exists a constant λ such that $w_x^+ w_y^- \geq \lambda > 1$ for every $x, y \in [n]$, and since the strong connectivity is a monotone increasing property of graphs, a simple coupling argument leads to the same conclusion for the Chung–Lu graph. Hence, this condition guarantees the existence and uniqueness of the invariant distribution π, and it will be part of the set of assumptions on the graph setting that will be given below, before stating the main results. Let us mention that we do not need the edge density to be of the order $\log(n)/n$ for a strongly connected component to exist. However, below that threshold, the latter will contain w.h.p. some dead ends (i.e., vertices with null out-degree), and this will prevent the random walk to be irreducible.

Provided that the stationary distribution is unique, the main goal of this work is to characterize the mixing time of the random walk, which is defined, for any initial state $x \in [n]$ and any precision $\varepsilon \in (0,1)$, as

$$\boldsymbol{t}_{\text{mix}}^{(x)}(\varepsilon) := \inf\{t > 0 : \|P^t(x,\cdot) - \pi\|_{\text{TV}} \leq \varepsilon\},$$

where $\|\mu - \nu\|_{\text{TV}} := \frac{1}{2}\sum_{x\in[n]} |\mu(x) - \nu(x)| = \sum_{x\in[n]} [\mu(x) - \nu(x)]^+$, is the total variation distance among the probability measures μ and ν. Here $[u]^+ := \max\{0, u\}$, for $u \in \mathbb{R}$.

We stress once more that the mixing time depends on the realization G of the graph, though the dependence is implicit in the notation. We will prove that our estimates on $\boldsymbol{t}_{\text{mix}}^{(x)}(\varepsilon)$ hold in $\mathbb{P}$-probability as $n \to \infty$.

In-degree distribution. One of the main hurdles to estimate the mixing time of simple random walks on digraphs is that the stationary measure π cannot be explicitly computed. In this respect, a useful tool is provided by the following probability measure on the set $[n]$,

$$\mu_{\text{in}}(x) := \frac{w_x^-}{\boldsymbol{w}}, \quad \text{for } x \in [n]. \tag{3.4}$$

The measure μ_{in} can be seen as an approximate averaged in-degree distribution. Specifically, under the upcoming assumption (3.5), we can deduce that for large n

$$\mathbb{E}[D_x^-] = \boldsymbol{w}\frac{\log(n)}{n} w_x^- (1 + o(1)), \qquad \sum_{x\in[n]} \mathbb{E}[D_x^-] = \boldsymbol{w}^2 \frac{\log(n)}{n}(1 + o(1)),$$

and then, taking the ratio among the two terms, we get

$$\mu_{\text{in}}(x) = \frac{\mathbb{E}[D_x^-]}{\sum_{x\in[n]} \mathbb{E}[D_x^-]}(1 + o(1)).$$

The measure μ_{in} will then simply referred to as in-degree distribution, and it will naturally appear through the proofs as a fundamental object in understanding the mixing mechanism of the dynamics.

3.1.2 Assumptions

We assume that:

1. There exist constants $M_0, M_1 > 1$ such that, for every $n \in \mathbb{N}$,

$$M_0 \leq w_x^+ \leq M_1 < +\infty, \quad \forall x \in [n]; \tag{3.5}$$

2. There exist constants $M_2 > 0$ and $0 < \eta < 1$ such that, for every $n \in \mathbb{N}$,

$$\sum_{x \in [n]} (w_x^-)^{2+\eta} \leq M_2 n. \tag{3.6}$$

Notice that, as a consequence of (3.5),

$$\boldsymbol{w} = \Theta(n) \quad \text{ and } \quad \mathbb{E}[D_x^+] = \Theta(\log n)\,,\ \forall x \in [n]\,.$$

Moreover, exploiting (3.6), we get that $\max_{x \in [n]} (w_x^-)^{2+\eta} \leq M_2 n$, and thus

$$w_x^- \leq (M_2 n)^{\frac{1}{2+\eta}} \leq (M_2 n)^{\frac{1}{2} - \frac{\eta}{6}}, \quad \forall x \in [n]\,, \tag{3.7}$$

which in turn implies, by (3.1), that

$$p_{\max} := \max_{x \neq y \in [n]} p_{x,y} = o(n^{-\frac{1}{2} - \frac{\eta}{7}}), \tag{3.8}$$

and

$$\mu_{\text{in}}^{\max} := \max_{x \in [n]} \mu_{\text{in}}(x) = O(n^{-\frac{1}{2} - \frac{\eta}{6}}). \tag{3.9}$$

In particular, following the terminology introduced in Caputo and Quattropani [2021a], the above assumptions imply that μ_{in} is a *widespread measure*.

3.1.3 Main results

Before stating the main results, and following the procedure traced in Bordenave et al. [2018], we need to introduce two fundamental quantities that will characterize the mixing time and the cutoff window of the dynamics. We define the entropy H of the Chung–Lu model as the mean logarithmic out-degree of a vertex sampled from μ_{in} (see (3.4)). Formally, we set

$$\mathrm{H} := \mathbb{E}\left[\sum_{x \in [n]} \mu_{\text{in}}(x) \log\left(D_x^+ \vee 1\right)\right], \tag{3.10}$$

and

$$\sigma^2 := \mathbb{E}\left[\sum_{x\in[n]} \mu_{\text{in}}(x)\log^2\left(D_x^+ \vee 1\right)\right] - \mathrm{H}^2\,. \tag{3.11}$$

We also define the *entropic time*

$$\boldsymbol{t}_{\text{ent}} := \frac{\log n}{\mathrm{H}}\,,$$

which we will show to be precisely the mixing time of the dynamics. In this sense, it is useful to state the following preliminary result which provides the asymptotic behavior of H and σ^2, as $n \to \infty$.

Proposition 3.2. *Under the assumptions* (3.5) *and* (3.6)*, it holds*

$$\mathrm{H} = \log\log n(1+o(1)), \qquad \sigma^2 = O(\log\log n).$$

While the proof of the above proposition is postponed to Subsection 3.3.2, we can immediately argue that the entropic time $\boldsymbol{t}_{\text{ent}}$ is asymptotically of order $\log n/\log\log n$. With that in mind, and with the usual convention that the discrete dynamics is evaluated in the integer part of each considered time, we can state our main results.

Theorem 3.3 (Cutoff)**.** *Let* $\beta \in (0,1)$*. It holds*

$$\min_{x\in[n]} \|P^{(1-\beta)\,\boldsymbol{t}_{\text{ent}}}(x,\cdot) - \pi\|_{\text{TV}} \xrightarrow[n\to+\infty]{\mathbb{P}} 1. \tag{3.12}$$

and

$$\max_{x\in[n]} \|P^{(1+\beta)\,\boldsymbol{t}_{\text{ent}}}(x,\cdot) - \pi\|_{\text{TV}} \xrightarrow[n\to+\infty]{\mathbb{P}} 0. \tag{3.13}$$

Remark 3.4. *By the monotonicity properties of the function* $t \mapsto \|P^t(x,\cdot) - \pi\|_{\text{TV}}$ *for* $x \in [n]$*, we get that* (3.12) *holds for any* $t \le (1-\beta)\boldsymbol{t}_{\text{ent}}$*, while* (3.13) *holds for any* $t \ge (1+\beta)\boldsymbol{t}_{\text{ent}}$*.*

The statement can be rephrased as follows: for every precision $\varepsilon \in (0,1)$,

$$\max_{x\in[n]} \left|\frac{\boldsymbol{t}_{\text{mix}}^{(x)}(\varepsilon)}{\boldsymbol{t}_{\text{ent}}} - 1\right| \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

This means that regardless of the starting point and the precision, the random walk takes, with high probability for n large enough, $\log n/\log\log n$ steps to mix.

This abrupt transition from 1 to 0 of the distance to stationarity can be further explored by zooming in around the cutoff point $\boldsymbol{t}_{\text{ent}}$, and in particular by taking an appropriate window of size $\mathbf{w}_{\text{ent}}$, with

$$\mathbf{w}_{\text{ent}} := \frac{\sigma}{\mathrm{H}}\sqrt{\boldsymbol{t}_{\text{ent}}}\,. \tag{3.14}$$

To avoid pathological situations, we will assume that σ^2 is non-degenerate in the following weak sense: there exists $\delta > 0$ such that

$$\sigma^2 \gg \frac{(\log\log n)^{2+\frac{\delta}{\delta+2}}}{(\log n)^{\frac{\delta}{\delta+2}}}. \tag{3.15}$$

Note that as $\delta \to \infty$, the r.h.s. reaches the order $(\log\log n)^3/\log n$, providing a non-degeneracy condition similar to that given in Bordenave et al. [2018]. This condition is expected to hold when the amount of inhomogeneity in the graph is sufficiently high. The next result shows that, inside this window and under this assumption, the cutoff shape approaches the tail distribution of the standard normal, as observed in Bordenave et al. [2018].

Theorem 3.5 (Cutoff window)**.** *Assume that the σ^2 satisfies the non-degeneracy condition (3.15). Then, for $t_\lambda := \boldsymbol{t}_{\mathrm{ent}} + \lambda \mathbf{w}_{\mathrm{ent}} + o(\mathbf{w}_{\mathrm{ent}})$ with $\lambda \in \mathbb{R}$ fixed, it holds*

$$\max_{x\in[n]} \left| \|P^{t_\lambda}(x,\cdot) - \pi\|_{\mathrm{TV}} - \frac{1}{\sqrt{2\pi}} \int_\lambda^{+\infty} e^{-\frac{u^2}{2}}\, du \right| \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

Remark 3.6. *Notice that the statements of Theorems 3.3 and 3.5 can be easily extended to Chung–Lu digraphs with random sequences of weights $(W_1^+,\dots,W_n^+)$ and $(W_1^-,\dots,W_n^-)$ which satisfy* a.s. *the constraints (3.5) and (3.6).*

3.2 Proof outline and main ingredients

3.2.1 General strategy

A main hurdle in the analysis of the mixing time of simple random walks on digraphs is the lack of an explicit formula for the stationary measure π. To cope with that, we will introduce an explicit probability measure $\widetilde{\pi}$ that well approximates π itself.

Using this idea, and looking first at an upper bound on the mixing time, by the triangle inequality we can write

$$\|P^t(x,\cdot) - \pi\|_{\mathrm{TV}} \le \|P^t(x,\cdot) - \widetilde{\pi}\|_{\mathrm{TV}} + \|\widetilde{\pi} - \pi\|_{\mathrm{TV}}, \quad \forall x \in [n]\,.$$

Note that if the first term in the r.h.s. is $o_{\mathbb{P}}(1)$ uniformly in $x \in [n]$, where $X_n = o_{\mathbb{P}}(Y_n)$ means that $|X_n/Y_n|$ vanishes in probability, then the same must hold for the second term since

$$\|\widetilde{\pi} - \pi\|_{\mathrm{TV}} = \|\widetilde{\pi} - \pi P^t\|_{\mathrm{TV}} = \sum_{x\in[n]} \pi(x)\|P^t(x,\cdot) - \widetilde{\pi}\|_{\mathrm{TV}}\,. \tag{3.16}$$

This is what we will prove for $t \ge (1+\beta)\boldsymbol{t}_{\mathrm{ent}}$, taking $\widetilde{\pi} := \mu_{\mathrm{in}} P^{h_\varepsilon}$, with $\varepsilon > 0$ and

$$h_\varepsilon := \frac{\varepsilon \log n}{20\mathrm{H}}\,. \tag{3.17}$$

More precisely, we will prove the following slightly weaker condition

$$\|P^t(x,\cdot) - \widetilde{\pi}\|_{\mathrm{TV}} = o_{\mathbb{P}}(1)\,, \quad \forall x \in V_\varepsilon^\star\,, \tag{3.18}$$

where $V_\varepsilon^\star$ is a subset of $[n]$ whose vertices have a locally tree-like out-neighborhood. This result will be sufficient to derive a proper upper bound on the mixing time as stated in (3.13) of Theorem 3.3.

As a further main tool to obtain (3.18), which also enters in the proof of the lower bound on the mixing time, we will introduce a suitable set of t-length paths, called *nice paths*, that will be shown to be typical trajectories of the simple random walk. Taking advantage of their properties, we will prove that, for any $\delta > 0$,

$$\|P^t(x,\cdot) - \widetilde{\pi}\|_{\mathrm{TV}} \le \mathcal{Q}_{x,t}\Big(\frac{1}{n\log^3 n}\Big) + 3\delta\,,$$

where for $x \in [n]$ and $\theta \in (0,1)$, $\mathcal{Q}_{x,t}(\theta)$ is the quenched probability that the mass of a path of length t selected by a random walk with initial point x is bigger than θ. Formally

$$\mathcal{Q}_{x,t}(\theta) := \mathbf{P}_x(\mathbf{m}(X_0, X_1, \dots, X_t) > \theta), \tag{3.19}$$

where $\mathbf{P}_x$ is the quenched measure associated to a random walk starting at x as in Subsection 3.1.1, and $\mathbf{m}(\cdot)$ is the mass of a path as given in (3.3).

A similar approach can be implemented to obtain a lower bound on the total variation distance as stated in (3.12). In particular, for $t = (1-\beta)\log n$ and $\theta = \log^a n/n$ (with a suitable $a \in \mathbb{N}$), it will lead to the inequality

$$\mathcal{Q}_{x,t}(\theta) \le \|\widetilde{\pi} - P^t(x,\cdot)\|_{\mathrm{TV}} + o_{\mathbb{P}}(1)\,. \tag{3.20}$$

The function $\mathcal{Q}_{x,t}(\theta)$ is thus one of the main characters of our analysis, and it will carry a very powerful limit result: in Theorem 3.21 we will observe that according to the choices (t,θ), it may vanish or saturate to 1. This dichotomy will actually conclude the proof of the cutoff regime (Theorem 3.3), and provide the main strategy for the proof of cutoff profile (Theorem 3.5).

We would like to emphasize again that the overall strategy of our proofs follows the entropic method developed by Bordenave, Caputo, and Salez in Bordenave et al. [2018, 2019] for analyzing random walks on sparse directed configuration models. While we draw on these ideas, our implementation occurs in a different connectivity regime, where average degrees diverge, and other significant modifications are necessary. In the case of directed configuration models, the analysis often relies on combinatorial computations, which are feasible due to the deterministic nature of in- and out-degrees. However, this approach is not applicable to the Chung–Lu setting, where the in- and out-degrees are themselves random. Instead, the Chung–Lu model benefits from the independence of edges, a property we crucially exploit in our analysis, along with appropriate concentration inequalities for the in- and out-degrees.

Finally, note that our results are consistent with Theorem 3 in Lubetzky and Sly [2010], which is set in the context of undirected regular random graphs in a weakly sparse setting. Although we are not aware of analogous results for the undirected Chung–Lu model in this regime, we believe that similar conclusions can be drawn for a broad class of undirected random graphs. However, in this setting, the speed of the random walk enters the game and needs to be properly analyzed (see Berestycki et al. [2018] for the study of sparse undirected graphs). It is also worth mentioning the analysis in Fountoulakis and Reed [2008], where the authors study the mixing time of random walks on the giant component of Erdős–Rényi graphs with average degree at most $O(\sqrt{\log n})$, which is below the assumptions of the present study. For average degrees $d = \Omega(\log(n)^2)$, we refer to Hildebrand [1996], where it is shown that the mixing time of the random walk is slightly above $\log(n)/\log(d)$, which turns to be the diameter of the graph and matches our result.

3.2.2 Typical paths and tree-like neighborhoods

We explain here the properties that a path of length t has to satisfy in order to be called *nice*.

Definition 3.7 (Nice path)**.** *Let* $\gamma = \frac{\varepsilon}{80}$*,* $\varepsilon \in (0,1)$*,* h_ε *as in* (3.17)*, and*

$$s := (1-\gamma)\boldsymbol{t}_{\text{ent}}, \qquad t := s + h_\varepsilon + 1 = (1+3\gamma)\boldsymbol{t}_{\text{ent}} + 1.$$

We say that a path $\mathfrak{p} = (x, x_1, x_2, \ldots, x_{t-1}, y)$ *of length* t *from* x *to* y *is nice if*

(i) the entire path is such that $\mathbf{m}(\mathfrak{p}) \leq \frac{1}{n\log^3 n}$*;*

(ii) the first s *steps are contained in certain tree* $\mathcal{T}_x(s)$*, defined below;*

(iii) the last h_ε *steps form the only path in* G *of length at most* h_ε *from its origin to* y*;*

(iv) it holds $P(x_s, x_{s+1}) = 1/D^+_{x_s} \geq \frac{1}{C\log n}$*, for some constant* $C > 0$*.*

Remark 3.8. *Definition 3.7 and the consequent machinery can be extended to times* $t = t_\lambda$*, lying in the critical window of Theorem 3.5. In that case we set* $s = t_\lambda - h_\varepsilon$*.*

To formalize the above properties, it remains to define the tree $\mathcal{T}_x(s)$.

Construction of the tree $\mathcal{T}_x(s)$. For a given realization of the graph G, a fixed root node $x \in [n]$ and a time $s \in \mathbb{N}$, we construct with an iterative procedure two sequences $(\mathcal{G}^\ell)_{\ell\geq 0}$ and $(\mathcal{T}^\ell)_{\ell\geq 0}$ such that, for every $\ell \geq 0$, $\mathcal{G}^\ell$ is a subgraph of G with ℓ edges, while $\mathcal{T}^\ell$ is a spanning tree of $\mathcal{G}^\ell$. The criterion adopted is similar to the one in [Cai et al., 2023, Sect. 3.2].

Set $\overline{\mathrm{H}} := (1+\gamma)\mathrm{H}$, where $\gamma = \frac{\varepsilon}{80}$ as in Definition 3.7. To initialize, let $\mathcal{G}^0 = \mathcal{T}^0 := \{x\}$. Then, for $\ell \geq 1$:

1. Let $\mathcal{E}^\ell$ be the set of edges with tails (i.e., starting points) belonging to $\mathcal{G}^{\ell-1}$, and which have not been yet visited by the first $\ell-1$ iterations of the algorithm. For an edge $e \in \mathcal{E}^\ell$, define the cumulative mass

$$\hat{\mathbf{m}}(e) := \mathbf{m}(\mathfrak{p}_{x,v_e^-}) \frac{1}{D^+_{v_e^-}} , \tag{3.21}$$

where v_e^- is the tail of e, and $\mathfrak{p}_{x,v_e^-}$ denotes the unique path in $\mathcal{T}^{\ell-1}$ from x to v_e^-. In particular, $\hat{\mathbf{m}}(e)$ corresponds to the probability that the random walk follows $\mathfrak{p}_{x,v_e^-}$ and then the edge e.

2. Choose $e_\ell \in \mathcal{E}^\ell$ such that:

   (a) $v_{e_\ell}^-$ is at distance at most $s-1$ from the root x,

   (b) $\hat{\mathbf{m}}(e_\ell) = \max_{e\in\mathcal{E}^\ell} \hat{\mathbf{m}}(e)$ and $\hat{\mathbf{m}}(e_\ell) \geq e^{-\overline{\mathrm{H}}s}$.

   If such edge does not exist, stop the procedure and set $\kappa_x \equiv \kappa_x(s) = \ell - 1$;

3. Generate $\mathcal{G}^\ell$ by adding e_ℓ to $\mathcal{G}^{\ell-1}$;

4. If step (2) does not break the tree structure of $\mathcal{T}^{\ell-1}$, generate $\mathcal{T}^\ell$ by adding e_ℓ to $\mathcal{T}^{\ell-1}$ and otherwise set $\mathcal{T}^\ell = \mathcal{T}^{\ell-1}$.

Note that $\kappa_x \equiv \kappa_x(s)$ is the last step of the iteration, and that it is finite as the graph itself is finite. We then set $\mathcal{G}_x(s) := \mathcal{G}^{\kappa_x}$ and $\mathcal{T}_x(s) := \mathcal{T}^{\kappa_x}$. We observe that $\mathcal{G}_x(s)$ is generated by all paths with mass at least $e^{-\overline{\mathrm{H}}s}$ and length at most s.

We will show that the properties of *nice paths* are satisfied w.h.p. for s as in Definition 3.7 and uniformly in all starting points $x \in V_\varepsilon^\star$, where $V_\varepsilon^\star \subset [n]$ is the random set of vertices mentioned in Eq. (3.18) and defined as follows.

For $h \in \mathbb{N}$ and $x \in [n]$, let us denote with $\mathcal{B}_x^+(h)$ (resp. $\mathcal{B}_x^-(h)$), the set of vertices $y \in [n]$ that are connected to x by an oriented path of length at most h and starting (resp. ending) at point x. They will be called out- (resp. in-)neighborhood of x of radius h. Then we set

$$V_\varepsilon^\star := \{x \in [n] : \mathcal{B}_x^+(h_\varepsilon) \text{ is a directed tree}\}. \tag{3.22}$$

As in Lubetzky and Sly [2010], vertices $x \in V_\varepsilon^\star$ are named h_ε-roots. We will prove that $V_\varepsilon^\star$ is attractive in a sense that will be specified in Lemma 3.20.

3.3 Tools

3.3.1 Annealed random walk

In this subsection we will give an alternative construction of the annealed law of a random walk. We will actually generalize this object to the joint annealed law of K independent random walks defined on the same random graph. This will be used in the forthcoming sections to compute the K-th moment of certain quenched statistics.

Let $K \in \mathbb{N}$. Given an initial distribution μ and a time T, we define iteratively the non-Markovian process $(X^{(k)})_{k\in\{1,\dots,K\}}$, where $X^{(k)} = (X_t^{(k)})_{0\le t\le T}$ is a random walk of length T whose evolution is, for every $k \ge 2$, conditioned to the previous $k-1$ walks. Formally, every random walk $X^{(k)}$ is defined by the following procedure:

(1) Set $X_0^{(k)} \sim \mu$;

Then for all $t \in \{1,\dots,T\}$:

(2)
- If $X_{t-1}^{(k)}$ was never visited before by the previous walks or for $s \le t-1$, generate its out-neighborhood $B_{X_{t-1}^{(k)}} := \mathcal{B}^+_{X_{t-1}^{(k)}}(1)$, according to the probability $\mathbb{P}$, and select a vertex v uniformly at random on $B_{X_{t-1}^{(k)}}$;
- If $X_{t-1}^{(k)}$ has been already visited, extract v uniformly at random from the previously generated out-neighborhood of $X_{t-1}^{(k)}$;

(3) Set $X_t^{(k)} = v$.

The key point of the above construction is that the law of $(X^{(k)})_{k\in\{1,\dots,K\}}$ corresponds to the annealed joint law $\mathbb{P}_\mu^{\mathrm{an},K}((X^{G,1},\dots,X^{G,K}) \in \cdot)$ of a system of K independent random walks $(X^{G,k})_{k\in\{1,\dots,K\}}$ (we recall that G denotes a realization of the graph). Indeed, given a measurable set of trajectories $A \subseteq [n]^{T\times K}$, we have

$$
\begin{aligned}
&\mathbb{P}_\mu^{\mathrm{an},K}((X^{G,1},\dots,X^{G,K}) \in A) = \mathbb{E}\big[\mathbf{P}_\mu^K((X^{G,1},\dots,X^{G,K}) \in A)\big] = \\
=&\mathbb{E}\Big[\sum_{\{x_t^j\}_{t,j}\in[n]^{T\times K}\cap A}\ \prod_{j=1}^{K}\mu(x_0^j)\prod_{t=0}^{T-1}P(x_t^j,x_{t+1}^j)\Big] = \\
=&\mathbb{E}\Big[\sum_{\{x_t^j\}_{t,j}\in[n]^{T\times K}\cap A}\ \prod_{j=1}^{K}\mu(x_0^j)\prod_{t=0}^{T-1}\mathbb{E}\left[P(x_t^j,x_{t+1}^j)\big|(B_{x_r^j})_{r<t}\,,\,(B_{x_r^\ell})_{r\le T,\ell<j}\right]\Big]\,,
\end{aligned}
$$

which characterizes the law of $(X^{(k)})_{k\in\{1,\dots,K\}}$.

Remark 3.9. *Notice that the annealed random walk has an applied interest: its defining algorithm constructs samples of independent random walks moving on a common structure. Understanding their self-repetition properties could provide information on the geometry of the graph, which is very important for statistical inference purposes.*

For a single random walk $X = (X_t)_{t\in\mathbb{N}}$ and any time $s \in \mathbb{N}$, we introduce the event that the vertex X_s was never visited before the step s, formally written as

$$\mathcal{L}_s = \{X_s \neq X_u,\ \forall u \in \{0,\dots,s-1\}\}\,,$$

where for $s = 0$, the event $\mathcal{L}_0$ should be understood as the whole sample space. Using this notation, we are going to prove a result which highlights the role of the measure μ_{in}, defined in (3.4), along the dynamics. Before giving the statement, we recall that $\mu_{\text{in}}^{\max} = \max_{x\in[n]} \mu_{\text{in}}(x) = O(n^{-\frac{1}{2}-\frac{\eta}{6}})$, as observed in (3.9).

Lemma 3.10. *For every initial distribution μ and any positive $s = O\left(n^{1/2}\right)$, it holds*

$$\mathbb{P}_\mu^{\text{an}}(X_s = z, \mathcal{L}_{s-1}) = \mu_{\text{in}}(z)\Big[1 + O\Big(\frac{1}{\sqrt[3]{\log n}}\Big)\Big].$$

Proof. If $s > 1$, and setting $z = z_s \in [n]$, we can write

$$\begin{aligned}
\mathbb{P}_\mu^{\text{an}}(X_s = z, \mathcal{L}_{s-1}) &= \sum_{\substack{z_0,\dots,z_{s-1}\in[n]\\ z_{s-1}\notin\{z_0,\dots,z_{s-2}\}}} \mu(z_0)\mathbb{E}\left[\prod_{i=0}^{s-1} \frac{\mathbf{1}_{\{z_i\to z_{i+1}\}}}{D_{z_i}^+}\right] \\
&= \sum_{\substack{z_0,\dots,z_{s-1}\in[n]\\ z_{s-1}\notin\{z_0,\dots,z_{s-2}\}}} \mu(z_0)\mathbb{E}\left[\prod_{i=0}^{s-2} \frac{\mathbf{1}_{\{z_i\to z_{i+1}\}}}{D_{z_i}^+}\right]\mathbb{E}\left[\frac{\mathbf{1}_{\{z_{s-1}\to z\}}}{D_{z_{s-1}}^+}\right],
\end{aligned} \tag{3.23}$$

Where we used that $\mathbf{1}_{\{z_{s-1}\to z_s\}}$ is independent of the other indicator functions, by definition of $\mathcal{L}_{s-1}$. From the concentration results on the out-degree D_x^+ that will be shown in Subsection 3.3.2, the conditional average appearing in the last display is given, up to lower order terms, by $(\mathbb{E}[D_{z_{s-1}}^+])^{-1} = (\boldsymbol{w}w_{z_{s-1}}^+ \log n/n)^{-1}(1 + O(1/\sqrt[3]{\log n}))$ (see Remark 3.17). Inserting this value in (3.23), using that $p_{z_i,z_{i+1}} = w_{z_i}^+ w_{z_{i+1}}^- \log n/n$, and from the explicit form of μ_{in}, we get

$$\begin{aligned}
&\sum_{\substack{z_0,\dots,z_{s-1}\in[n]\\ z_{s-1}\notin\{z_0,\dots,z_{s-2}\}}} \mu(z_0)\mathbb{E}\left[\prod_{i=0}^{s-2} \frac{\mathbf{1}_{\{z_i\to z_{i+1}\}}}{D_{z_i}^+}\right]\mu_{\text{in}}(z)\Big[1 + O\Big(\frac{1}{\sqrt[3]{\log n}}\Big)\Big] \\
&\qquad = \mathbb{P}_\mu^{\text{an}}(\mathcal{L}_{s-1})\mu_{\text{in}}(z)\Big[1 + O\Big(\frac{1}{\sqrt[3]{\log n}}\Big)\Big].
\end{aligned}$$

We now observe that, for every $i \le s-1$, thanks to (3.8) and our hypothesis on s,

$$1 - o(n^{-\frac{\eta}{7}}) = 1 - sp_{\max} \le \mathbb{P}_\mu^{\text{an}}(\mathcal{L}_i) \le 1.$$

Then the claimed statement holds for all $s > 1$.

If $s = 1$, being $\mathcal{L}_0$ the whole sample space, we get more directly, by the same estimates,

$$\mathbb{P}_\mu^{\text{an}}(X_1 = z) = \sum_{z_0\in[n]} \mu(z_0)\mathbb{E}\left[\frac{\mathbf{1}_{\{z_0\to z\}}}{D_{z_0}^+}\right] = \mu_{\text{in}}(z)\Big[1 + O\Big(\frac{1}{\sqrt[3]{\log n}}\Big)\Big].$$

$\square$

Remark 3.11. *By Lemma 3.10 and Remark 3.17,*

$$\mathbb{P}_\mu^{\mathrm{an}}(X_s = z, \mathcal{L}_{s-1}) = \mathbb{P}_\mu^{\mathrm{an}}(\mathcal{L}_{s-1})\mu_{\mathrm{in}}(z)(1+\epsilon_z),$$

where $0 < \epsilon_z = O(1/\sqrt[3]{\log n})$ *and* $\mathbb{P}_\mu^{\mathrm{an}}(\mathcal{L}_{s-1}^{\complement}) = o(n^{-\frac{\eta}{7}})$*. As a consequence*

$$\begin{aligned}1 = &\sum_{z\in[n]} \mathbb{P}_\mu^{\mathrm{an}}(X_s = z, \mathcal{L}_{s-1}) + \mathbb{P}_\mu^{\mathrm{an}}(\mathcal{L}_{s-1}^{\complement}) \\ =&\mathbb{P}_\mu^{\mathrm{an}}(\mathcal{L}_{s-1})\big(1 + \sum_{z\in[n]} \mu_{\mathrm{in}}(z)\epsilon_z\big) + \mathbb{P}_\mu^{\mathrm{an}}(\mathcal{L}_{s-1}^{\complement}) = 1 + \sum_{z\in[n]} \mu_{\mathrm{in}}(z)\epsilon_z + o(n^{-\frac{\eta}{7}})\ ,\end{aligned}$$

which leads to $\sum_{z\in[n]} \mu_{\mathrm{in}}(z)\epsilon_z = o(n^{-\frac{\eta}{7}})$*. Then, we conclude,*

$$\begin{aligned}2\|\,\mathbb{P}_\mu^{\mathrm{an}}(X_s = \cdot) - \mu_{\mathrm{in}}\|_{\mathrm{TV}} &\le \sum_{z\in[n]} |\,\mathbb{P}_\mu^{\mathrm{an}}(X_s = z, \mathcal{L}_{s-1}) - \mu_{\mathrm{in}}(z)| + \mathbb{P}_\mu^{\mathrm{an}}(\mathcal{L}_{s-1}^{\complement}) \\ &\le \sum_{z\in[n]} \mu_{\mathrm{in}}(z)|\epsilon_z - o(n^{-\frac{\eta}{7}})| + \mathbb{P}_\mu^{\mathrm{an}}(\mathcal{L}_{s-1}^{\complement}) \\ &\le \sum_{z\in[n]} \mu_{\mathrm{in}}(z)(|\epsilon_z| + |o(n^{-\frac{\eta}{7}})|) + \mathbb{P}_\mu^{\mathrm{an}}(\mathcal{L}_{s-1}^{\complement}) = o(n^{-\frac{\eta}{7}}).\end{aligned} \tag{3.24}$$

Let us now define, for every $0 < s < t$ the event $\mathcal{A}_{s,t}$ that the trajectory $(X_u)_{s\le u<t}$ has no self-intersections, formally given by

$$\mathcal{A}_{s,t} \equiv \mathcal{A}_{s,t}^X := \{X_u \ne X_v,\ \forall u \ne v \in \{s, \dots, t-1\}\}\,. \tag{3.25}$$

We set also $\mathcal{A}_t := \mathcal{A}_{0,t}$.

The next result shows that, if the initial measure μ is $\mathrm{Unif}([n])$, then the event $\mathcal{A}_T$ is indeed typical for a time $T = \log^2 n$, which is asymptotically much bigger than $\boldsymbol{t}_{\mathrm{ent}}$. This will be crucial to prove the convergence result inside the cutoff window.

Lemma 3.12. *Let* $T := \log^2 n$*. If* $\mu = \mathrm{Unif}([n])$*, then there exists a constant* $C_1 > 0$ *such that*

$$\mathbb{P}_\mu^{\mathrm{an}}(\mathcal{A}_T^{\complement}) \le C_1 \log^4 n/n\,.$$

Proof. Let τ be the first self-intersection time of X, given by

$$\tau := \min\{s > 0\,:\, \exists u < s \text{ such that } X_s = X_u\}\,,$$

and write

$$\mathbb{P}_\mu^{\mathrm{an}}(\mathcal{A}_T^{\complement}) = \mathbb{P}_\mu^{\mathrm{an}}(\tau < T) = \sum_{t=1}^{T-1} \mathbb{P}_\mu^{\mathrm{an}}(\tau = t), \tag{3.26}$$

where

$$\mathbb{P}_\mu^{\mathrm{an}}(\tau = t) = \sum_{z\in[n]} \Big(\mathbb{P}_\mu^{\mathrm{an}}(X_0 = X_t = z, \tau = t) + \sum_{0<s<t} \mathbb{P}_\mu^{\mathrm{an}}(X_s = X_t = z, \tau = t)\Big). \tag{3.27}$$

We estimate separately the two terms appearing in the above summation.

The first term can be written as

$$\begin{aligned}\mathbb{P}^{\rm an}_\mu(X_0 = X_t = z, \tau = t) &= \mathbb{P}^{\rm an}_\mu(X_t = z, \tau = t | X_0 = z) \cdot \mathbb{P}^{\rm an}_\mu(X_0 = z)\\ &= \mathbb{P}^{\rm an}_z(X_t = z, \tau = t)\mu(z) \le \mathbb{P}^{\rm an}_z(X_t = z, \mathcal{L}_{t-1})\mu(z)\\ &= \frac{1}{n}\mu_{\rm in}(z)(1 + o(1)),\end{aligned} \tag{3.28}$$

where the last identity follows from Lemma 3.10 and using that $\mu = \mathrm{Unif}([n])$. Inserting this value in (3.27) and summing over z, we conclude that this term provides an overall contribution to $\mathbb{P}^{\rm an}_\mu(\tau = t)$ equal to $1/n + o(1/n)$.

Let us turn to the second term. For all $s < t \le T$, we introduce the event

$$\mathcal{B}_{s,t} \equiv \mathcal{B}^X_{s,t} := \{X_v \ne X_u,\ \forall u \in \{0, \dots, s-1\} \text{ and } v \in \{s, \dots, t-1\}\}, \tag{3.29}$$

corresponding to the event that the trajectory $(X_v)_{v\in[s,t)}$ does not intersect the trajectory $(X_u)_{v\in[0,s)}$. Note that, in this notation, $\mathcal{A}_t = \mathcal{A}_s \cap \mathcal{A}_{s,t} \cap \mathcal{B}_{s,t}$, and we can write

$$\begin{aligned}\mathbb{P}^{\rm an}_\mu(X_s = X_t = z, \tau = t) &\le \mathbb{P}^{\rm an}_\mu(X_s = X_t = z, \mathcal{A}_t)\\ &= \mathbb{P}^{\rm an}_\mu(X_s = X_t = z, \mathcal{A}_s \cap \mathcal{A}_{s,t} \cap \mathcal{B}_{s,t})\\ &= \sum_{\substack{v\in([n]\setminus z)^s\\ \text{self-avoiding}}} \mathbb{P}^{\rm an}_\mu(X_t = z, \mathcal{A}_{s,t} | (X_k)_{0\le k\le s} = (v, z),\, \mathcal{B}_{s,t})\\ &\qquad\qquad \times \mathbb{P}^{\rm an}_\mu((X_k)_{0\le k\le s} = (v, z),\, \mathcal{B}_{s,t}).\end{aligned} \tag{3.30}$$

Thanks to the conditioning, the first factor can be written as $\widetilde{\mathbb{P}}^{\rm an}_z(X_{t-s} = z, \mathcal{A}_{t-s})$ where $\widetilde{\mathbb{P}}^{\rm an}_z(\cdot) = \widetilde{\mathbb{E}}[\mathbf{P}_z(\cdot)]$ denotes the annealed measure induced by a Chung–Lu probability measure $\widetilde{\mathbb{P}}$ on a graph with $n - s$ nodes. To the sake of readability we do not stress the dependence of $\widetilde{\mathbb{P}}$ on the vector $v \in ([n] \setminus z)^s$. We conclude observing that, thanks to Lemma 3.10,

$$\widetilde{\mathbb{P}}^{\rm an}_z(X_{t-s} = z, \mathcal{A}_{t-s}) \le \widetilde{\mathbb{P}}^{\rm an}_z(X_{t-s} = z, \mathcal{L}_{t-s}) = \mu_{\rm in}(z)(1 + o(1))$$

Plugging this identity in (3.30), summing over $v \in ([n] \setminus z)^s$, and applying once more Lemma 3.10, we end up with

$$\begin{aligned}\mathbb{P}^{\rm an}_\mu(X_s = X_t = z, \tau = t) &\le \mu_{\rm in}(z)\, \mathbb{P}^{\rm an}_\mu(X_s = z,\, \mathcal{A}_s \cap \mathcal{B}_{s,t})\\ &\le \mu_{\rm in}(z)\, \mathbb{P}^{\rm an}_\mu(X_s = z,\, \mathcal{A}_s) \le \mu_{\rm in}(z)\, \mathbb{P}^{\rm an}_\mu(X_s = z,\, \mathcal{L}_{s-1})\\ &= \mu_{\rm in}(z)^2(1 + o(1))\end{aligned}$$

Inserting this value in (3.27), summing over $s < t$ and $z \in [n]$, and noting that, by assumption (3.6), there exists a finite constant C_1 such that

$$\sum_{z\in[n]} \mu_{\rm in}(z)^2 \le M_2 n/\boldsymbol{w}^2 \le \frac{C_1}{n}, \tag{3.31}$$

we conclude that the contribution to $\mathbb{P}^{\rm an}_\mu(\tau = t)$ of this second term is at most $C_1 \frac{T-1}{n}$. The claimed statement follows including these estimates in (3.26). □

Remark 3.13. *Note that the bound of order* $\log^4 n/n$ *is due to the specific choice of the time* T*. The result can be generally stated for any time* $T \geq \log(n)^2$ *which grows poly-logarithmically in* n*, providing an estimate of order* $O(T^2/n)$*. The requirement over the initial measure can be similarly weakened by replacing* $\mathrm{Unif}([n])$ *with a measure* μ *sufficiently widespread over* $[n]$*, so that* $\max_{x\in[n]} \mu(x) = O(T/n)$ *and the term in* (3.28) *can be properly controlled.*

3.3.2 Properties of the random graph

In this subsection we consider some non-trivial properties of the environment which are the ground floor to understand the typical behavior of random walk paths. We will state two main results about the in- and out-neighborhood of a given vertex, and provide the proof of Proposition 3.2 regarding the entropy asymptotics.

Concentration of out-degrees and entropy. Our first two results concern with the out-degree properties of the graph. They are straightforward consequences of the Chernoff bounds, which we provide below for the reader's convenience (see Prop. 2.21, van der Hofstad [2016]).

Let $X_i \sim \mathrm{Ber}(p_i)$, $i = 1, \dots, n$, be independent Bernoulli random variables of parameter $p_i \in (0,1)$ and let $X = \sum_{i=1}^n X_i$. Then, for every choice of $t > 0$,

$$\begin{aligned} \mathbb{P}(X \geq \mathbb{E}[X] + t) &\leq \exp\left(-\frac{t^2}{2(\mathbb{E}[X] + t/3)}\right), \\ \mathbb{P}(X \leq \mathbb{E}[X] - t) &\leq \exp\left(-\frac{t^2}{2\mathbb{E}[X]}\right). \end{aligned} \tag{3.32}$$

The above Chernoff bounds, applied to the random variables $(D_x^+)_{x\in[n]}$, yields the following bounds on Δ_+ and δ_+ (maximum and minimum out-degree).

Lemma 3.14. *There exists* $C > 1$ *such that the event* $\mathcal{E}^+ := \{\delta_+ \geq 2\} \cap \{\Delta_+ \leq C \log n\}$ *satisfies*

$$\mathbb{P}(\mathcal{E}^+) = 1 - o(1).$$

Proof. Fix a single vertex $x \in [n]$. It holds

$$\mathbb{P}(D_x^+ < 2) = \prod_{y\neq x}(1 - p_{x,y}) + \sum_{z\neq x} p_{x,z} \prod_{y\neq x,z}(1 - p_{x,y}),$$

and recalling that $\log(1-t) \leq -t$ for every $|t| < 1$,

$$\mathbb{P}(D_x^+ < 2) \leq e^{-\sum_{y\neq x} p_{x,y}} + \sum_{z\neq x} p_{x,z} e^{-\sum_{y\neq x,z} p_{x,y}} = O(n^{-w_x^+} \log n).$$

Since $w_x^+ > 1$ for every $x \in [n]$, by a union bound we get $\mathbb{P}(\delta_x^+ < 2) = o(1)$.

To bound below Δ_+, we apply the Chernoff bounds (3.32) to get

$$\mathbb{P}(D_x^+ > C\log n) \le \exp\left(-\frac{(C\log n - \mathbb{E}[D_x^+])^2}{2\left(\mathbb{E}[D_x^+] + \frac{1}{3}(C\log n - \mathbb{E}[D_x^+])\right)}\right),$$

and note that we can choose C sufficiently large to obtain a uniform estimate in x, so that the r.h.s. is of order $n^{-\gamma}$, for any $\gamma > 0$. Then, with a union bound on $x \in [n]$,

$$\mathbb{P}(\Delta_+ \le C\log n) = 1 - o(1).$$

□

Being $w_x^\pm > 1$, δ_+ could be proved to be greater than any constant, but we take $\delta_+ \ge 2$ in analogy to strongly connected configuration models.

Lemma 3.15. *There exists a constant $c > 0$, independent of n, such that, for every vertex $x \in [n]$,*

$$\mathbb{P}(D_x^+ \le c\log n) = o(1)\,.$$

Proof. Applying the Chernoff bounds (3.32) with $X = D_x^+$ and $t := \mathbb{E}[D_x^+] - c\log n > 0$ it holds

$$\mathbb{P}(D_x^+ \le c\log n) \le \exp\left(-\frac{(\mathbb{E}[D_x^+] - c\log n)^2}{2\mathbb{E}[D_x^+]}\right). \tag{3.33}$$

By assumption (3.5), for every $x \in [n]$ it holds $\mathbb{E}[D_x^+] = \Theta(\log n)$, with asymptotic constant uniformly bounded in n. Then, there exists $c > 0$, independent of n, such that

$$\frac{1}{\log n}\cdot\frac{(\mathbb{E}[D_x^+] - c\log n)^2}{2\mathbb{E}[D_x^+]} = \Theta(1)\,,\quad \forall x \in [n].$$

This completes the proof. □

Remark 3.16. *Since $w_x^- > 1$ for every $x \in [n]$, for a point-wise estimate we can simply take $c = 1$. In general, to perform a union bound in (3.33) and prove that $\delta_+ > c\log n$ w.h.p., it must hold, for $x \in [n]$,*

$$\frac{1}{\log n}\cdot\frac{(\mathbb{E}[D_x^+] - c\log n)^2}{2\mathbb{E}[D_x^+]} = \alpha(x)(1 + o(1)),$$

for a constant $\alpha(x)$ such that $\alpha(x) > 1$ uniformly in $x \in [n]$ and $n \in \mathbb{N}$. This can happen only if, for large n and for every $x \in [n]$, $(\boldsymbol{w}w_x^+/n - c)^2 > 2\boldsymbol{w}w_x^+/n$. Since for every $n \in \mathbb{N}$, $c \in (0, \boldsymbol{w}w_x^+/n)$, passing to the roots we derive the equivalent condition that $c < \boldsymbol{w}w_x^+/n - \sqrt{2\boldsymbol{w}w_x^+/n}$ for large n and for every $x \in [n]$.

However, this condition is not always satisfied under our general hypotheses. For instance, on the Erdős–Rényi graph with parameter $\lambda\log n/n$, where $1 < \lambda < \sqrt{2}$, it holds that $\boldsymbol{w}w_x^+/n \equiv \lambda$, and the above condition is satisfied only if c is such that $0 < c < \lambda - \sqrt{2\lambda} < 0$, yielding a contradiction. The above strategy is then insufficient to deal with this specific case.

Remark 3.17. *The Chernoff bounds* (3.32) *provide a precise estimate on the average of the reciprocal of out-degrees. To see this, it is sufficient to plug* $X = D_x^+$ *and* $t = m\mathbb{E}[D_x^+]$ *into* (3.32)*. Since* $\mathbb{E}[D_x^+] = \Theta(\log n)$*, the choice* $m = 1/\sqrt[3]{\log n}$ *implies*

$$\mathbb{E}\left[\frac{1}{D_x^+}\right] = \frac{1}{\mathbb{E}[D_x^+]}\Big[1 + O\Big(\frac{1}{\sqrt[3]{\log n}}\Big)\Big].$$

Notice that, thanks to Jensen's inequality, the multiplicative error term has to be greater than 1. We conclude this subsection providing the proof of Proposition 3.2 about the entropy H. It is a straightforward application of the two previous lemmas.

Proof of Proposition 3.2. From the definition of the entropy H given in (3.10), we can conveniently rewrite

$$\mathrm{H} = \sum_{x\in[n]} \mu_{\mathrm{in}}(x) \sum_{i=2}^{n} \log i \mathbb{P}(D_x^+ = i)\,. \tag{3.34}$$

By Lemmata 3.14-3.15, for every fixed vertex $x \in [n]$,

$$\mathbb{P}(D_x^+ > C\log n) = o(1/n), \qquad \mathbb{P}(D_x^+ < c\log n) = o(1),$$

where $C > 1$ and $c = c(x) > 0$ uniformly in n. Hence

$$\log(c\log n) + o(1) \le \sum_{i=2}^{n} \log i \mathbb{P}(D_x^+ = i) \le \log(C\log n) + o(1/n),$$

which together (3.34), implies that $\mathrm{H} = \log\log n(1 + o(1))$.

From the definition of σ^2 given in (3.11), we can write

$$\sigma^2 = \sum_{x\in[n]} \mu_{\mathrm{in}}(x) \sum_{i=2}^{n} (\log i)^2 \mathbb{P}(D_x^+ = i) - \mathrm{H}^2\,.$$

Since for every $C \in (0 + \infty)$ it holds $(\log(C\log n))^2 = (\log\log n)^2 + 2\log C \log\log n + \log^2 C$, from the previous displays, and inserting the derived estimate of H, we conclude that $\sigma^2 = O(\log\log n)$. □

The entropy H provides an average observable of the system. In the forthcoming sections it will be shown to be deeply connected with the dynamics of the random walk. More precisely, we will deduce from Theorem 3.21 that the probability mass of a typical random walk path of length t is $e^{-\mathrm{H}t + O(\sqrt{\mathrm{H}t})}$.

Size of in-neighborhoods. We now focus on the analysis of the in-neighborhood properties of the graph, that will turn to be fundamental in understanding the spread of the random walk on the environment.

Recall that for $x \in [n]$ and $s \in \mathbb{N}$, $\mathcal{B}_x^+(s)$ and $\mathcal{B}_x^-(s)$ denote, respectively, the out- and in-neighborhood of x with depth s. Following the general proof strategy traced in Cai et al. [2023], we are going to show that, w.h.p. and uniformly in x, the size of an in-neighborhood of radius $\varepsilon \boldsymbol{t}_{\mathrm{ent}}/20$ is at most $n^{1/2+\varepsilon}$.

Lemma 3.18. *Let* $h_\varepsilon = \frac{\varepsilon \log n}{20\mathrm{H}}$ *as in* (3.17)*, and define the event*

$$\mathcal{S}_\varepsilon^- := \{\forall x \in [n], |\mathcal{B}_x^-(h_\varepsilon)| \le n^{1/2+\varepsilon}\}. \tag{3.35}$$

Then $\mathbb{P}(\mathcal{S}_\varepsilon^-) = 1 - o(1)$.

Proof. The idea is to provide a suitable upper bound on $\mathbb{P}(|\mathcal{B}_x^-(h_\varepsilon)| > n^{1/2+\varepsilon})$, and then conclude the proof by a union bound. In this spirit, we claim that, for n large enough,

$$\mathbb{E}[|\mathcal{B}_x^-(h_\varepsilon)|^2 \cdot \mathbf{1}_{\mathcal{E}^+}] \le w_x^- n^\varepsilon \log^3 n\,, \tag{3.36}$$

where $\mathcal{E}^+$ is the typical event described in Lemma 3.14. Assuming its validity, we readily get, by Markov's inequality, that

$$\mathbb{P}(|\mathcal{B}_x^-(h_\varepsilon)| > n^{1/2+\varepsilon}\,,\, \mathcal{E}^+) \le \frac{\mathbb{E}[|\mathcal{B}_x^-(h_\varepsilon)|^2 \cdot \mathbf{1}_{\mathcal{E}^+}]}{n^{1+2\varepsilon}} \le \frac{w_x^- \log^3 n}{n^{1+\varepsilon}}\,.$$

From Lemma 3.14, applying a union bound on $x \in [n]$ and by the assumption (3.6), we conclude that for large n

$$\begin{aligned}\mathbb{P}(\mathcal{S}_\varepsilon^{-\complement}) &= \mathbb{P}(\mathcal{S}_\varepsilon^{-\complement} \cap \mathcal{E}^+) + o(1) \le \sum_{x\in[n]} \mathbb{P}(|\mathcal{B}_x^-(h_\varepsilon)| > n^{1/2+\varepsilon}\,,\, \mathcal{E}^+) + o(1)\\ &\le \frac{\log^3 n}{n^{1+\varepsilon}} \sum_{x\in[n]} w_x^- + o(1) = o(1)\,,\end{aligned}$$

which proves the statement.

It remains to show inequality (3.36). Let $\mathcal{B}_x^\pm = \mathcal{B}_x^\pm(h_\varepsilon)$ and write

$$\mathbb{E}[|\mathcal{B}_x^-|^2 \cdot \mathbf{1}_{\mathcal{E}^+}] = \sum_{y\in[n]} \sum_{z\in[n]} \mathbb{P}(x \in \mathcal{B}_y^+, x \in \mathcal{B}_z^+, \mathcal{E}^+)\,,$$

where

$$\mathbb{P}(x \in \mathcal{B}_y^+, x \in \mathcal{B}_z^+, \mathcal{E}^+) \le \mathbb{P}(x, z \in \mathcal{B}_y^+, \mathcal{E}^+) + \mathbb{P}(x \in \mathcal{B}_y^+, x \in \mathcal{B}_z^+, z \notin \mathcal{B}_y^+, \mathcal{E}^+). \tag{3.37}$$

We start by estimating the first term on the r.h.s. of the last display. Note that, from the independence of the edge connectivity and applying Lemma 3.14, we can write

$$\mathbb{P}(x, z \in \mathcal{B}_y^+, \mathcal{E}^+) = \mathbb{P}(x \in \mathcal{B}_y^+, \mathcal{E}^+)\mathbb{P}(z \in \mathcal{B}_y^+ | \mathcal{E}^+) = \mathbb{P}(x \in \mathcal{B}_y^+, \mathcal{E}^+)\mathbb{P}(z \in \mathcal{B}_y^+, \mathcal{E}^+)(1+o(1))\,,$$

and it is then enough to bound $\mathbb{P}(x \in \mathcal{B}_y^+\,,\, \mathcal{E}^+)$ for general $x \in [n]$.

On the event $\mathcal{E}^+$, the out-neighborhood $\mathcal{B}_y^+$ contains at most $(C \log n)^{h_\varepsilon}$ vertices. Moreover, the probability that a vertex $u \in [n] \setminus \{x\}$ is connected to x is

$$p_{u,x} = w_u^+ w_x^- \frac{\log n}{n} \le M_1 w_x^- \frac{\log n}{n},$$

where M_1 is the constant given in the assumption (3.5). Let A_x denote the subset of $[n]$, of size $(C\log n)^{h_\varepsilon}$, whose vertices maximize the parameters $(p_{u,x})_{u\in[n]\setminus\{x\}}$. Then, for large n,

$$\begin{aligned}\mathbb{P}(x\in\mathcal{B}_y^+\,,\,\mathcal{E}^+) &\le \mathbb{P}\Bigg(\bigcup_{u\in\mathcal{B}_y^+\setminus\{x\}}\{u\to x\}\cap\mathcal{E}^+\Bigg)\le\mathbb{P}\Bigg(\bigcup_{u\in A_x}\{u\to x\}\cap\mathcal{E}^+\Bigg)\\ &\le (C\log n)^{h_\varepsilon}M_1 w_x^-\frac{\log n}{n}\le w_x^- n^{\frac{\varepsilon}{10}}\frac{\log n}{n}\,.\end{aligned}$$

We now bound the second term in (3.37). Note that, given that $x\in\mathcal{B}_y^+$ and $z\notin\mathcal{B}_y^+$, the event $x\in\mathcal{B}_z^+$ can be obtained if either x is the closest vertex to y in $\mathcal{B}_y^+\cap\mathcal{B}_z^+$, or there exists $u\neq x$ which is the closest vertex to y in $\in\mathcal{B}_y^+\cap\mathcal{B}_z^+$ and that is connected to x by a directed path.

Reasoning as before, and for large n, the first scenario has probability less than $(w_x^- n^{\frac{\varepsilon}{10}}\frac{\log n}{n})^2$, while the second scenario is included in the event $E_{y,z,u}=\{u\in\mathcal{B}_y^+\cap\mathcal{B}_z^+\}\cap\{x\in\mathcal{B}_u^+\}$ that has probability

$$\mathbb{P}(E_{y,z,u}\cap\mathcal{E}^+)\le w_x^-(w_u^-)^2 n^{\frac{3\varepsilon}{10}}\frac{\log^3 n}{n^3}\,.$$

All in all, and by assumption (3.6), we get

$$\mathbb{P}(x\in\mathcal{B}_y^+,x\in\mathcal{B}_z^+,z\notin\mathcal{B}_y^+\,,\,\mathcal{E}^+)\le w_x^-w_z^-n^{\frac{\varepsilon}{5}}\frac{\log^2 n}{n^2}+M_2w_x^-n^{\frac{3\varepsilon}{10}}\frac{\log^3 n}{n^2}\,.$$

Summing over y, $z\in[n]$, and using that $\boldsymbol{w}=\Theta(n)$, we get that for large n

$$\mathbb{E}[|\mathcal{B}_x^-|^2\cdot\mathbf{1}_{\mathcal{E}^+}]\le w_x^- n^\varepsilon\log^3 n\,,$$

which concludes the proof of the claimed inequality (3.36), and then of the lemma. □

Tree excess of out-neighborhoods. Following Bordenave et al. [2019], we introduce a quantity that measures how much subgraphs look like trees. Given a graph $S=(V,E)$, we define its tree excess $\mathtt{Tx}(S)$ as the minimum number of edges to remove in order to obtain a directed tree, that is

$$\mathtt{Tx}(S):=1+|E|-|V|\,. \tag{3.38}$$

Then, for every $s\ge 0$, we define the **bad** event $\mathcal{G}^+(s)$ as the set of graphs such that there exists a vertex having an out-neighborhood of depth s with tree-excess greater than 1, that is

$$\mathcal{G}^+(s):=\bigcup_{x\in[n]}\{\mathtt{Tx}(\mathcal{B}_x^+(s))\ge 2\}\,.$$

Lemma 3.19. *Let h_ε be as in (3.17). Then, for all ε sufficiently small, it holds*

$$\mathbb{P}(\mathcal{G}^+(2h_\varepsilon))=o(1)\,.$$

Proof. First note that, for any $x \in [n]$, the event $\{\mathtt{Tx}(\mathcal{B}_x^+(s)) \geq 2\}$ corresponds to the event that, while drawing iteratively $\mathcal{B}_x^+(s)$, at least two vertices are explored at least twice.

Let $C > 1$ be a constant such that $\mathbb{P}(\mathcal{E}^+) = 1 - o(1)$, as in Lemma 3.14, so that, being $\{\Delta \leq C \log n\} \supset \mathcal{E}^+$, it holds that

$$\mathbb{P}(\mathcal{G}^+(2h_\varepsilon)) = \mathbb{P}(\mathcal{G}^+(2h_\varepsilon) \cap \{\Delta_+ \leq C \log n\}) + o(1).$$

On the event $\{\Delta_+ \leq C \log n\}$, the ball $\mathcal{B}_x^+(s)$ has size at most $(C \log n)^{2h_\varepsilon}$, and hence the probability of the event $\{\mathtt{Tx}(\mathcal{B}_x^+(s)) \geq 2\}$ can be bounded above using, as a counter of vertices which are explored at least twice, a binomial random variable $\mathrm{Bin}(m, q)$, where $m = (C \log n)^{2h_\varepsilon}$ is the maximum size of $\mathcal{B}_x^+(s)$, and q bounds above the maximum probability of choosing an already explored vertex.

In particular, letting $p_{\max} := \max_{x,y \in [n]} p_{x,y}$ and with a union bound on the vertices $y \in \mathcal{B}_x^+(s)$, we set $q = (C \log n)^{2h_\varepsilon} p_{\max}$ and get

$$\begin{aligned}\mathbb{P}(\mathtt{Tx}(\mathcal{B}_x^+(s)) \geq 2 \,|\, \Delta_+ < C \log n) &\leq \mathbb{P}\left(\mathrm{Bin}\left(m, q\right) \geq 2\right) \\ &\leq \left((C \log n)^{4h_\varepsilon} p_{\max}\right)^2 .\end{aligned}$$

Since $p_{\max} = O(n^{-\frac{1}{2} - \frac{\eta}{7}})$, due to (3.8), and inserting the explicit value of h_ε, the r.h.s. of the above inequality turns to be $O(n^{-1 + \frac{4\varepsilon}{5} - \frac{2\eta}{7}})$. Observing that

$$\mathbb{P}(\mathtt{Tx}(\mathcal{B}_x^+(s)) \geq 2\,,\, \Delta_+ < C \log n) = \mathbb{P}(\{\mathtt{Tx}(\mathcal{B}_x^+(s)) \geq 2\} \cap \{\Delta_+ < C \log n\})(1 + o(1)),$$

choosing ε sufficiently small, e.g. such that $\frac{4\varepsilon}{5} < \frac{2}{7}\eta$, we conclude the proof by a union bound over $x \in [n]$. □

3.4 Typical properties of random walk trajectories

3.4.1 Mass of a typical trajectory

Having at hand some remarkable properties of the random environment, we switch to consider their impact on the random walk trajectories. The goal of this subsection is to characterize the typical mass of a random walk of length $t = \Theta(\boldsymbol{t}_{\mathrm{ent}})$. In particular, Theorem 3.21 below can be interpreted as a quenched law of large numbers for this quantity (or rather its logarithm). This last result will be then refined to a central limit theorem, which applies to all trajectories of length t, with t taken in an appropriate critical window (see Theorem 3.22 below).

We start with a simple lemma, that is a direct adaptation of Lemma 3.1 in Cai et al. [2023] and that will be useful in the next computations. Recall the definition of the vertex-set $V_\varepsilon^\star$ given in (3.22), whose elements are called h_ε-roots. We are going to show that w.h.p. with respect to the graph setting, the quenched probability that the random walk does not belong to V_ε after t steps decays at least exponentially in t.

Lemma 3.20. *Let h_ε be as in (3.17). Then, for all ε sufficiently small and all $t \le h_\varepsilon$,*

$$\mathbb{P}(\max_{x\in[n]} \mathbf{P}_x^G(X_t \notin V_\varepsilon^\star) \le 2^{-t}) = 1 - o(1).$$

Proof. First note that, in the notation introduced in Subsection 3.3.2, we can rewrite

$$V_\varepsilon^\star = \{y \in [n] : \mathtt{Tx}(\mathcal{B}_y^+(h_\varepsilon)) = 0\}.$$

In particular, due to Lemma 3.19, we can restrict ourselves, with an error of order $o(1)$, to the event

$$(\mathcal{G}^+(2h_\varepsilon))^\complement = \bigcap_{x\in[n]} \{\mathtt{Tx}(\mathcal{B}_x^+(2h_\varepsilon)) \le 1\}\,.$$

In other words, under this event, the out-neighborhood $\mathcal{B}_x^+(2h_\varepsilon)$ is a directed tree except for at most one directed edge, for all $x \in [n]$. If $\mathcal{B}_x^+(2h_\varepsilon)$ is a tree, then also $\mathcal{B}_{X_t}^+(h_\varepsilon)$ is a tree and hence $X_t \in V_\varepsilon^\star$. If $\mathcal{B}_x^+(2h_\varepsilon)$ is not a tree, then it contains precisely one cycle and we can identify the closest node to x on this cycle, say y, that will be at a distance $s < 2h_\varepsilon$ from x. Note that if $s < t$, then necessarily $\mathcal{B}_{X_t}^+(h_\varepsilon)$ is a tree, as the contrary would imply the existence of a second cycle in $\mathcal{B}_x^+(2h_\varepsilon)$, which is impossible under $(\mathcal{G}^+(2h_\varepsilon))^\complement$. Instead, if $t \le s$, the event $\{X_t \notin V_\varepsilon^\star\}$ is realized only if the random walk follows the unique directed path from x to y for t steps. In view of Lemma 3.14, we can further restrict on the event $\mathcal{E}^+$, which ensures that $\delta_+ \ge 2$, and on this event we derive the bound $\mathbf{P}_x^G(X_t \notin V_\varepsilon^\star) \le 2^{-t}$, that holds w.h.p. and concludes the proof. □

Before stating and proving the main results of this section, let us introduce some notation.

Let $(D_k)_{k\ge1}$ be independent copies of D_V^+, the random out-degree of a random vertex $V \in [n]$ sampled from μ_{in}. This sequence is defined w.r.t. a probability measure that with a little abuse of notation will be simply denoted by $\mathbb{P}$. Moreover, for $t \in \mathbb{N}$, set

$$S_t := \sum_{k=1}^t L_k\,, \qquad \text{where} \qquad L_k := \log(D_k \vee 1)\,.$$

Then, for every $\theta \in (0,1)$ and $t \in \mathbb{N}$, we define

$$q_t(\theta) := \mathbb{P}\left(\prod_{k=1}^t \frac{1}{D_k \vee 1} > \theta\right) = \mathbb{P}\left(S_t < -\log(\theta)\right). \tag{3.39}$$

Note that $q_t(\theta)$ corresponds to the probability that a path made of t i.i.d. samples from the in-degree distribution has mass at least θ. Under suitable hypotheses, we will show that the quenched probability $\mathcal{Q}_{x,t}(\theta)$, given in (3.19), is well approximated by $q_t(\theta)$. This is the crucial idea in order to prove the next result.

Theorem 3.21 (Quenched Law of Large Number)**.** *Let $\mathcal{Q}_{x,t}(\theta)$ be the quenched probability given in (3.19), and assume that $t = \Theta(\boldsymbol{t}_{\text{ent}})$ and $\theta \in (0,1)$ are such that*

$$-\frac{\log\theta}{\mathrm{H}t} \xrightarrow{n\to+\infty} \rho. \tag{3.40}$$

Then,

$$\max_{x\in[n]} |\mathcal{Q}_{x,t}(\theta) - \mathbf{1}_{\{\rho>1\}}| \xrightarrow{\mathbb{P}} 0\,.$$

Note that, since $\mathrm{H}t = \Theta(\log n)$, the assumption (3.40) implies that $\log\theta = \Theta(\log n)$. A possible choice could be $\theta = n^{-\rho}$, with possible multiplicative poly-log corrections.

Proof. Our proof follows the strategy given in [Cai et al., 2023, Prop 3.2]. For $\ell = 3\log\log n$, we define

$$\bar{\mathcal{Q}}_{x,t}(\theta) := \sum_{y\in[n]} P^{\ell}(x,y)\mathcal{Q}_{y,t}(\theta).$$

This has the following interpretation. The first ℓ steps do not affect the total mass of the trajectory, but in view of Lemma 3.20, they are sufficient to let the walk move w.h.p. to a h_ε-root vertex. Hence, we let the random walk move for ℓ steps and then start recording the mass of the trajectory. For $\varepsilon \in (0, \eta/2)$, with $\eta \in (0,1)$ as in (3.6), we claim that

$$\max_{x\in V_\varepsilon^\star} |\bar{\mathcal{Q}}_{x,t}(\theta) - q_t(\theta)| \xrightarrow{\mathbb{P}} 0. \tag{3.41}$$

Before proving the claimed convergence, we explore the asymptotic properties of $q_t(\theta)$, and then we complete the proof assuming the validity of (3.41). As a first step, note that since $\{L_k\}_{k\geq 1}$ are i.i.d., and in view of Proposition 3.2, it holds that

$$\mathbb{E}(S_t) = \mathrm{H}t = \log n(1+o(1))\,, \qquad \mathbb{V}\mathrm{ar}(S_t) = \sigma^2 t = O(\log n)\,.$$

From the hypothesis (3.40), it turns that $-\log\theta = \rho\mathbb{E}[S_t](1+o(1))$, so that we may expect the event in the definition of $q_t(\theta)$ to be typical or rare according to the value of ρ. Formally:

(i) if $\rho > 1$ then, for large n, it holds $-\log\theta - \mathbb{E}[S_t] > 0$ and

$$1 - q_t(\theta) = \mathbb{P}\left(S_t \geq -\log\theta\right) = \mathbb{P}\left(S_t - \mathbb{E}[S_t] \geq -\log\theta - \mathbb{E}[S_t]\right)$$

(ii) if $\rho < 1$ then, for large n, it holds $\log\theta + \mathbb{E}[S_t] > 0$ and

$$q_t(\theta) = \mathbb{P}\left(S_t < -\log\theta\right) = \mathbb{P}\left(-S_t + \mathbb{E}[S_t] \geq \log\theta + \mathbb{E}[S_t]\right)$$

In both cases, we can bound above the expression on the right-hand side of the last two displays by Chebyshev's inequality, and get

$$\begin{aligned} &\mathbb{P}\left(|S_t - \mathbb{E}[S_t]| \geq |\log\theta + \mathbb{E}[S_t]|\right) \leq \frac{\mathbb{V}\mathrm{ar}(S_t)}{(\log\theta + \mathbb{E}[S_t])^2} = o(1) \\ \implies \quad & q_t(\theta) \xrightarrow{\mathbb{P}} \begin{cases} 1 & \text{if } \rho > 1 \\ 0 & \text{if } \rho < 1 \end{cases}\,. \end{aligned} \tag{3.42}$$

Going back to the proof of our main statement, let us first observe that since the mass of a path of length ℓ is always in $[\Delta_+^{-\ell}, \delta_+^{-\ell}]$, it holds that

$$\begin{aligned}
\mathcal{Q}_{x,t}(\theta) &\le \mathbf{P}_x^G(\mathbf{m}(X_\ell, X_{\ell+1}, \dots, X_t) > \theta\delta_+^\ell) = \bar{\mathcal{Q}}_{x,t-\ell}(\theta\delta_+^\ell) \\
&\le \mathbf{P}_x^G(\mathbf{m}(X_\ell, X_{\ell+1}, \dots, X_t) > \theta\delta_+^\ell | X_\ell \in V_\varepsilon^\star) + \mathbf{P}_x^G(X_\ell \notin V_\varepsilon^\star) \\
&\le \max_{y\in V_\varepsilon^\star} \mathcal{Q}_{y,t-\ell}(\theta\delta_+^\ell) + \mathbf{P}_x(X_\ell \notin V_\varepsilon^\star) \\
&\le \max_{y\in V_\varepsilon^\star} \bar{\mathcal{Q}}_{y,t-2\ell}(\theta\delta_+^{2\ell}) + \mathbf{P}_x^G(X_\ell \notin V_\varepsilon^\star) \\
&\le \max_{y\in V_\varepsilon^\star} \bar{\mathcal{Q}}_{y,t}(\theta\delta_+^{2\ell}\Delta_+^{-2\ell}) + \mathbf{P}_x^G(X_\ell \notin V_\varepsilon^\star)\,.
\end{aligned}$$

By Lemma 3.20 and assuming the validity of (3.41), we get that

$$\max_{x\in[n]} \mathcal{Q}_{x,t}(\theta) \le q_t(\theta\delta_+^{2\ell}\Delta_+^{-2\ell}) + o_{\mathbb{P}}(1)\,.$$

Since $q_t(\cdot)$ is decreasing, and both w.h.p. $\Delta_+ \le C\log n$ and $\delta_+ \ge 2$ are valid, we conclude that

$$\max_{x\in[n]} \mathcal{Q}_{x,t}(\theta) \le q_t(\theta 2^\ell (C\log n)^{-\ell}) + o_{\mathbb{P}}(1). \tag{3.43}$$

Similarly, we first observe that by definition

$$\begin{aligned}
\mathcal{Q}_{x,t}(\theta) &\ge \mathbf{P}_x^G(\mathbf{m}(X_\ell, X_{\ell+1}, \dots, X_t) > \theta\Delta_+^\ell) = \bar{\mathcal{Q}}_{x,t-\ell}(\theta\Delta_+^\ell) \\
&\ge \mathbf{P}_x^G(\mathbf{m}(X_\ell, X_{\ell+1}, \dots, X_t) > \theta\Delta_+^\ell | X_\ell \in V_\varepsilon^\star)\mathbf{P}_x^G(X_\ell \in V_\varepsilon^\star) \\
&\ge \min_{x\in V_\varepsilon^\star} \mathcal{Q}_{x,t-\ell}(\theta\Delta_+^\ell)\mathbf{P}_x^G(X_\ell \in V_\varepsilon^\star) \\
&\ge \min_{x\in V_\varepsilon^\star} \bar{\mathcal{Q}}_{x,t-2\ell}(\theta\Delta_+^{2\ell})\mathbf{P}_x^G(X_\ell \in V_\varepsilon^\star)\,.
\end{aligned}$$

By Lemma 3.20 and assuming again the validity of (3.41), we obtain

$$\begin{aligned}
\min_{x\in[n]} \mathcal{Q}_{x,t}(\theta) &\ge \min_{x\in V_\varepsilon^\star} \bar{\mathcal{Q}}_{x,t-2\ell}(\theta\Delta_+^{2\ell})(1 - 2^{-\ell} - o_{\mathbb{P}}(1)) \\
&\ge q_{t-2\ell}(\theta\Delta_+^{2\ell}) + o_{\mathbb{P}}(1) \;\ge q_t(\theta\Delta_+^{2\ell}) + o_{\mathbb{P}}(1)\,.
\end{aligned}$$

Since $q_t(\theta)$ is decreasing in t and $\Delta_+ \le C\log n$ w.h.p. , we conclude that

$$\min_{x\in[n]} \mathcal{Q}_{x,t}(\theta) \ge q_t(\theta(C\log n)^{2\ell}) + o_{\mathbb{P}}(1). \tag{3.44}$$

At last note that, setting $\theta' = \theta(C\log n)^{\pm 2\ell}$, then $\log\theta' = \log\theta + O((\log\log n)^2)$. Since the asymptotic value of $q_t(\cdot)$ is not sensitive to perturbations θ' such that $|\log\theta' - \log\theta| = O((\log\log n)^2)$, Eqs. (3.43)-(3.44), together with (3.42), conclude the proof of our statement.

Let us finally prove the claimed convergence (3.41). To this aim, we are going to show that, for all $\delta > 0$,

$$\mathbb{P}(\mathbf{1}_{x\in V_\varepsilon^\star}\bar{\mathcal{Q}}_{x,t}(\theta) \geq q_t(\theta) + \delta) = o(n^{-1}), \tag{3.45}$$

and then we apply a union bound over $x \in V_\varepsilon^\star$. This will give only half of (3.41), but actually the same argument applies to $1 - \bar{\mathcal{Q}}_{x,t}(\theta)$ and $1 - q_t(\theta)$, completing the proof.

For any fixed $K \geq 1$, by Markov's inequality we get

$$\mathbb{P}\left(\mathbf{1}_{x\in V_\varepsilon^\star}\bar{\mathcal{Q}}_{x,t}(\theta) \geq q_t(\theta) + \delta\right) \leq \frac{\mathbb{E}\left[\mathbf{1}_{x\in V_\varepsilon^\star}\left(\bar{\mathcal{Q}}_{x,t}(\theta)\right)^K\right]}{\left(q_t(\theta)+\delta\right)^K}. \tag{3.46}$$

We now follow the strategy of the proof given in [Cai et al., 2023, Prop. 3.2]. Consider the annealed measure $\mathbb{P}_x^{\mathrm{an},K}$ associated to the process $(X^{(k)})_{k\in\{1,\dots,K\}}$ defined in Subsection 3.3.1, for $T = t + \ell$, where $\ell = 3\log\log n$ as above. The process consists of K random walks of length $t + \ell$ and initial measure δ_x, realized one after the other together with the partial graph structure that they explore. Let $K = \lfloor \log^2(n) \rfloor$ and, for every $1 \leq j \leq K$, define the event B_j through the following conditions:

(i) the union of the first j trajectories up to time ℓ, that is $(X_s^{(1)}, \dots, X_s^{(j)})_{s\leq\ell}$, forms a directed tree;

(ii) for every $i \leq j$, the last t steps of the i-th walk, that is $(X_s^{(i)})_{s\in[\ell+1,\ell+t]}$, define a path $\mathfrak{p}$ of mass $\mathbf{m}(\mathfrak{p}) > \theta$;

(iii) The vertices in the first j trajectories have out-degree at least 2.

By definition, note that the event $\{x \in V_\varepsilon^\star\}$ is contained in the event that the K trajectories form a tree up to depth ℓ. Hence

$$\mathbb{E}[\mathbf{1}_{x\in V_\varepsilon^\star}(\bar{\mathcal{Q}}_{x,t}(\theta))^K] \leq \mathbb{P}_x^{\mathrm{an},K}(B_K) = \mathbb{P}_x^{\mathrm{an},K}(B_1)\prod_{j=2}^{K}\mathbb{P}_x^{\mathrm{an},K}(B_j \mid B_{j-1}). \tag{3.47}$$

Note that, given B_{j-1}:

1. either the j-th walk follows one of the previously traced trajectories up to time ℓ, thus keeping unchanged the tree structure of depth ℓ around x.

2. or the j-th walk explores a new vertex before time ℓ. In that case, the event B_j takes place if the j-th walk keeps exploring new vertices at least up to time ℓ, in order to preserve the whole tree structure, and then moves its last t steps on a path $\mathfrak{p}$ with mass $\mathbf{m}(\mathfrak{p}) > \theta$.

Since the out-degree of these vertices is at least 2 by the conditioning, the first scenario happens, for all $j \leq K$, with conditional probability which is at most

$$(K-1)2^{-\ell} \leq K2^{-\ell} = e^{2\log\log n - \ell\log 2} = o(1).$$

To estimate the probability of the second scenario, first note that, at each step, the conditional probability to visit an already explored vertex is less than $K(t+\ell)p_{\max}$. Summing this term for all the $\ell + t$ steps of the path, we obtain that the conditional probability that the j-th walk visits an already explored vertex, and create a cycle along the whole process, is less than $(t+\ell)^2 K p_{\max} = o(1)$, for all $j \leq K$. Hence the tree structure is preserved w.h.p. along the whole trajectory.

Moreover, on the event that the j-th trajectory always visits new vertices, the conditional law of its last t steps corresponds to the annealed law of a random walk of length t defined on a reduced Chung–Lu graph, which is obtained by removing the vertices explored by the whole process before its last t steps, on the event that it has no self-intersections. In particular, from Lemma 3.10 and Eq. (3.24), each step of this random walk can be chosen approximately as a sample of μ_{in}. In other words, after exiting the already visited trajectories, the rest of the path up to step $t+\ell$, can be coupled with an i.i.d. sample from μ_{in} with an overall total variation cost which is of order $O((t+\ell)^2 K p_{\max}) = o(1)$. The second scenario is then satisfied with probability $q_t(\theta) + o(1)$.

Altogether, this shows that, for all $\delta > 0$ and for all $j \leq K$,

$$\mathbb{P}_x^{\text{an},K}(B_j \mid B_{j-1}) \leq q_t(\theta) + \frac{\delta}{2},$$

that, thanks to Eqs. (3.46)-(3.47), implies (3.45). This ends the proof of the claimed convergence (3.41) and of the theorem. □

Let us now consider a time window of size $\mathbf{w}_{\text{ent}} := \frac{\sigma}{\mathrm{H}}\sqrt{\boldsymbol{t}_{\text{ent}}}$, as given in (3.14). Then it holds the following.

Theorem 3.22 (Central Limit Theorem). *Let $t_\lambda := \boldsymbol{t}_{\text{ent}} + \lambda\mathbf{w}_{\text{ent}} + o(\mathbf{w}_{\text{ent}})$, with $\lambda \in \mathbb{R}$ fixed, and assume that $\theta \in (0,1)$ is such that*

$$\frac{\log\theta + \mathrm{H}t_\lambda}{\sigma\sqrt{t_\lambda}} \xrightarrow[n\to+\infty]{} \lambda, \tag{3.48}$$

where σ^2 satisfies the non-degeneracy condition (3.15). Then

$$\max_{x\in[n]}\left|\mathcal{Q}_{x,t_\lambda}(\theta) - \frac{1}{\sqrt{2\pi}}\int_\lambda^\infty e^{-u^2/2}\,du\right| \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

Note that, since $t_\lambda = \boldsymbol{t}_{\text{ent}}(1+o(1))$ and $\mathrm{H}t_\lambda = \log n + \lambda\sigma\sqrt{\boldsymbol{t}_{\text{ent}}}$, the assumption (3.48) implies that $\log\theta = -\log n(1+o(1))$. A possible choice could be $\theta = n^{-1}$, with possible multiplicative poly-log corrections.

Proof. To ease the notation, let $t = t_\lambda$. In view of the convergence (3.41), we first focus on the probability $q_t(\theta)$. By Eq. (3.39), we can write

$$q_t(\theta) = \mathbb{P}\left(\frac{S_t - \mathrm{H}t}{\sigma\sqrt{t}} < -\frac{\log(\theta) + \mathrm{H}t}{\sigma\sqrt{t}}\right).$$

Looking at the argument of that probability, while the r.h.s. converges to $-\lambda$ due to assumption (3.48), we will prove that the l.h.s. converges in distribution to a Normal. We can indeed check that the Lyapunov condition of the Lindeberg-Feller Central Limit Theorem holds (see, e.g., Klenke [2020], Lemma 15.41). Specifically, we need to prove that there exists $\delta > 0$ such that

$$\lim_{n\to+\infty} \frac{1}{\mathrm{Var}(S_t)^{1+\delta/2}} \sum_{k=1}^{t} \mathbb{E}[|L_k - \mathrm{H}|^{2+\delta}] = 0\,. \tag{3.49}$$

We observe that, due to Lemmata 3.14 and 3.15, and by our choice of t, for all $\delta > 0$,

$$\sum_{k=1}^{t} \mathbb{E}[{L_k}^{2+\delta}] = t\mathbb{E}[{L_1}^{2+\delta}] = \log n(\log\log n)^{1+\delta}(1 + o(1))\,.$$

Using that $\mathbb{E}[|L_k - \mathrm{H}|^{2+\delta}] \le 2^{1+\delta}\left(\mathbb{E}[|L_k|^{2+\delta}] + \mathrm{H}^{2+\delta}\right)$, and thanks to Proposition 3.2, we then get that the numerator of (3.49) is $O(\log n(\log\log n)^{1+\delta})$.

On the other hand, let $\delta > 0$ be such that the non-degeneracy condition (3.15) on σ^2 is satisfied. Then

$$\mathbb{V}\mathrm{ar}(S_t)^{1+\delta/2} = (t\sigma^2)^{1+\delta/2} \gg \log n(\log\log n)^{1+\delta}\,,$$

and the Lyapunov condition (3.49) is verified. As a consequence,

$$\lim_{n\to+\infty} q_t(\theta) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{-\lambda} e^{-\frac{u^2}{2}}\,du = \frac{1}{\sqrt{2\pi}}\int_{\lambda}^{+\infty} e^{-\frac{u^2}{2}}\,du.$$

The thesis now follows thanks to the convergence (3.41), together with the bounds (3.43) and (3.44), and to the fact that the asymptotic value of q_t is not sensitive to perturbations θ' such that $|\log\theta' - \log\theta| = O((\log\log n)^2)$. □

3.4.2 Tree-like trajectories

The goal of this subsection is to analyze the random kernel of the random walk in order to prove that the properties characterizing nice paths, listed in Definition 3.7, hold w.h.p. as $n \to \infty$. We will first show that, for all times $s \le (1-\gamma)\boldsymbol{t}_{\mathrm{ent}}$, where $\gamma = \frac{\varepsilon}{80}$ as in Definition 3.7, the random walk trajectories of length s live w.h.p. in the tree $\mathcal{T}_x(s)$ given in Subsection 3.2.2. Accordingly to Remark 3.8, this result can be extended with few little adjustments to times $s = t_\lambda - h_\varepsilon$, with t_λ lying in the critical window of Theorem 3.5. We will briefly comment at the end of the subsection.

We start with a preliminary result. Recall the notation introduced in Subsection 3.2.2 and the procedure to construct the tree $\mathcal{T}_x(s) \subset \mathcal{G}_x(s)$, which involves the sequences of graphs $(\mathcal{G}^\ell)_{\ell\geq 0}$ and $(\mathcal{T}^\ell)_{\ell\geq 0}$, and the sequence of edges $(e_\ell)_{\ell\geq 0}$. In particular, remind that $\mathcal{T}_x(s) := \mathcal{T}^{\kappa_x}$, where κ_x is the index of the last iteration of the algorithm.

Lemma 3.23. *For all $1 \leq \ell \leq \kappa_x$, let e_ℓ denote the edge chosen by the ℓ-th iteration of the algorithm defining $\mathcal{T}_x(s)$. Then, on the event $\mathcal{E}^+$, it holds that*

$$e^{-\overline{\mathrm{H}}s} \leq \hat{\mathbf{m}}(e_\ell) \leq \frac{2}{2+\ell}\,, \tag{3.50}$$

where $\hat{\mathbf{m}}(e_\ell)$ was given in (3.21), and $\overline{\mathrm{H}} = (1+\gamma)\mathrm{H}$. As a consequence, $\kappa_x \leq 2e^{\overline{\mathrm{H}}s}$.

Proof. See the proof of [Bordenave et al., 2018, Lemma 11], which applies to the present setting without substantial changes. □

With this result at hand, we can prove the following proposition, which shows that, w.h.p., a random walk starting from a vertex in $V_\varepsilon^\star$ performs a trajectory in $\mathcal{T}_x(s)$. To state the result, let us denote by $\mathcal{P}(x, y, s, H)$ the set of paths from x to y of length s, in a subgraph H of G .

Proposition 3.24. *For all ε, $\gamma \in (0,1)$ and $s \leq (1-\gamma)\boldsymbol{t}_{\mathrm{ent}}$, it holds*

$$\min_{x\in V_\varepsilon^\star}\left(\sum_{y\in[n]}\sum_{\mathfrak{p}\in\mathcal{P}(x,y,s,\mathcal{T}_x(s))}\mathbf{m}(\mathfrak{p})\right)\xrightarrow[n\to+\infty]{\mathbb{P}} 1.$$

Proof. Note that, by the definition of $\mathcal{T}_x(s)$, a path $\mathfrak{p} \in \mathcal{P}(x, y, s, G)$ is not in $\mathcal{P}(x, y, s, \mathcal{T}_x(s))$ if one of the two following conditions holds:

(1) $\mathbf{m}(\mathfrak{p}) \leq e^{-\overline{\mathrm{H}}s} = 1/n^{1-\gamma^2}(1 + o(1))$.

(2) $\mathfrak{p}$ has edges in $\mathcal{G}_x(s) \setminus \mathcal{T}_x(s)$.

For $j = 1, 2$, we denote with $\mathcal{P}_{x,y}^{j,*}$ the set of paths in $\mathcal{P}(x, y, s, G)$ for which condition (j) does not hold, and observe that by definition

$$\sum_{y\in[n]}\sum_{\mathfrak{p}\in\mathcal{P}_{x,y}^{1,*}}\mathbf{m}(\mathfrak{p}) \geq \mathcal{Q}_{x,s}(1/n^{1-\gamma^2})\,, \quad \forall x \in [n]\,.$$

Since $\frac{\overline{\mathrm{H}}s}{\mathrm{H}s} = 1 + \gamma > 1$, Theorem 3.21 applies and we get that

$$\min_{x\in[n]}\left\{\sum_{y\in[n]}\sum_{\mathfrak{p}\in\mathcal{P}_{x,y}^{1,*}}\mathbf{m}(\mathfrak{p})\right\}\xrightarrow[n\to+\infty]{\mathbb{P}} 1\,, \tag{3.51}$$

which proves that condition (1) is not likely to be satisfied. To estimate the probability that condition (2) is satisfied, let us define iteratively $(M_\ell)_{\ell=0}^{\kappa_x}$ setting

$$M_0 := 0\,, \qquad M_\ell := M_{\ell-1} + \hat{\mathbf{m}}(e_\ell)\mathbf{1}(\ell \leq \kappa_x)\mathbf{1}(v_{e_\ell}^+ \in V(\mathcal{G}^{\ell-1})), \quad \forall \ell \in \{1, \dots, \kappa_x\}\,,$$

where $V(H)$ denotes the vertex set of a graph H and v_e^+ denotes the head of an edge e. Note that M_ℓ represents the total probability mass that is excluded from $\mathcal{G}^\ell$ in the generation of $\mathcal{T}^\ell$. We recall that e_ℓ is the edge selected by the ℓ-th iteration of the algorithm. In particular

$$M_{\kappa_x} = \sum_{y\in[n]} \sum_{\mathfrak{p}\in\mathcal{P}^{2,*}_{x,y}} \mathbf{m}(\mathfrak{p}).$$

We want to show that, for all $\delta > 0$,

$$\mathbb{P}(\exists x \in V_\varepsilon^\star : M_{\kappa_x} > \delta) = o(1).$$

To this aim, we first prove

$$\mathbb{P}(M_{\kappa_x} > \delta, \mathcal{E}^+) = o(n^{-1}), \tag{3.52}$$

so that Lemma 3.14 and a union bound over $x \in V_\varepsilon^\star$ are sufficient to conclude the proof.

Let $\ell_\varepsilon = 2^{h_\varepsilon}$. Remember that condition (1) above is satisfied with vanishing probability for $s = h_\varepsilon = \Theta(\boldsymbol{t}_{\text{ent}})$ and observe that $(\hat{\mathbf{m}}(e_\ell))_{\ell\geq 0}$ is decreasing in ℓ. Moreover notice that, being $x \in V_\varepsilon^\star$, $\mathcal{T}_x(h_\varepsilon)$ is a tree. Combining these facts, it follows that w.h.p. $\mathcal{T}_x(h_\varepsilon) = \mathcal{G}_x(h_\varepsilon) = \mathcal{B}_x^+(h_\varepsilon)$. In conclusion, w.h.p. , $\kappa_x = |\mathcal{T}_{\kappa_x}| \geq |\mathcal{B}_x^+(h_\varepsilon)| \geq 2^{h_\varepsilon} = \ell_\varepsilon$.

As a by-product of the previous lines, we get that in the first ℓ_ε steps of the construction of $\mathcal{T}_{\kappa_x}$, no mass is thrown away and then $M_\ell = 0$ for all $\ell \leq \ell_\varepsilon$. Moreover, due to (3.50), on the event $\mathcal{E}^+$

$$M_\ell - M_{\ell-1} \leq \frac{2}{2+\ell_\varepsilon} \leq 2^{-h_\varepsilon+1} \leq 1\,, \qquad \forall \ell \geq \ell_\varepsilon + 1\,. \tag{3.53}$$

Let $\mathcal{F}_\ell$ denote the σ-field associated to the first ℓ generation steps of $\mathcal{T}_x(s)$. By the previous estimates, it turns out that

$$\mathbb{E}[(M_\ell - M_{\ell-1})\mathbf{1}_{\mathcal{E}^+} \,|\, \mathcal{F}_{\ell-1}] \leq \frac{2}{2+\ell} \cdot \mathbb{P}(v_{e_\ell}^+ \in V(\mathcal{G}^{\ell-1})\,,\, \mathcal{E}^+ | \,\mathcal{F}_{\ell-1})\,, \quad \mathbb{P}\text{−}a.s.\,, \qquad \forall \ell \geq \ell_\varepsilon + 1\,, \tag{3.54}$$

where

$$\mathbb{P}(v_{e_\ell}^+ \in V(\mathcal{G}^{\ell-1})\,,\, \mathcal{E}^+ | \,\mathcal{F}_{\ell-1}) \leq \max_{y\in V(\mathcal{G}^{\ell-1})} \sum_{z\in V(\mathcal{G}^{\ell-1})} p_{y,z} \leq M_1 \frac{\log n}{n} \sum_{z\in V(\mathcal{G}^{\ell-1})} w_z^-\,. \tag{3.55}$$

To estimate the r.h.s. of the display, note that, for any $S \subset [n]$ and taking ζ such that $6\zeta < \eta$, with η as given in assumption (3.6), we can apply Hölder's inequality and get, for $p = 2 + 6\zeta$,

$$\sum_{z\in S} w_z^- \leq \left[\sum_{z\in S} (w_z^-)^p\right]^{\frac{1}{p}} |S|^{1-\frac{1}{p}} \leq \left[\frac{M_2 n}{|S|}\right]^{\frac{1}{2+6\zeta}} |S|. \tag{3.56}$$

We take $S = V(\mathcal{G}^{\ell-1})$ and observe that $\frac{1}{2+6\zeta} < \frac{1}{2} - \zeta$. Since $|V(\mathcal{G}^{\ell-1})| \leq \kappa_x \leq 2n^{1-\gamma^2}$, where the last inequality is due to Lemma 3.23, we obtain that

$$\sum_{z \in V(\mathcal{G}^{\ell-1})} w_z^- = o(n^{1-\xi}),$$

for $\xi > 0$ sufficiently small, depending on the given ζ and γ. Inserting this estimate in (3.55) and then in (3.54), we conclude that, for any $\ell \leq \kappa_x$,

$$\mathbb{E}[(M_\ell - M_{\ell-1})\mathbf{1}_{\mathcal{E}^+} \,|\, \mathcal{F}_{\ell-1}] = \frac{1}{\ell}\, o\left(\frac{\log n}{n^\xi}\right),$$

and in a similar way that

$$\mathbb{E}[(M_\ell - M_{\ell-1})^2\, \mathbf{1}_{\mathcal{E}^+} \,|\, \mathcal{F}_{\ell-1}] = \frac{1}{\ell^2}\, o\left(\frac{\log n}{n^\xi}\right).$$

Consequently,

$$a := \sum_{\ell=1}^{\kappa_x} \mathbb{E}[M_\ell - M_{\ell-1}\mathbf{1}_{\mathcal{E}^+} \,|\, \mathcal{F}_{\ell-1}] = o\left(\frac{\log^2 n}{n^\xi}\right), \tag{3.57}$$

$$b := \sum_{\ell=1}^{\kappa_x} \mathbb{E}[(M_\ell - M_{\ell-1})^2\, \mathbf{1}_{\mathcal{E}^+} \,|\, \mathcal{F}_{\ell-1}] = o\left(\frac{\log n}{n^\xi}\right), \tag{3.58}$$

where we used the fact that $\sum_{\ell=1}^{\kappa_x} \ell^{-1} = O(\log \kappa_x)$. For $\ell \in \{0, \dots, \kappa_x\}$, we define

$$Z_{\ell+1} := \frac{c_\xi}{\delta}(M_{\ell+1} - M_\ell - \mathbb{E}[(M_{\ell+1} - M_\ell)\mathbf{1}_{\mathcal{E}^+}|\mathcal{F}_\ell])\mathbf{1}_{\mathcal{E}^+}, \tag{3.59}$$

where $c_\xi := 2/\xi + 2$. Thanks to (3.53), $|Z_{\ell+1}| \leq 1$ for large n. Since $\kappa_x \geq \ell_\varepsilon$, we also define

$$\phi_u := \sum_{i=\ell_\varepsilon}^{u} Z_{i+1}, \qquad \forall u \in \{\ell_\varepsilon, \dots, \kappa_x\}.$$

The sequence $(\phi_u)_{\ell_\varepsilon \leq u \leq \kappa_x}$ is a martingale. Observe that $M_{\kappa_x} = a + \frac{\delta}{c_\xi}\phi_{\kappa_x}$. Thanks to the estimates (3.57) and (3.58), we can assume that $a \leq \frac{\delta}{c_\xi}$ for large enough n. Hence, recalling that $c_\xi - 2 = 2/\xi$, we can write

$$\mathbb{P}\Big(M_{\kappa_x} \geq \frac{c_\xi - 1}{c_\xi}\delta, \mathcal{E}^+\Big) \leq \mathbb{P}\Big(\phi_\ell \geq \frac{2}{\xi} \text{ for some } \ell \geq \ell_\varepsilon\,,\, \mathcal{E}^+\Big). \tag{3.60}$$

At last, let us consider the conditional variance

$$b' := \sum_{i=1}^{\ell} \mathbb{V}\mathrm{ar}(Z_i|\mathcal{F}_i).$$

On $\mathcal{E}^+$, thanks to (3.58)-(3.59), $b' \le (c_\xi/\delta)^2 b = o\left(\frac{\log n}{n^\xi}\right)$ uniformly in ℓ. Choosing $c(n) = \frac{\log n}{n^\xi}$, for all $n \gg 1$ it holds

$$\mathbb{P}\Big(b' > c(n) \text{ for some } \ell \ge \ell_\varepsilon \,,\, \mathcal{E}^+\Big) = 0\,,$$

and thus

$$\mathbb{P}\Big(\phi_\ell \ge \frac{2}{\xi} \text{ for some } \ell \ge \ell_\varepsilon \,, \mathcal{E}^+\Big) = \mathbb{P}\Big(\phi_\ell \ge \frac{2}{\xi}, b' \le c(n) \text{ for some } \ell \ge \ell_\varepsilon \,, \mathcal{E}^+\Big)\,.$$

As in [Cai et al., 2023, Lemma 3.3], we apply [Freedman, 1975, Theorem 1.6] to bound the r.h.s. with

$$e^{\frac{2}{\xi}} \left(\frac{c(n)}{\frac{2}{\xi} + c(n)}\right)^{\frac{2}{\xi}+c(n)} = o(n^{-1}).$$

Inserting this bound in (3.60), we obtain (3.52) which concludes the proof. □

Remark 3.25. *The statement of this proposition can be easily generalized to time $s = t_\lambda - h_\varepsilon$, with t_λ lying in the critical window of Theorem 3.5. Indeed, with this specific choice, it holds $s = (1-4\gamma)\boldsymbol{t}_{\text{ent}}(1+o(1))$ and $n^{1-4\gamma} \le e^{\overline{\text{H}}s} \le n^{1-\gamma^2}$. All the estimates involving s, and specifically Lemma 3.23 and the convergence (3.51), come true without substantial changes.*

3.5 Proof of the main results

3.5.1 Upper bound

We now prove Eq. (3.13) of Theorem 3.3. We start by rearranging in a more convenient form the total variation distance of the statement. For $h = h_\varepsilon$ as in (3.17), let

$$\widetilde{\pi} := \mu_{\text{in}} P^h\,,$$

and write, by the triangle inequality,

$$\|P^t(x,\cdot) - \pi\|_{\text{TV}} \le \|P^t(x,\cdot) - \widetilde{\pi}\|_{\text{TV}} + \|\widetilde{\pi} - \pi\|_{\text{TV}}.$$

If the first term in the r.h.s. is $o_{\mathbb{P}}(1)$ uniformly in $x \in [n]$, then by the triangle inequality and (3.16), the same must hold for the second term. Let $t' = t + \ell$, with $\ell = \log\log n$ and $t = (1+\beta)\boldsymbol{t}_{\text{ent}}$. Applying the Markov property,

$$\begin{aligned}\max_{x\in[n]} \|P^{t'}(x,\cdot) - \widetilde{\pi}\|_{\text{TV}} &= \max_{x\in[n]} \|\sum_{y\in[n]} P^\ell(x,y)(P^t(y,\cdot) - \widetilde{\pi}\,)(\mathbf{1}_{\{y\in V_\varepsilon^\star\}} + \mathbf{1}_{\{y\notin V_\varepsilon^\star\}})\|_{\text{TV}} \\ &\le \max_{x\in[n]} \frac{1}{2} \sum_{z\in[n]} \sum_{y\in V_\varepsilon^\star} P^\ell(x,y)|P^t(y,z) - \widetilde{\pi}(z)| + \max_{x\in[n]} \frac{1}{2} \sum_{y\in[n]\setminus V_\varepsilon^\star} 2P^\ell(x,y) \\ &\le \max_{y\in V_\varepsilon^\star} \|P^t(y,\cdot) - \widetilde{\pi}\|_{\text{TV}} + \max_{x\in[n]} \mathbf{P}_x^G(X_\ell \notin V_\varepsilon^\star),\end{aligned}$$

where the first inequality is obtained by the triangle inequality and bounding the total variation distance by 2, while the second inequality is obtained by changing the order of the sums and maximizing over y (so that x disappears).

The second term is arbitrarily small due to Lemma 3.20. We then focus on the first term. For sake of readiness we will keep calling x the maximizing variable and we will bound $\max_{x\in V_\varepsilon^\star}\|P^t(x,\cdot)-\widetilde{\pi}\|_{\mathrm{TV}}$. For every $x,y\in[n]$, let $\widetilde{P}^t(x,y)$ be the probability to go from x to y in t steps following a nice path, that is

$$\widetilde{P}^t(x,y)=\sum_{\mathfrak{p}\in\widetilde{\mathcal{P}}(x,y,t,G)}\mathbf{m}(\mathfrak{p})\,, \tag{3.61}$$

where $\widetilde{\mathcal{P}}(x,y,t,G)$ is the set of nice paths from x to y of length t in G. Moreover we set

$$\widetilde{q}(x):=1-\sum_{y\in[n]}\widetilde{P}^t(x,y)\,.$$

Then, for all $\delta>0$, it holds

$$\|P^t(x,\cdot)-\widetilde{\pi}\|_{\mathrm{TV}}\le\sum_{y\in[n]}\left[(1+\delta)\widetilde{\pi}(y)+\frac{\delta}{n}-\widetilde{P}^t(x,y)\right]^+. \tag{3.62}$$

To handle the term in the r.h.s. above, we apply Proposition 3.26 below in order to remove the positive part $[u]^+=\max\{0,u\}$ in (3.62). From this statement, we get that for all $\delta>0$, and w.h.p., (3.62) becomes

$$\sum_{y\in[n]}\left[(1+\delta)\widetilde{\pi}(y)+\frac{\delta}{n}-\widetilde{P}^t(x,y)\right]=2\delta+\widetilde{q}(x), \tag{3.63}$$

It is now sufficient to provide an upper bound on $\widetilde{q}(x)$, uniformly over $x\in V_\varepsilon^\star$. This can be derived by bounding above the probability that some conditions in Definition 3.7 fail. Condition (i) fails, by definition, with quenched probability $\mathcal{Q}_{x,t}\left(\frac{1}{n\log^3 n}\right)$, for all $x\in[n]$.
Condition (ii) holds with quenched probability $1-o_{\mathbb{P}}(1)$ for all $x\in V_\varepsilon^\star$, by Proposition 3.24.
Condition (iii) is satisfied with quenched probability bounded below by $\mathbf{P}_x^G(X_{s+1}\in V_\varepsilon^\star)$. Taking the minimum over $x\in V_\varepsilon^\star$, and thanks to Lemma 3.20, we conclude that (iii) holds with quenched probability $1-o_{\mathbb{P}}(1)$, uniformly for $x\in V_\varepsilon^\star$.
At last, condition (iv) holds w.h.p. for all $x\in[n]$ due to Lemma 3.14.

In conclusion,

$$\max_{x\in V_\varepsilon^\star}\widetilde{q}(x)\le\max_{x\in V_\varepsilon^\star}\mathcal{Q}_{x,t}\left(\frac{1}{n\log^3 n}\right)+o_{\mathbb{P}}(1)\,. \tag{3.64}$$

Note that for $\theta=\frac{1}{n\log^3 n}$ and $t=(1+\beta)\boldsymbol{t}_{\mathrm{ent}}$ condition (3.40) is satisfied with $\rho<1$. Hence, Theorem 3.21 applies to the r.h.s. in the last display, and ends the proof. □

It now remains to state and prove the result that was applied in order to reduce (3.62) to (3.63). Set $\beta=3\gamma=\frac{3\varepsilon}{80}$. In the notation introduced above, it holds the following.

Proposition 3.26. *Let $t = s + h + 1$ with $s = (1-\gamma)\boldsymbol{t}_{\text{ent}}$, $\gamma > 0$ as in Definition 3.7 and $h \equiv h_\varepsilon$ as in (3.17). Then, for all $\delta > 0$,*

$$\mathbb{P}\left(\max_{x \in V_\varepsilon^\star} \widetilde{P}^t(x,y) \le (1+\delta)\widetilde{\pi}(y) + \frac{\delta}{n}, \forall y \in [n]\right) = 1 - o(1).$$

Proof. To prove the statement, we will perform a time-gluing procedure among the first s steps of the walk (which is confined w.h.p. in the tree $\mathcal{T}_x(s)$) and the last h steps (where the path to a target end point y is unique). Thanks to a partial conditioning on the starting and ending subpaths (of length resp. s and h), we will be able to prove a concentration result for the trajectories of length t which will lead to the desired inequality.

Let us stress that the entire strategy closely follows the proof of Proposition 3.6 in Cai et al. [2023], given in the context of directed configuration models, but requires significant adaptations to address the directed Chung–Lu framework. In particular, while the analysis for the directed configuration model relies heavily on combinatorial computations, which reflect the nature of the model where the in- and out-degrees are deterministic, our approach leverages the independence of connections between vertices, along with appropriate concentration inequalities for the in- and out-degrees. This shift is particularly evident in the computations beginning with (3.68).

Given $x \ne y \in [n]$, let $\mathcal{F} = \mathcal{F}(x,y)$ denote the partial environment obtained after the generation of $\mathcal{T}_x(s)$ and $\mathcal{B}_y^-(h)$. Consider κ_x and κ_y to be the number of matchings needed to generate respectively the two subgraphs. It holds $\kappa_x = |\mathcal{T}_x(s)| - 1$ and $\kappa_y \le |\mathcal{B}_y^-(h)| - 1$.

Let $V_\mathcal{F}^-$ denote the set of vertices in $\partial\mathcal{B}_y^-(h)$ such that there exists a unique path of length h to y, and $V_\mathcal{F}^+$ be the set of unmatched vertices at depth s in $\mathcal{T}_x(s)$. Note that, by construction,

$$\sum_{z \in V_\mathcal{F}^+} \mathbf{m}(\mathfrak{p}_{x,z}) \le 1, \tag{3.65}$$

and

$$\sum_{v \in V_\mathcal{F}^-} \mu_{\text{in}}(v)\mathbf{m}(\mathfrak{p}_{v,y}) \le \mu_{\text{in}} P^h(y) = \sum_{v \in [n]} \mu_{\text{in}}(v) P^h(v,y). \tag{3.66}$$

With this notation, we develop $\widetilde{P}^t(x,y)$, the probability to follow a nice path of length t from x to y, as

$$\widetilde{P}^t(x,y) = \sum_{z \in V_\mathcal{F}^+} \sum_{v \in V_\mathcal{F}^-} \mathbf{m}(\mathfrak{p}_{x,z}) \frac{1}{D_z^+} \mathbf{m}(\mathfrak{p}_{v,y}) \mathbf{1}_{\{z \to v\}} \mathbf{1}_{\{\mathfrak{p} \text{ is a nice path}\}}, \tag{3.67}$$

where $\mathfrak{p} = \mathfrak{p}_{x,z} \cup (z,v) \cup \mathfrak{p}_{v,y}$, with a little abuse of notation. Note that, in this representation of $\widetilde{P}^t(x,y)$, the last indicator highlights the validity of conditions (i) and (iv) of definition 3.7 of nice paths, since (ii) and (iii) are satisfied by construction.

We want study the conditional expectation of (3.67) on the partial environment $\mathcal{F}$. By linearity, we are reduced to analyze the random variables $\mathbf{1}_{\{z \to v\}}/D_z^+$ for $z \in V_\mathcal{F}^+$, $v \in$

$V_{\mathcal{F}}^-$. Since the Bernoulli variables $\mathbf{1}_{\{z\to v\}}$ are independent from the partial environment $\mathcal{F}$, it holds

$$\mathbb{E}\left[\frac{\mathbf{1}_{\{z\to v\}}}{D_z^+}\big|\mathcal{F}\right]=p_{z,v}\mathbb{E}\left[\frac{1}{D_z^+}\,\big|\mathcal{F},\mathbf{1}_{\{z\to v\}}=1\right], \tag{3.68}$$

Some of the indicator functions defining the out-degree of z may have been sampled during the generation of the partial environment $\mathcal{F}=\mathcal{F}(x,y)$. However it holds

$$D_z^+\geq\sum_{\substack{w\in[n]:\\(z,w)\notin\mathcal{F}\cup(z,v)}}\mathbf{1}_{\{z\to w\}}+\mathbf{1}_{\{z\to v\}}=:Y_z^v+\mathbf{1}_{\{z\to v\}}\,.$$

Thus, we can write

$$\mathbb{E}\left[\frac{\mathbf{1}_{\{z\to v\}}}{D_z^+}\big|\mathcal{F}\right]\leq p_{z,v}\mathbb{E}\left[\frac{1}{Y_z^v+1}\big|\mathcal{F}\right]=\frac{p_{z,v}}{\mathbb{E}[Y_z^v|\mathcal{F}]}(1+o(1)),$$

where the last equality follows by Remark 3.17. For $y\in[n]$, we then define the events $\mathcal{W}_y:=\{\kappa_y\leq n^{\frac{1}{2}+\varepsilon}\}$ and $\mathcal{W}:=\cap_{y\in[n]}\mathcal{W}_y$. Since $\mathcal{W}\supseteq\mathcal{S}_\varepsilon^-$, where $\mathcal{S}_\varepsilon^-$ is defined in (3.35), by Lemma 3.18 we get

$$\mathbb{P}(\mathcal{W})\geq\mathbb{P}(\mathcal{S}_\varepsilon^-)=1-o(1)\,. \tag{3.69}$$

On $\mathcal{W}_y$, the number of Bernoulli variables removed from D_z^+ in the definition of Y_z^v is at most $n^{\frac{1}{2}+\varepsilon}$. Moreover, thanks to (3.8), the connection parameter is at most $p_{\max}=o(n^{-\frac{1}{2}-\frac{\eta}{7}})$. We assume since now on that $\mathcal{F}\in\mathcal{W}_y$. Then, if $\varepsilon<\eta/7$, we get that $\mathbb{E}[Y_z^v\,|\mathcal{F}]=\mathbb{E}[D_z^+](1+o(1))=\boldsymbol{w}w_z^+\log n/n(1+o(1))$, and consequently we may conclude that

$$\mathbb{E}\left[\frac{\mathbf{1}_{\{z\to v\}}}{D_z^+}\,\big|\mathcal{F}\right]\leq\frac{w_v^-}{\boldsymbol{w}}(1+o(1))=\mu_{\text{in}}(v)(1+o(1)). \tag{3.70}$$

Taking the conditional average in (3.61) and plugging there (3.70), we obtain that

$$\begin{aligned}\mathbb{E}[\widetilde{P}^t(x,y)\,|\,\mathcal{F}\,]&\leq\sum_{z\in V_{\mathcal{F}}^+}\sum_{v\in V_{\mathcal{F}}^-}\mathbf{m}(\mathfrak{p}_{x,z})\mathbb{E}\left[\frac{\mathbf{1}_{\{z\to v\}}}{D_z^+}\,\big|\mathcal{F}\right]\mathbf{m}(\mathfrak{p}_{v,y})\\&\leq\sum_{z\in V_{\mathcal{F}}^+}\sum_{v\in V_{\mathcal{F}}^-}\mathbf{m}(\mathfrak{p}_{x,z})\mu_{\text{in}}(v)\mathbf{m}(\mathfrak{p}_{v,y})(1+o(1))\\&\leq\sum_{v\in V_{\mathcal{F}}^-}\mu_{\text{in}}(v)\mathbf{m}(\mathfrak{p}_{v,y})(1+o(1))\leq\mu_{\text{in}}P^h(y)(1+o(1)),\end{aligned}$$

where the last lines follows from (3.65) and (3.66). This implies that for every $\delta>0$ and for n large enough, it holds

$$\left(1+\frac{\delta}{2}\right)\mathbb{E}\left[\widetilde{P}^t(x,y)\big|\mathcal{F}\right]\leq(1+\delta)\,\mu_{\text{in}}P^h(y)=(1+\delta)\,\widetilde{\pi}(y)\,. \tag{3.71}$$

Let us consider the random variables

$$X_z := \sum_{v\in V_{\mathcal{F}}^-} \mathbf{m}(\mathfrak{p}_{x,z})\frac{1}{D_z^+}\mathbf{m}(\mathfrak{p}_{v,y})\mathbf{1}_{\{z\to v\}}\mathbf{1}_{\{\mathfrak{p} \text{ is a nice path}\}}\,, \qquad z\in V_{\mathcal{F}}^+\,,$$

where $\mathfrak{p} = \mathfrak{p}_{x,z}\cup(z,v)\cup\mathfrak{p}_{v,y}$. These random variables are conditionally independent. Moreover, thanks to condition (i) of Definition 3.7, we have

$$\mathbf{m}(\mathfrak{p}_{x,z})\frac{1}{D_z^+}\mathbf{m}(\mathfrak{p}_{v,y}) \le \frac{1}{n\log^3 n},$$

and thanks to requirement (iv) of Definition 3.7, it holds

$$|\{v\in V_{\mathcal{F}}^- : \mathfrak{p}_{x,z}\cup(z,v)\cup\mathfrak{p}_{v,y} \text{is nice}\}| \le C\log n.$$

Then, X_z is uniformly bounded in $z\in V_{\mathcal{F}}^+$ by the quantity

$$M = M(n) := \frac{C\log n}{n\log^3 n} = \frac{C}{n\log^2 n}.$$

For $a>0$ and M as above, we can apply the Bernstein inequality to the conditional probability measure $\mathbb{P}(\,\cdot\,|\mathcal{F})$, and get

$$\mathbb{P}\left(\widetilde{P}^t(x,y) - \mathbb{E}\big[\widetilde{P}^t(x,y)|\mathcal{F}\big] \ge a\;\middle|\mathcal{F}\right) \le \exp\left(-\frac{a^2}{2M(\mathbb{E}[\widetilde{P}^t(x,y)|\mathcal{F}]+a)}\right). \tag{3.72}$$

Reasoning as in [Bordenave et al., 2019, Prop. 14], we write $r = n\mathbb{E}[\widetilde{P}^t(x,y)|\mathcal{F}]$ and let $a = \frac{\delta}{n}(\frac{r}{2}+1)$. Then the r.h.s. of (3.72) turns to

$$\exp\left(-\frac{\delta^2(r+2)^2}{4Mn(r(2+\delta)+2\delta)}\right) \le \exp\left(-\frac{c(\delta)C}{Mn}\right) = \exp\left(-c(\delta)\log^2 n\right),$$

where $c(\delta)>0$, is obtained optimizing over $r\ge 0$. In this notation, we rewrite (3.72) as

$$\mathbb{P}\left(\widetilde{P}^t(x,y) \ge \left(1+\frac{\delta}{2}\right)\mathbb{E}\left[\widetilde{P}^t(x,y)\middle|\mathcal{F}\right] + \frac{\delta}{n}\;\middle|\mathcal{F}\right) \le \exp\left(-c(\delta)\log^2 n\right). \tag{3.73}$$

In conclusion, by (3.71) and (3.73), we get that, for all $\mathcal{F}\in\mathcal{W}_y$,

$$\mathbb{P}\left(\widetilde{P}^t(x,y) \ge (1+\delta)\widetilde{\pi}(y) + \frac{\delta}{n}\;\middle|\mathcal{F}\right) = \exp\left(-c(\delta)\log^2 n\right) = o(n^{-3}). \tag{3.74}$$

We are almost done. Reasoning as in [Cai et al., 2023, Prop. 3.6], for $x\in V_\varepsilon^\star$ and $y\in[n]$, let

$$\mathcal{Z}_{x,y} := \left\{\widetilde{P}^t(x,y) \ge (1+\delta)\widetilde{\pi}(y) + \frac{\delta}{n}\right\}.$$

With a little abuse of notation we can write

$$\mathbb{P}(\cup_{x\in V_\varepsilon^\star, y\in[n]}\mathcal{Z}_{x,y}\cap\mathcal{W}) \le n^2 \max_{x\in V_\varepsilon^\star, y\in[n]} \mathbb{P}(\mathcal{Z}_{x,y}\cap\mathcal{W}_y) \le n^2 \max_{x\in V_\varepsilon^\star, y\in[n]}\max_{\mathcal{F}\in\mathcal{W}_y}\mathbb{P}(\mathcal{Z}_{x,y}|\mathcal{F})\,,$$

where the last probability is precisely the l.h.s. in (3.74). Then, having in mind (3.69),

$$\mathbb{P}\big(\cap_{x\in V_\varepsilon^\star, y\in[n]} \mathcal{Z}_{x,y}\big) \geq 1-\mathbb{P}(\cup_{x\in V_\varepsilon^\star, y\in[n]}\mathcal{Z}_{x,y}\cap \mathcal{W})-\mathbb{P}(\mathcal{W}^\complement)=1-o(1)\,,$$

which concludes the proof. □

Remark 3.27. *This proof works well also for times t_λ, lying in the critical window of Theorem 3.5, for which $s=(1-4\gamma)\boldsymbol{t}_{\rm ent}(1+o(1))$, as explained in 3.25.*

3.5.2 Lower bound

We now prove Eq. (3.12) of Theorem 3.3. One possible achieve it consists in achieving inequality (3.20), and then applying the law of large number stated in Theorem 3.21. The bored reader can skip to Subsection 3.5.3 for this approach. We present here an alternative proof, in the spirit of Cai et al. [2023], which exploits the equivalent construction of the annealed random walk described in 3.3.1.

The idea of the proof is that, on one hand, the stationary distribution π is w.h.p. well distributed on $[n]$, in a sense that is specified by Lemma 3.28 below (see also the stronger result stated in the Proposition 3.29). On the other hand, after $t=(1-\beta)\boldsymbol{t}_{\rm ent}$ steps, the random walk concentrates on a set of size at most $n^{1-\beta^2}$ which cannot cover the entire graph, and hence the mixing is far to be achieved at this timescale.

Formally, for $\beta\in(0,1)$, let $t=(1-\beta)\boldsymbol{t}_{\rm ent}$ and let $\mathcal{P}^\beta_{x,y}$ denote the set of paths from x to y of lenght t and with probability mass bigger or equal than $1/n^{1-\beta^2}$. An easy check shows that

$$\sum_{y\in[n]}|\mathcal{P}^\beta_{x,y}|\leq n^{1-\beta^2}\,, \tag{3.75}$$

and hence the set $S_x:=\{y\in[n]:\mathcal{P}^\beta_{x,y}\neq\emptyset\}$ satisfies $|S_x|\leq n^{1-\beta^2}$. From the notation of distance in total variation, we can write

$$\min_{x\in[n]}\|P^t(x,\cdot)-\pi\|_{\rm TV}\geq \min_{x\in[n]}\big(P^t(x,S_x)-\pi(S_x)\big)\geq \min_{x\in[n]}P^t(x,S_x)-\max_{x\in[n]}\pi(S_x).$$

Note that, by definition of S_x and of the quenched probability $\mathcal{Q}_{x,t}(\theta)$ in (3.19), it holds that

$$P^t(x,S_x)\geq \mathcal{Q}_{x,t}(n^{1-\beta^2})\,,\qquad \forall x\in[n]\,.$$

We can then apply Theorem 3.21 with $\theta=n^{1-\beta^2}$ and $t=(1-\beta)\boldsymbol{t}_{\rm ent}$, so that the condition $\rho=\lim_{n\to\infty}-\frac{\log\theta}{{\rm H}t}=1+\beta>1$ is satisfied, and conclude that

$$\min_{x\in[n]}P^t(x,S_x)\geq 1-o_{\mathbb{P}}(1)\,.$$

Going back to (3.75), it now remains to show that $\max_{x\in[n]}\pi(S_x)$ is negligible. We stress that, by monotonicity of the total variation distance, we may assume $\beta^2<\eta$, where η is such that (3.6) holds. Then, we can apply the following lemma with $\delta:=\beta^2/6$, which provides the desired estimate and ends the proof of the lower bound. □

Lemma 3.28. *For all $\delta \in (0, \frac{\eta}{6})$, with $\eta \in (0,1)$ as in (3.6), it holds*

$$\mathbb{P}(\forall S \subset [n] \text{ such that } |S| \leq n^{1-6\delta} \,:\, \pi(S) \leq n^{-\delta/2}) = 1 - o(1).$$

Proof. Let us first define, for any $y \in [n]$ and $t' \in \mathbb{N}$,

$$\mu_{t'}(y) := \frac{1}{n} \sum_{x \in [n]} P^{t'}(x, y)\,. \tag{3.76}$$

By the properties of the total variation distance (see [Levin and Peres, 2017, 4.4]), it holds that

$$\|P^{ks}(x, \cdot) - \pi\|_{\mathrm{TV}} \leq (2\|P^{s}(x, \cdot) - \pi\|_{\mathrm{TV}})^{k}\,,$$

for any $k, s \in \mathbb{N}$. Thanks to the upper bound (3.13), it holds $\|P^{2\boldsymbol{t}_{\mathrm{ent}}}(x, \cdot) - \pi\|_{\mathrm{TV}} \leq 1/2e$. Then, choosing $k = \log^2 n$, $s = 2\boldsymbol{t}_{\mathrm{ent}}$, and setting $T = ks$, we get

$$\|P^{T}(x, \cdot) - \pi\|_{\mathrm{TV}} \leq (2\|P^{2\boldsymbol{t}_{\mathrm{ent}}}(x, \cdot) - \pi\|_{\mathrm{TV}})^{\log^{3/2} n} \leq e^{-\log^{3/2} n}\,,$$

which implies

$$\max_{v \in [n]} |\pi(v) - \mu_T(v)| = o(e^{-\log^{3/2}(n)})\,, \tag{3.77}$$

As a consequence, we can prove the thesis for μ_T in place of π.

To prove the statement, it is now sufficient to show that, given $L := \lceil n^{1-6\delta} \rceil$, then

$$\max_{S : |S| = L} \mathbb{P}(\mu_T(S) \geq n^{-\delta}) = o(n^{-L})\,.$$

So let $S \subset [n]$ with $|S| = L$, set $K = \delta^{-1} L$, and consider the annealed measure $\mathbb{P}^{\mathrm{an},K}_{\mathrm{unif}}$ associated to the process $(X^{(k)})_{k \in \{1,\dots,K\}}$ defined as in Subsection 3.3.1, for $T = \log^2 n \cdot \boldsymbol{t}_{\mathrm{ent}}$. For every $j \leq K$, let B_j be the event defined by the following property: the first j walks end in S. It holds

$$\mathbb{E}[\mu_T(S)^K] = \mathbb{P}^{\mathrm{an},K}_{\mathrm{unif}}(B_K) = \mathbb{P}^{\mathrm{an},K}_{\mathrm{unif}}(B_1) \prod_{j=2}^{K} \mathbb{P}^{\mathrm{an},K}_{\mathrm{unif}}(B_j | B_{j-1})\,.$$

Since $6\delta < \eta$, we can apply the same argument used in (3.56), with $p = 2 + 6\delta$, and get

$$\sum_{v \in S} w_v^- = O(n^{\frac{1}{2} - \delta} L^{\frac{1}{2} + \delta}) = O(n^{1 - 3\delta - 6\delta^2})\,,$$

where we used that $\frac{1}{2+6\delta} < \frac{1}{2} - \delta$ and that $|S| = L = \lceil n^{1-6\delta} \rceil$.

Given B_{j-1}, the j-th trajectory can end in S if it replicates from the beginning one of the previous $j-1$ trajectories (this happens with probability at most $\frac{KT}{n}$), or if it enters at least once the set S or the set formed by the $j-1$ trajectories. Non-fresh vertices (i.e., the ones belonging to the previous trajectories) affect only logarithmically the order of L, and the probability of entering S from a fresh vertex at a given step is bounded by

$$\max_{x\in[n]}\sum_{v\in S}p_{x,v}\le M_1\frac{\log n}{n}\sum_{v\in S}w_v^-=o(n^{-3\delta}).$$

Since the j-th trajectory has $T=O(\log^3 n)$ steps, we conclude that

$$\mathbb{P}^{\mathrm{an},K}_{\mathrm{unif}}(B_j|B_{j-1})\le\frac{KT}{n}+T\,o\big(n^{-3\delta}\big)=o(n^{-2\delta}).$$

This proves that $\mathbb{E}[\mu_T(S)^K]=o(n^{-2\delta K})=o(n^{-2L})$. By Markov's inequality, and being $K=\delta^{-1}L$, we obtain

$$\mathbb{P}(\mu_T(S)\ge n^{-\delta})\le\frac{\mathbb{E}[\mu_T(S)^K]}{n^{-L}}=o(n^{-L})\,,$$

and conclude with a union bound on the $O(n^L)$ sets S with $|S|=L$. □

3.5.3 Cutoff window

We are going to provide upper and lower bounds on the total variation distance which appears in the Theorem 3.5. **We first prove the upper bound.**
Recall the notation introduced in the Theorems 3.21, 3.22 and in Eq. (3.39), and take a reference time $t_\lambda:=\boldsymbol{t}_{\mathrm{ent}}+\lambda\mathbf{w}_{\mathrm{ent}}+o(\mathbf{w}_{\mathrm{ent}})$ with $\lambda\in\mathbb{R}$ fixed. Since $\mathbb{V}\mathrm{ar}(S_{t_\lambda})=\sigma^2 t_\lambda$, choosing $\theta=\frac{1}{n}$, it holds that

$$\frac{\mathrm{H}t_\lambda+\log\theta}{\sqrt{\mathbb{V}\mathrm{ar}(S_{t_\lambda})}}=\frac{\lambda\sigma\sqrt{\frac{\log n}{\log\log n}}(1+o(1))}{\sigma\sqrt{\frac{\log n}{\log\log n}}(1+o(1))}\xrightarrow[n\to+\infty]{}\lambda\,,$$

and we are then under the hypothesis (3.48) of Theorem 3.22. Thanks to this result, together with the inequality (3.64), we get that for every $\delta>0$ and w.h.p.

$$\max_{x\in V_\varepsilon^\star}\widetilde{q}(x)\le\int_\lambda^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}\,du+\delta\,.$$

Applying Proposition 3.26, we may conclude that for every $\delta>0$ and w.h.p.

$$\|P^{t_\lambda}(x,\cdot)-\widetilde{\pi}\|_{\mathrm{TV}}\le 2\delta+\widetilde{q}(x)\le\int_\lambda^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}\,du+3\delta\,.$$

For the lower bound, we first observe that, for $\theta\in(0,1)$,

$$P^{t_\lambda}(x,y)\ge\sum_{\mathfrak{p}\in\mathcal{P}(x,y,t_\lambda,G)}\mathbf{m}(\mathfrak{p})\mathbf{1}_{\mathbf{m}(\mathfrak{p})\le\theta}\,.$$

Then, for every distribution ν on $[n]$,

$$\nu(y)-\sum_{\mathfrak{p}\in\mathcal{P}(x,y,t_\lambda,G)}\mathbf{m}(\mathfrak{p})\mathbf{1}_{\mathbf{m}(\mathfrak{p})\le\theta}\le\big[\nu(y)-P^{t_\lambda}(x,y)\big]^+ +\nu(y)\mathbf{1}_{P^{t_\lambda}(x,y)>\theta}.$$

Summing over $y \in [n]$, using that there are less than $1/\theta$ vertices such that $P^{t_\lambda}(x,y) > \theta$, and by the Cauchy-Schwarz inequality, we get

$$\mathcal{Q}_{x,t_\lambda}(\theta) \le \|\nu - P^{t_\lambda}(x,\cdot)\|_{\mathrm{TV}} + \sqrt{\frac{1}{\theta}\sum_{y\in[n]} \nu^2(y)}\,. \tag{3.78}$$

We just need to show that for suitable choices of ν and $\theta \in [0,1]$, (3.78) implies the claimed statement.

1. A quite straightforward proof of this fact can be done under a further assumption on the weigths $(w_x^-)_{x\in[n]}$. Explicitly, let $w^-_{\max}(n) := \max_{x\in[n]} w_x^-$, and assume that

$$w^-_{\max}(n) = o(e^{\sqrt{\log n}})\,. \tag{3.79}$$

Choosing $\nu = \widetilde{\pi}$ and $\theta = w^-_{\max}(n)\frac{\log^4 n}{n}$, we want to show that

$$\frac{1}{\theta}\mathbb{E}\left[\sum_{x\in[n]} \widetilde{\pi}^2(y)\right] = o\left(1\right)\,. \tag{3.80}$$

Then Markov's inequality will be sufficient to conclude that

$$\mathcal{Q}_{x,t_\lambda}(\theta) \le \|\nu - P^{t_\lambda}(x,\cdot)\|_{\mathrm{TV}} + o_{\mathbb{P}}(1)\,,$$

and the desired lower bound will be a consequence of the central limit Theorem 3.22.

To prove (3.80), first note that, since $\widetilde{\pi} = \mu_{\mathrm{in}} P^h$,

$$\mathbb{E}\left[\sum_{x\in[n]} \widetilde{\pi}^2(y)\right] = \mathbb{P}^{\mathrm{unif}}_{\mu_{\mathrm{in}}}(X^{(1)}_h = X^{(2)}_h)\,,$$

where $X^{(1)}$ and $X^{(2)}$ are two random walks as defined in Subsection 3.3.1, and with initial distribution μ_{in}. Thanks to assumptions (3.7) and (3.79), the probability that they start from the same vertex is less than

$$\mu^{\max}_{\mathrm{in}} = o(w^-_{\max}(n)/n) = o(1),$$

On the other hand, the probability that $X^{(2)}$ meets $X^{(1)}$ at a certain step $0 < s \le h$ is less than $(h+1)^2 p_{\max}$, where $p_{\max} = o(w^-_{\max}(n)\log n/n)$. Thanks to the assumption (3.79), we globally get

$$\frac{1}{\theta}\mathbb{E}\left[\sum_{x\in[n]} \widetilde{\pi}^2(x)\right] = \frac{1}{\theta}\mathbb{P}^{\mathrm{unif}}_{\mu_{\mathrm{in}}}(X^{(1)}_h = X^{(2)}_h) = O\left(\frac{1}{\log n}\right),$$

and thus conclude the proof of (3.80) and of the lower bound.

2. To get rid of assumption (3.79), we may proceed in a similar way but choosing $\nu = \pi$ and $\theta = \frac{1}{n}\log^8 n$, so to stay again under the hypothesis (3.48) of Theorem 3.22. To recover the analogue of (3.80), with π instead of $\widetilde{\pi}$, we will need to apply the Proposition 3.29 below. Given this, we can easily recover the estimate (3.80), and then conclude with the application of Theorem 3.22. ☐

Proposition 3.29. *In the above notation and setting, it holds that*

$$\mathbb{E}\left[\sum_{x\in[n]} \pi^2(x)\right] \le C_1 \frac{\log^6 n}{n},$$

where $C_1 > 0$ is the finite constant given in Lemma 3.12.

Proof. Let $T = \log^2 n$. Let $X^{(1)}$ and $X^{(2)}$ be two independent random walks of length T, moving on the same random graph and with initial distribution Unif($[n]$). Note that, according to Subsection (3.3.1), their joint annealed law is equivalently described by the measure $\mathbb{P}^{\mathrm{an},2}_{\mathrm{unif}}$, that for the sake of readability we simply write $\mathbb{P}^{\mathrm{an}}$.

Denoting by μ_T their common distribution at time T, as in (3.76), we can immediately argue from (3.77) that being $T \gg \boldsymbol{t}_{\mathrm{ent}}$, then

$$\mathbb{E}\left[\sum_{x\in[n]} (\pi(x))^2\right] = \mathbb{E}\left[\sum_{x\in[n]} (\mu_T(x))^2\right](1+o(1)) = \mathbb{P}^{\mathrm{an}}(X^{(1)}_T = X^{(2)}_T)(1+o(1)).$$

We then focus on the probability on the r.h.s. of the above display, that will be estimated using similar ideas to those appeared in Lemmata 3.12 and 3.10.

At first, let $\mathcal{T}$ denote the first time such that the trajectory of $X^{(2)}$ meets that of $X^{(1)}$, formally given by $\mathcal{T} := \min\{s > 0 \,:\, \exists u \le T \text{ such that } X^{(2)}_s = X^{(1)}_u\}$, so that

$$\mathbb{P}^{\mathrm{an}}(X^{(1)}_T = X^{(2)}_T) \le \mathbb{P}^{\mathrm{an}}(\mathcal{T} \le T) = \sum_{t=0}^{T} \mathbb{P}^{\mathrm{an}}(\mathcal{T} = t). \tag{3.81}$$

Since the initial measure is uniform over $[n]$, we immediately get that $\mathbb{P}^{\mathrm{an}}(\mathcal{T} = 0) \le \frac{T}{n}$.

For $t = 1, \dots, T$, it is instead convenient to consider the events

$$\mathcal{A}^j_s \equiv \mathcal{A}^{X^{(j)}}_s, \quad \text{for } j = 1,2 \text{ and } s \in \{0, \dots T\},$$

given in (3.25), and to introduce, for any $s,t \in \{0,\dots T\}$, the events

$$\mathcal{B}^{1,2}_{s,t} := \left\{X^{(2)}_v \ne X^{(1)}_u,\ \forall u \in \{0,\dots,s-1\} \text{ and } v \in \{s,\dots,t-1\}\right\},$$

which are analogues of the events defined in (3.29). With this notation, we can first write

$$\mathbb{P}^{\mathrm{an}}(\mathcal{T} = t) \le \mathbb{P}^{\mathrm{an}}(\mathcal{T} = t, \mathcal{A}^2_t) + \mathbb{P}^{\mathrm{an}}((\mathcal{A}^2_t)^{\complement}), \tag{3.82}$$

where $\mathbb{P}^{\rm an}((\mathcal{A}_t^2)^{\complement}) \le \mathbb{P}^{\rm an}((\mathcal{A}_T^2)^{\complement}) \le C_1 \log^4 n/n$ due to Lemma (3.12), and then express the first summand as

$$\mathbb{P}^{\rm an}(\mathcal{T}=t,\mathcal{A}_t^2) = \sum_{s=0}^{T}\sum_{z\in[n]} \mathbb{P}^{\rm an}(X_t^{(2)}=X_s^{(1)}=z,\mathcal{B}_{T,t}^{1,2}\cap\mathcal{A}_t^2)\,. \tag{3.83}$$

Conditioning over the whole trajectory of $X^{(1)}$, we get

$$\begin{aligned}
&\mathbb{P}^{\rm an}(X_t^{(2)}=X_s^{(1)}=z,\mathcal{B}_{T,t}^{1,2}\cap\mathcal{A}_t^2)\\
&=\sum_{\substack{v\in[n]^{T+1}:\\ v_s=z}} \mathbb{P}^{\rm an}\left(X_t^{(2)}=z,\,\mathcal{A}_t^2 | (X_s^{(1)})_{s\le T}=v,\mathcal{B}_{T,t}^{1,2}\right)\mathbb{P}^{\rm an}\left((X_s^{(1)})_{s\le T}=v,\mathcal{B}_{T,t}^{1,2}\right)=\\
&=\sum_{\substack{v\in[n]^{T+1}:\\ v_s=z}} \widetilde{\mathbb{P}}^{\rm an}(X_t=z,\mathcal{A}_t)\,\mathbb{P}^{\rm an}\left((X_s^{(1)})_{s\le T}=v,\mathcal{B}_{T,t}^{1,2}\right)\,,
\end{aligned}$$

where $\widetilde{\mathbb{P}}^{\rm an}(\cdot) := \widetilde{\mathbb{E}}[\mathbf{P}_{\rm unif}^{G}(\cdot)]$ denotes the annealed measure induced by a Chung–Lu probability measure $\widetilde{\mathbb{P}}$ on a graph with vertex-set $[n]\setminus\{v_k\}_{k\in[0,T]\setminus\{s\}}$ and X is a simple random walk with initial uniform distribution. To the sake of readability we do not stress the dependence of $\widetilde{\mathbb{P}}$ on the path v.

Thanks to Lemma 3.10, $\widetilde{\mathbb{P}}^{\rm an}(X_t=z,\mathcal{A}_t) \le \widetilde{\mathbb{P}}^{\rm an}(X_t=z,\mathcal{L}_{t-1}) = \mu_{\rm in}(z)(1+o(1))$ uniformly over the paths $v\in[n]^{T+1}$ so that, inserting this value in the last display, we get

$$\mathbb{P}^{\rm an}(X_t^{(2)}=X_s^{(1)}=z,\mathcal{B}_{T,t}^{1,2}\cap\mathcal{A}_t^2) \le \mu_{\rm in}(z)\,\mathbb{P}^{\rm an}(X_s^{(1)}=z,\mathcal{B}_{T,t}^{1,2})(1+o(1))\,.$$

As a further application of Lemmata 3.10 and 3.12, it holds that

$$\mathbb{P}^{\rm an}(X_s^{(1)}=z,\mathcal{B}_{T,t}^{1,2}) \le \mathbb{P}^{\rm an}(X_s^{(1)}=z) \le \mathbb{P}^{\rm an}(X_s^{(1)}=z,\mathcal{A}_s^1)+\mathbb{P}^{\rm an}((\mathcal{A}_s^1)^{\complement}) \le \mu_{\rm in}(z)(1+o(1))\,,$$

and altogether, going back to Eq. (3.83) and replacing the value of T, we obtain

$$\mathbb{P}^{\rm an}(\mathcal{T}=t,\mathcal{A}_t^2) \le \sum_{s=0}^{T}\sum_{z\in[n]}\mu_{\rm in}(z)^2(1+o(1)) = O\left(\frac{T}{n}\right)\,,$$

where in the last identity we used the approximation (3.31). We conclude that the leading term in (3.82) is indeed provided by $\mathbb{P}^{\rm an}({\mathcal{A}_t^2}^{\complement})$, so that

$$\mathbb{P}^{\rm an}(\mathcal{T}=t) \le C_1\frac{\log^4 n}{n}(1+o(1))\,,$$

which inserted in (3.81) yields the claimed inequality. □

Chapter 4

Mixing trichotomy for the simple random walk on directed block models

This chapter presents the contents of Bianchi et al. [2025]. We characterize the mixing time of the simple random walk on a directed random graph exhibiting a community structure. The first tentative to study the mixing time of random walks on random graphs with many communities is due to Ben-Hamou [2020], who studied the *non-backtracking* random walk on a variant of the configuration model incorporating a 2-community structure. In Hermon et al. [2025], the authors extended the analysis to the simple random walk and allowed multiple communities. These results reveal that the community structure can create a configurational bottleneck in the set of random walk trajectories, depending on the strength of interactions between communities, and disrupt the entropic picture. In fact, the mixing behavior displays a phase transition among a subcritical regime, where the interaction strength is sufficiently high so that the random walk exhibits cutoff at the same time as in the single-community case, and a supercritical regime, where the low interaction strength results in a smooth–rather than abrupt–mixing, and the convergence to equilibrium is driven by the occurrence of the inter-community transitions. In both these works, the intensity of inter-community connections is modeled via a parameter $\alpha \in [0, 1]$, and the critical scaling for this parameter is shown to correspond to the inverse of the entropic time of a single community. Nevertheless, in Ben-Hamou [2020], Hermon et al. [2025] the authors do not attempt a refined analysis of the total variation distance profile in the supercritical regime, being their focus on the cutoff/non-cutoff transition. This chapter makes the heuristic picture discussed so far as explicit as possible by means of a simple (but natural) model. We rely on the results from Bianchi and Passuello [2025], contained in the previous chapter, concentration properties of the vertex degrees, the characterization of the stationary distribution given in Cooper and Frieze [2012], and the properties of quasi-stationary distributions stated in Aldous [1982].

4.1 Setup and results

In this section, we introduce the necessary notation and preliminaries, and we state our main results. We begin in Section 4.1.1 by providing a precise definition of the graph model. Then, in Section 4.1.2, we review key concepts essential for stating and interpreting our main result, Theorem 4.3, which is presented and discussed formally in Section 4.1.3. In Section 4.1.4, we comment on our choice of the graph model and explore potential extensions of the framework considered in this chapter. The proof of the main result is developed in several sections, from Section 4.2 to Section 4.6. A detailed road map that describes the structure of this technical part is provided in Section 4.1.5.

4.1.1 Model

We consider the following model, which we call *Directed Block Model* and denote by $\mathrm{DBM}(n, m, p, \alpha)$:

(1) Consider $m \in \mathbb{N} \setminus \{1\}$ independent directed Erdős–Rényi random graphs with n vertices and connection probability $p \in (0, 1)$, that is, any ordered couple of vertices presents an oriented edge with probability p. We call these graphs $G_1 = (V_1, E_1)$, $\dots$, $G_m = (V_m, E_m)$. More precisely, for each $i \le m$, the vertices in V_i will be labeled by the integers in $[n]$, with the superscript (i) identifying their community of membership.

(2) For each edge of each graph, throw a coin with a success probability $\alpha \in [0, 1/2)$. If it is a head, rewire the edge as follows: if the edge is in the graph G_i with $i \le m$ and it goes from $x^{(i)}$ to $y^{(i)}$, for $x, y \in [n]$, remove the edge $(x^{(i)}, y^{(i)})$, choose $j \in [m] \setminus \{i\}$ uniformly at random, and let the new edge be $(x^{(i)}, y^{(j)})$.

We denote by $G = (V, E)$ such a graph on the mn vertices. Notice that, rather than a technical constraint, the requirement $\alpha < \frac{1}{2}$ is a physical assumption that guarantees some sort of *community structure*, since the majority of edges out-going a given vertex point towards its same community (i.e., are not *rewired*). Throughout the chapter, we will write $\mathbb{P}$ (resp. $\mathbb{E}$) to denote the probability measure (resp. expectation) encoding the randomness of the two-step graph generation process just described. As usual in the random graph literature, we will be interested in the asymptotic regime in which $n \to \infty$, and all the asymptotic notation will refer to that limit. As will become clear soon, we will keep m as a fixed number We will let the parameters p and α depend on n. Since we want to understand how the mixing behavior depends on the relation between α and n, the dependence of α on n will be considered later. We will say that an event occurs with high probability (or simply w.h.p.), if the probability of its occurrence is a function of n that converges to 1 in the limit $n \to \infty$. Besides the usual Landau notation, we will also write that (cfr. p. vii), for two positive sequences a_n and b_n, it holds $a_n \ll b_n$ (resp. $a_n \lesssim b_n$, $a_n \asymp b_n$, and $a_n \sim b_n$) if $\lim_{n\to+\infty} a_n/b_n = \ell$ and $\ell = 0$ (resp. $0 \le \ell < +\infty$, $0 < \ell < +\infty$, and $\ell = 1$).

Assumption 4.1. *We make the following assumptions on the parameters of the model.*

- *We consider $p = \frac{\lambda \log(n)}{n}$ for some $\lambda > 1$ to ensure that each graph is strongly connected with high probability (see Section 4.2 below).*
- *We also assume $\lambda \asymp 1$, to have a logarithmic average degree. Due to the result of Chapter 3, this assumption ensures that the random walk in each of the graphs (before the rewiring) exhibits a cutoff with high probability.*
- *We consider $m \asymp 1$.*

Let us now introduce some further notation. For each vertex $x \in V$, we denote by D_x^+ the out-degree of x, and write

$$D_x^+ = O_x^+ + I_x^+ ,$$

where

$$\begin{aligned} O_x^+ &:= \#\{\text{out-edges of } x \text{ pointing to other graphs}\} \\ I_x^+ &:= \#\{\text{out-edges of } x \text{ not affected by the rewiring}\} \end{aligned}$$

In the same fashion, we can denote by $D_x^- = O_x^- + I_x^-$ the in-degree of $x \in V$.

Notice that while the out-degree D_x^+ is the same before and after the rewiring, the in-degree D_x^- could be different, and for this reason, we introduce the symbol $D_{x,i}^-$, to denote the in-degree of a vertex $x \in V_i$ before the rewiring procedure. It will also be convenient to define the function $c : V \to \{1, \dots, m\}$ that maps each vertex to its community.

We will first consider the random walk on G_i for $i \leq m$, i.e., on a graph *before* the rewiring, and we write $\mathbf{P}_\mu^{G_i}(\cdot)$ for the associated probability measure when the initial position of the walk has distribution μ on V_i. Recall that the latter is a *random* measure, since it depends on the realization of the environment G_i, and for this reason we will refer to $\mathbf{P}_\mu^{G_i}(X_t \in \cdot)$ as *quenched law* at time t. If the graph G_i turns out to be strongly connected and aperiodic (which is the case with high probability), we will call π_i the stationary distribution of the random walk on G_i.

Similarly, we will consider the random walk on the whole graph G (after the rewiring procedure), and we will denote by $\mathbf{P}_\mu^{G}(X_t \in \cdot)$ its law at time t when the initial position of the walk has distribution μ (on V). If $\mu = \delta_x$ for some $x \in V$, we can denote it by $P^t(x, \cdot)$. We denote by π its stationary distribution, if unique. We will show below that, in the setting we are dealing with, the measures π and $(\pi_i)_{i \leq m}$ are w.h.p. unique. To avoid ambiguities, we will conventionally set them to coincide with the uniform distribution (on the corresponding vertex sets) in the unlikely event that they are not uniquely defined.

4.1.2 Preliminaries

Recall that, for $i \leq m$, G_i denotes an Erdős–Rényi random digraph with vertex set V_i and connection probability $p = \lambda \frac{\log n}{n}$, with $\lambda > 1$ and $\lambda \asymp 1$. For any $i \leq m$, we will let P_i denote the transition kernel of the simple random walk (SRW in short) on the digraph G_i, and we will call

$$\mathrm{H} := \mathbb{E}\left[\log(D_x^+ \vee 1)\right] = -\sum_{y \in V_i} \mathbb{E}\left[P_i(x,y)\log(P_i(x,y))\right] \tag{4.1}$$

the average row entropy of P_i. Let us stress that, thanks to the symmetry of the graph model, the above quantity is independent of the choice of $i \le m$ and $x \in V_i$ and thus equivalent to the one in Eq. (3.10). Moreover, a first order estimate of H, for $n \to \infty$, can be easily deduced to be (see Proposition 3.2 in Chapter 3)

$$\mathrm{H} \sim \log\log(n)\,.$$

Let us denote the *entropic time* associated to the entropy in (4.1) with

$$\boldsymbol{t}_{\mathrm{ent}} := \frac{\log(n)}{\mathrm{H}}\,. \tag{4.2}$$

We now re-state, in the new notation, Theorem 3.3 from Chapter 3, which establishes a uniform cutoff for the random walk on G_i, taking place at the entropic time (4.2).

Theorem 4.2. *Let $\beta > 0$ and $\beta \neq 1$. Then for $i \le m$,*

$$\max_{x \in V_i} \left| \|\mathbf{P}_x^{G_i}(X_{\beta \boldsymbol{t}_{\mathrm{ent}}} \in \cdot) - \pi_i\|_{\mathrm{TV}} - \mathbf{1}_{\{\beta<1\}} \right| \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

Theorem 4.2 is in the same spirit as the findings in Bordenave et al. [2018], Cai et al. [2023] for the directed configuration model. In Section 4.5, its proof will be readapted to analyze the random walk on the whole graph $\mathrm{DBM}(n,m,p,\alpha)$ when α is large enough, and to show the occurrence of a similar cutoff behavior as described in Eq. (4.3) below. We point out that its proof does not require an explicit point-wise knowledge of the stationary measures $(\pi_i)_{i \le m}$, in contrast to what usually happens in the reversible setting.

4.1.3 Main results

We are now in a good position to present our main results.

Theorem 4.3. *Let G be a realization of the random digraph $\mathrm{DBM}(n,m,p,\alpha)$ defined in Section 4.1.1, and $\boldsymbol{t}_{\mathrm{ent}}$ the entropic time given in (4.2). The following mixing trichotomy takes place.*

- ***Subcritical case (Fig. 4.1):*** *if $\alpha^{-1} \ll \boldsymbol{t}_{\mathrm{ent}}$ and $\alpha \le \frac{1}{2}$, then, for all $\beta > 0$ with $\beta \neq 1$,*

$$\max_{x \in V} \left| \|\mathbf{P}_x^{G}(X_{\beta \boldsymbol{t}_{\mathrm{ent}}} \in \cdot) - \pi\|_{\mathrm{TV}} - \mathbf{1}_{\{\beta<1\}} \right| \xrightarrow[n\to+\infty]{\mathbb{P}} 0\,. \tag{4.3}$$

- ***Critical case (Fig. 4.1):*** *if $\alpha^{-1} \sim C\boldsymbol{t}_{\mathrm{ent}}$ for some constant $C > 0$, then, for all $\beta > 0$ with $\beta \neq 1$,*

$$\max_{x \in V} \left| \|\mathbf{P}_x^{G}(X_{\beta \boldsymbol{t}_{\mathrm{ent}}} \in \cdot) - \pi\|_{\mathrm{TV}} - \mathbf{1}_{\{\beta<1\}} - \tfrac{m-1}{m} \mathrm{e}^{-\frac{\beta}{C}\frac{m}{m-1}} \mathbf{1}_{\{\beta>1\}} \right| \xrightarrow[n\to+\infty]{\mathbb{P}} 0\,. \tag{4.4}$$

- ***Supercritical case (Fig. 4.2):*** *if* $\alpha^{-1} \gg \boldsymbol{t}_{\rm ent}$ *and* $\alpha^{-1} \ll \lambda n \log(n)$*, then*
  - *(local equilibrium at* $\boldsymbol{t}_{\rm ent}$*) for any* $\beta \neq 1$

$$\max_{x\in V} \left| \|\mathbf{P}_x^G(X_{\beta \boldsymbol{t}_{\rm ent}} \in \cdot) - \pi\|_{\rm TV} - \mathbf{1}_{\{\beta<1\}} - \tfrac{m-1}{m}\mathbf{1}_{\{\beta>1\}} \right| \xrightarrow[n\to+\infty]{\mathbb{P}} 0\,, \tag{4.5}$$

  - *(whole mixing at* α^{-1}*) for any* $\beta > 0$

$$\max_{x\in V} \left| \|\mathbf{P}_x^G(X_{\beta\alpha^{-1}} \in \cdot) - \pi\|_{\rm TV} - \tfrac{m-1}{m} \mathrm{e}^{-\frac{\beta m}{m-1}} \right| \xrightarrow[n\to+\infty]{\mathbb{P}} 0\,. \tag{4.6}$$

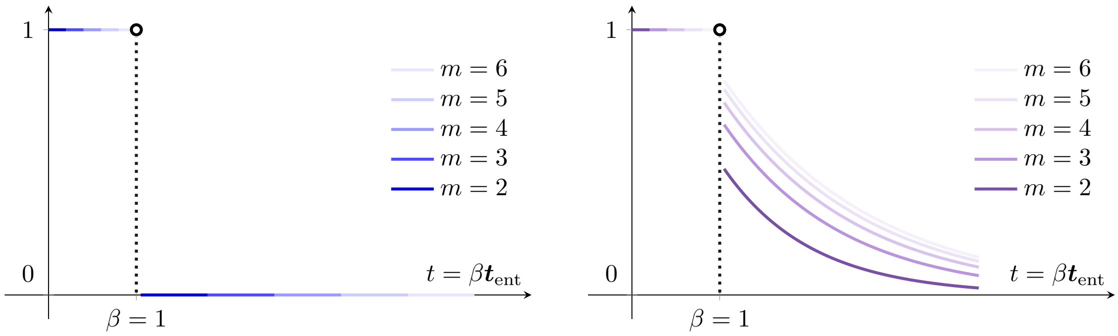

Figure 4.1: Plot of the (theoretical) limiting mixing profile in the subcritical case (left) and critical case (right) with $C = 2$ and $m = 2, 3, 4, 5, 6$.

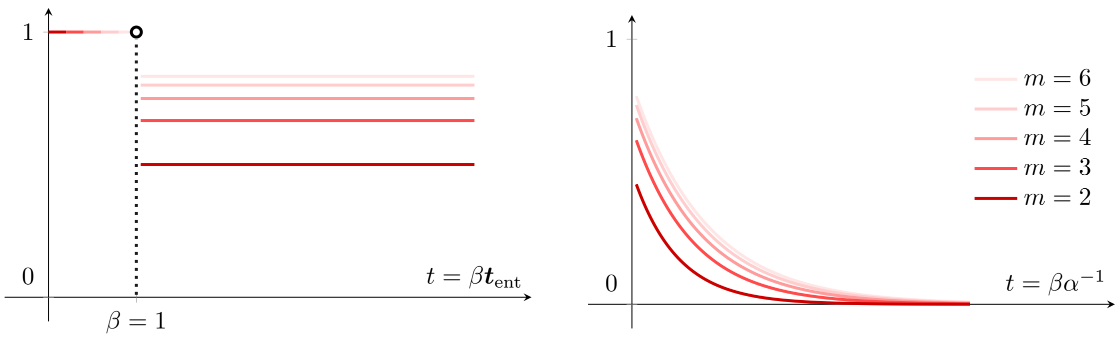

Figure 4.2: Plot of the (theoretical) limiting mixing profile in the supercritical case, with $m = 2, 3, 4, 5, 6$, in the two timescales $t \asymp \boldsymbol{t}_{\rm ent}$ (left) and $t \asymp \alpha^{-1}$ (right).

Theorem 4.3 represents a neat example of the trichotomy phenomenon, and the mechanisms underlying this behavior are easy to read through the mathematical statement and with the help of Figures 4.1 and 4.2. In the subcritical case, the mixing behavior of the walk is totally unaffected by the presence of a macroscopic community structure, since the inter-community jumps occur on a much shorter timescale compared to the entropic time, which represents the time needed to reach the local equilibrium in a single community. In contrast, in the supercritical phase, the random walk abruptly reaches the local equilibrium of the community where it started, in the time $\boldsymbol{t}_{\mathrm{ent}}$. Then, the process essentially behaves as a mean-field random walk on the communities, i.e., as a Markov chain with transition matrix

$$Q = \begin{pmatrix} 1-\alpha & \frac{\alpha}{m-1} & \cdots & \frac{\alpha}{m-1} \\ \frac{\alpha}{m-1} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{\alpha}{m-1} \\ \frac{\alpha}{m-1} & \cdots & \frac{\alpha}{m-1} & 1-\alpha \end{pmatrix}, \tag{4.7}$$

which is in fact easily checked to exhibit the mixing behavior in (4.6). Finally, in the critical case, the two behaviors interpolate, giving rise to the half-cutoff picture, as shown on the right side of Figure 4.1. In light of this interpretation, the reader might already foresee that the proof of Theorem 4.3 essentially reduces to establishing a form of *homogenization* property for the random environment. This will be achieved through different methods in the (sub)critical and supercritical cases, and by identifying two distinct sub-regimes within the supercritical phase. We will provide more details on the organization of the proofs in Section 4.1.5.

4.1.4 Comments on the graph model

As mentioned in the beginning of this chapter, this work aims to establish a simple yet natural framework for studying the mixing trichotomy induced by the presence of a bottleneck in the state space. In this sense, the model we consider is somehow the *minimal one* to this aim, both in terms of notation and of technicalities required for a rigorous proof. However, it would be possible to extend our findings to more general versions of the model presented in Section 4.1.1. In this section, we present some remarks in the direction of such generalizations.

Weakening the assumptions on λ. Let us stress that our assumptions on the connectivity parameter λ are not expected to be *sharp* to prove a mixing trichotomy. For example, $\lambda > 1$ is sufficient but not necessary to guarantee the strong connectivity of communities, and it would be enough to have $(\lambda-1)\log(n) \to +\infty$ as $n \to +\infty$ Cooper and Frieze [2012]. On the other hand, we do not expect $\lambda \asymp 1$ to be sharp either; rather, we anticipate that any $\lambda = n^{o(1)}$ would yield similar results. In fact, as long as $\lambda = n^{o(1)}$, it should be possible to show that the random walk on a single community exhibits a cutoff at the entropic time, which itself forms a divergent sequence in terms of n. However, since this observation is not rigorously stated in any of the aforementioned works, we retain this generalization here.

Less rigid rewirings. We chose to define the rewiring procedure in the model as being strictly tied to the "labels" of the vertices. The same result can be obtained assuming that the edges are rewired uniformly at random. Although this would require only minor adjustments, it would introduce additional notation to describe the rewiring process. To maintain clarity, we choose to omit this generalization. We believe that the argument presented in the chapter is robust to more general non-singular rewiring mechanisms, provided that the rewiring is uniform among the other communities.

Removing dependencies between inter- and intra-community out-degrees. The construction in Section 4.1.1, introduces, for every $x \in V$, a correlation between the random variables O_x^+ and I_x^+, which is not present in the context of the classical stochastic block model. However, by randomizing the number of vertices of the graph, it is possible to remove this kind of dependencies. In particular, assuming that each community has a number $N \sim \mathrm{Pois}(n)$ of vertices, for every $n \in \mathbb{N}$, it holds that $O_x^+ \sim \mathrm{Bin}(D_x^+, \alpha) \sim \mathrm{Pois}(n\alpha p)$ and $I_x^+ \sim \mathrm{Bin}(D_x^+, 1-\alpha) \sim \mathrm{Pois}(n(1-\alpha)p)$. Moreover, the last two variables turn out to be independent as desired. Being $|N - n| = O(\sqrt{n})$ w.h.p., the arguments given in the proofs remain valid and lead to the same results.

Heterogeneous communities. We believe that, with some extra work, the results presented in this work could be generalized to the case in which the communities have different intra-community connectivities, say $\lambda_1, \dots, \lambda_m$, with $\min_{i \le m} \lambda_i > 1$, different rewiring parameters $(\alpha_{i,j})_{i,j \le m}$ with $\max_{j,k \le m} \frac{\alpha_{i,j}}{\alpha_{i,k}} \lesssim 1$, or with different sizes, say $n_1, \dots, n_m$, with $\max_{i,j \le m} \frac{n_i}{n_j} \lesssim 1$. In particular, while the result in Theorem 4.3 should remain unaffected by the heterogeneity of the λ_j's, we believe that in the latter two cases the matrix Q has to be modified, and the (super)critical exponential profile should change to a more general mixing profile, depending on the asymptotic behavior of the corresponding auxiliary Markov chain. Also notice that when the community sizes are different, the rewiring procedure described in Section 4.1.1 is not well defined, and should be suitably modified, e.g., by allowing a rewired edge to point to a uniformly random vertex in the new chosen community, as mentioned above.

Diverging number of communities. Although not the focus of our study, Theorem 4.3 suggests that even when the number of communities is slowly diverging, the characterization of the critical and supercritical regimes should remain unchanged. However, if $m \gg 1$ grows sufficiently fast, the mixing behavior of the model could undergo significant changes. For instance, choosing the parameters (α, m, λ) so that the average number of rewired edges within each community is $\asymp \log(m)$, we expect a cutoff behavior even in the supercritical regime $\alpha^{-1} \gg \boldsymbol{t}_{\mathrm{ent}}$.

In particular, rescaling the time by α, we expect that the dynamics is well approximated by a simple random walk on a coarse-grained graph obtained by collapsing each community to a single point and erasing multiple edges. In that case, the coarse-grained graph is precisely a directed Erdős–Rényi random graph on m vertices in the weakly sparse regime, for which cutoff is now well known to take place. In conclusion, in that setting and for $\alpha^{-1} \gg \boldsymbol{t}_{\mathrm{ent}}$, we expect to observe a cutoff on the timescale $t_{\mathrm{ent},m} = \alpha^{-1} \log(m)/\mathrm{H}_m \gg 1$, where $\mathrm{H}_m \sim \log\log(m)$ is the row-entropy of the random walk on the coarse-grained graph.

4.1.5 Organization of the proof

The proof of Theorem 4.3 is articulated on several parts, depending on the different regimes for the parameter α. We now provide a road map to the forthcoming sections. We start in Section 4.2 by presenting some preliminary facts that will be needed throughout the analysis of both the sub- and supercritical case. In Section 4.3 we deal with the subregion of the supercritical regime ($\alpha \ll \boldsymbol{t}_{\rm ent}$), which we will call *weakly supercritical*. Here $\boldsymbol{t}_{\rm ent} \ll \alpha^{-1} \ll \sqrt{n}\log^{-2}(n)$ and the parameter α is sufficiently large to allow a control on the total variation distance to equilibrium for both the timescales appearing in Eq. (4.5) and Eq. (4.6). The arguments rely on the analysis of the so-called annealed random walk (recalled in Section 4.2). Section 4.4 is the technical core of our work, and it deals with the remaining subregion of the supercritical case, i.e. $\sqrt{n}\log^{-2}(n) \lesssim \alpha^{-1} \ll \lambda n\log(n)$, which we will refer to as *strongly supercritical*. In this case, α is too small and arguments based on *annealed random walks* are doomed to fail, requiring us to employ alternative technical tools to establish a form of *homogenization property* of the graph. This, in turn, offers a level of control over the total variation distance comparable to that achieved in Section 4.3 for the *weakly supercritical* case. Section 4.5 deals with the analogue of the results in the previous sections but for the subcritical and the critical case. Here, the arguments can be seen as an adaptation of those in Bordenave et al. [2018], Cai et al. [2023], and Chapter 3 to the DBM case. Finally, in Section 4.6 we collect all the estimates obtained throughout the chapter and conclude the proof of Theorem 4.3.

4.2 Approximations and auxiliary processes

In this section we aim at presenting some background material and the first statements about the behavior of the random walk on *short timescales*, i.e., for $t \ll \sqrt{n}\log(n)^{-2}$. These results will be used later in the chapter in all three regimes. In particular, in Section 4.2.1 we present some approximation of the stationary distribution of the *local* random walks in terms of the in-degree sequences; in Section 4.2.2 we recall some classical facts about the local structure of a (weakly) sparse graph seen from a vertex; while in Section 4.2.3 we introduce the *annealed random walk* and use it as a tool to provide some key approximation for the law of the inter-community jumps performed by the random walk, see Proposition 4.10.

4.2.1 Approximating the stationary distribution

After the statement of Theorem 4.2, we claimed that the proof does not require explicit knowledge of the stationary distribution. Nevertheless, a uniform first-order approximation for the latter in our setup has been obtained by Cooper and Frieze, as stated by the following theorem. Recall that $D^-_{x,i}$ denotes the in-degree of a vertex $x \in V_i$ before the rewiring procedure, so that the next statement, valid for m disjoint Erdős–Rényi digraph of parameter p, holds.

Theorem 4.4 (Cooper and Frieze [2012]). *The local stationary distributions are approximated at first order and uniformly by the (normalized) local in-degree sequences. In formulas,*

$$\max_{i\le m}\max_{x\in V_i}\left|\frac{\pi_i(x)}{\frac{D^-_{x,i}}{pn^2}}-1\right|\xrightarrow[n\to+\infty]{\mathbb{P}}0\,.$$

The latter result is partially based on the fact that, given the choice $\lambda>1$, uniform concentration properties for the degrees of vertices of G are true.

Proposition 4.5. *There exist two positive constants C_1 and C_2 such that, w.h.p. ,*

$$\max_{i\le m}\max_{x\in V_i}\max\{D^+_x,D^-_x,D^-_{x,i}\}\le C_2\lambda\log(n),$$
$$\min_{i\le m}\min_{x\in V_i}\min\{D^+_x,D^-_x,D^-_{x,i}\}\ge C_1\lambda\log(n)\,.$$

Proof. The thesis follows immediately from concentration bounds for binomial random variables (see e.g. Lemmata 3.14 and 3.15) and an application of the union bound. For instance, the uniform lower bound on $(D^-_{x,i})_{x\in[n]}$ can be derived as follows. For every $t>0$ and $c>0$ it holds

$$\mathbb{P}\left(D^-_{x,i}<c\log(n)\right)=e^{ct\log(n)}\left(1+p(e^{-t}-1)\right)^n\le e^{ct\log(n)+np(e^{-t}-1)},$$

and taking $t=\log\left(\frac{\lambda}{c}\right)$ we have

$$\mathbb{P}\left(D^-_{x,i}<c\log(n)\right)\le e^{\log(n)\left(c\log\left(\frac{\lambda}{c}\right)+c-\lambda\right)}=n^{c\log\left(\frac{\lambda}{c}\right)+c-\lambda}.$$

Since as $c\to0$ the exponent in the r.h.s. converges to $-\lambda<-1$, there exists $c\in(0,\lambda)$ such that the union bound can be performed. To conclude, it suffices to set $C_1=c/\lambda$. □

As a consequence we may conclude that w.h.p. the stationary distribution π_i is uniform, up to corrections bounded away from zero and infinity, and this will be helpful in the characterization of the stationary distribution of the random walk on the whole G. This fact can then be used to deduce the following byproduct of Theorems 4.2, 4.4, and Proposition 4.5, which allows us to control the ℓ^∞-distance to equilibrium of local random walks.

Corollary 4.6. *Let $S=3\boldsymbol{t}_{\rm ent}\log(n)$. Then*

$$\max_{i\le m}\max_{x,y\in V_i}\left|\frac{\mathbf{P}^{G_i}_x(X_S\in y)}{\pi_i(y)}-1\right|=o_{\mathbb{P}}(1).$$

Proof. Recall the subadditivity property of the total variation distance, which implies that

$$\mathbb{P}\left(\max_{i\le m}\max_{x\in V_i}\|\mathbf{P}^{G_i}_x(X_{2k\boldsymbol{t}_{\rm ent}}\in\cdot)-\pi_i\|\le \mathrm{e}^{-k}\right)=1-o(1)\,.$$

Then, by choosing $k=\lfloor\frac{3}{2}\log(n)\rfloor$,

$$\max_{i\le m}\max_{x\in V_i}\left\|\mathbf{P}^{G_i}_x(X_S\in\cdot)-\pi_i\right\|_{\rm TV}=o_{\mathbb{P}}(n^{-1})\,,$$

and the desired result follows by Theorem 4.4 and Proposition 4.5. □

4.2.2 Local neighborhoods are tree-like

The digraphs $(G_i)_{i\leq m}$ and G are w.h.p. locally tree-like in the sense that has been described in Chapter 3: the out-neighborhood $\mathcal{B}_x^+(4\varepsilon \boldsymbol{t}_{\text{ent}})$ of depth $4\varepsilon \boldsymbol{t}_{\text{ent}}$ (growing with n) of any vertex $x \in V$, locally look like trees for small ε, as the next proposition, inherited from Lemmata 3.18 and 3.19, clarifies.

Lemma 4.7. *For a digraph $S = (V, E)$, let $\mathtt{Tx}(S) = |E| - |V| + 1$ denote the tree-excess of S as in Eq. (3.38) and, for $\varepsilon > 0$,*

$$\mathcal{G}_\varepsilon := \{\forall x \in V, \mathtt{Tx}(\mathcal{B}_x^+(4\varepsilon \boldsymbol{t}_{\text{ent}})) < 2\}.$$

Let $(X_t)_{t\in\mathbb{N}}$ denote the SRW on G_i or G and, for $t \leq 2\varepsilon \boldsymbol{t}_{\text{ent}}$, consider the set and related event

$$V_\varepsilon^\star := \left\{x \in V \mid \mathtt{Tx}(\mathcal{B}_x^+(4\varepsilon \boldsymbol{t}_{\text{ent}})) = 0\right\}, \qquad \mathcal{V}_\varepsilon := \{\max_{x\in V} \mathbf{P}_x^G(X_t \notin V_\varepsilon^\star) \leq 2^{-t}\}. \quad (4.8)$$

For a sufficiently small $\varepsilon > 0$ (independent of n), it holds $\mathbb{P}(\mathcal{G}_\varepsilon \cap \mathcal{V}_\varepsilon) = 1 - o(1)$.

Proof. The proof for G_i follows from Lemmata 3.18 and 3.19 from Chapter 3 and can be adapted to G. □

This means that out-neighborhoods of depth $4\varepsilon \boldsymbol{t}_{\text{ent}}$ are at most trees with an extra edge, and if they are not trees, each step of SRW (up to $2\varepsilon \boldsymbol{t}_{\text{ent}}$ steps) w.h.p. halves the quenched probability to have extra edges.

4.2.3 Annealed random walk

A key analysis tool is the so-called *annealed random walk*. Let us recall its definition and outline its construction in our setting.

By annealed random walk one simply means the non-Markovian process in which the underlying random graph is constructed together with the random walk on it. More precisely, at time $t = 0$ all vertices are *unrevealed*, in the sense that its out-neighbors are unknown. Then, for any $t \geq 0$, assuming $X_t = x \in V$,

1. If x is revealed, go directly to step (3).
2. If x is unrevealed, reveal it by generating its out-neighborhood, i.e.:
   - (i) for each vertex $y \in V_{c(x)}$ toss a coin with success probability p;
   - (ii) for each $y \in V_{c(x)}$ which resulted in a success in (i), toss a coin with success probability α;
   - (iii) for each $y \in V_{c(x)}$ which resulted in a success in (i) and a failure in (ii), add an arrow from x to y;
   - (iv) for each $y \in V_{c(x)}$ which resulted in a success both in (i) and (ii), sample a label i u.a.r. in $\{1, \ldots, m\} \setminus \{c(x)\}$ and add an arrow from x to the vertex in G_i corresponding to y (i.e., the vertex with the same label of y).

3. Let X_{t+1} be one of the out-neighbors of x sampled u.a.r..

Clearly, such a random process will eventually stop to reveal new vertices. Moreover, the joint law of the revealed vertices coincides with their joint law under $\mathbb{P}$. The striking feature of the annealed random walk is that, for any initial probability distribution μ on V, its law $\mathbb{P}^{\rm an}_\mu((X_0,\dots,X_t)\in\cdot)$ coincides with the expectation of the quenched law, i.e., for any $t\ge 0$

$$\mathbb{P}^{\rm an}_\mu\big((X_0,\dots,X_t)\in\cdot\big)=\mathbb{E}\big[\mathbf{P}^G_\mu\big((X_0,\dots,X_t)\in\cdot\big)\big]\,.$$

In what follows we will be interested in the event

$$\mathcal{C}_t=\{(X_0,\dots,X_t)\text{ is cycle-free}\}\,,$$

where by cycle-free we simply mean that the random walk does visit any vertex more than once.

Lemma 4.8. *For any $x\in V$, if $t\ll\sqrt{n}$,*

$$\mathbb{P}^{\rm an}_x(X_t=z,\mathcal{C}_t)=\frac{Q^t(c(x),c(z))}{n}\left(1+O\left(\tfrac{t^2}{n}\right)\right)\,,\qquad\forall z\in V\,,$$

where

$$Q^t(i,j)=\frac{1+(m\mathbf{1}_{i=j}-1)(1-\frac{m}{m-1}\alpha)^t}{m},\qquad i,j\le m,\quad t\ge 0,\tag{4.9}$$

Remark 4.9. *Notice that (4.9) corresponds to the t-step transition probability of the Markov chain with transition matrix Q as in (4.7). This can be verified computing the powers of the diagonal form of Q.*

Proof. Fix $z_0=x\in V$ and $z_t=z\in V$. We will say that a sequence $(z_1,\dots,z_{t-1})$ in V is *cycle free* if $(z_0,\dots,z_t)$ does not contain repeated vertices. Start by writing

$$\begin{aligned}\mathbb{P}^{\rm an}_x(X_t=z,\mathcal{C}_t)&=\sum_{\substack{(z_1,\dots,z_{t-1})\in V^{t-1}\\ \text{cycle-free}}}\mathbb{E}\left[\prod_{i=0}^{t-1}\frac{\mathbf{1}_{\{z_i\to z_{i+1}\}}}{D^+_{z_i}}\right]\\&=\sum_{\substack{(z_1,\dots,z_{t-1})\in V^{t-1}\\ \text{cycle-free}}}\prod_{i=0}^{t-1}\mathbb{P}(z_i\to z_{i+1})\mathbb{E}\left[\frac{1}{D^+_{z_i}}\middle|\mathbf{1}_{\{z_i\to z_{i+1}\}}=1\right].\end{aligned}$$

Conditionally on $\mathbf{1}_{\{z_i\to z_{i+1}\}}=1$, the out-degree of z_i has the same distribution as $1+D$, where $D\sim\mathrm{Bin}(n-2,p)$. Then

$$\mathbb{E}\left[\frac{1}{D^+_{z_i}}\middle|\mathbf{1}_{\{z_i\to z_{i+1}\}}=1\right]=\frac{1}{np}\left(1+o\left(n^{-1}\right)\right)\,,$$

and as a consequence

$$\mathbb{P}_x^{\mathrm{an}}(X_t = z, \mathcal{C}_t) = \big(1 + o\big(\tfrac{t}{n}\big)\big) \sum_{\substack{(z_1,\dots,z_{t-1}) \in V^{t-1} \\ \text{cycle-free}}} (np)^{-t} \prod_{i=0}^{t-1} \mathbb{P}(z_i \to z_{i+1})\,.$$

Moreover,

$$\mathbb{P}(z_i \to z_{i+1}) = \begin{cases} p\frac{\alpha}{m-1} & \text{if } c(z_i) \neq c(z_{i+1}) \\ p(1-\alpha) & \text{if } c(z_i) = c(z_{i+1}) \end{cases},$$

and therefore $\mathbb{P}_x^{\mathrm{an}}(X_t = z, \mathcal{C}_t)$ can be expressed as

$$\big(1 + o\big(\tfrac{t}{n}\big)\big) \sum_{\substack{(z_1,\dots,z_{t-1}) \in V^{t-1} \\ \text{cycle-free}}} \frac{1}{n^t} \prod_{i=0}^{t-1} \left(\frac{\alpha}{m-1} \mathbf{1}_{c(z_i) \neq c(z_{i+1})} + (1-\alpha) \mathbf{1}_{c(z_i) = c(z_{i+1})} \right).$$

Since the term within the brackets is bounded above by 1 uniformly in (z_i, z_{i+1}), we could take the sum over all the possible sequences $(z_1, \dots, z_{t-1}) \in V^t$, so having n^{t-1} rather than $(n-2)\cdots(n-t)$ summands. Overall, $\mathbb{P}_x^{\mathrm{an}}(X_t = z, \mathcal{C}_t)$ can be expressed as

$$\left(1 + O\big(\tfrac{t^2}{n}\big)\right) \sum_{(z_1,\dots,z_{t-1}) \in V^{t-1}} \frac{1}{n^t} \prod_{i=0}^{t-1} \left(\frac{\alpha}{m-1} \mathbf{1}_{c(z_i) \neq c(z_{i+1})} + (1-\alpha) \mathbf{1}_{c(z_i) = c(z_{i+1})} \right),$$

and since $t = o(\sqrt{n})$, the prefactor does not change the first order term. Now we notice that the sum over V^{t-1} collapses into a sum over $[m]^{t-1}$. More precisely, using the shorthand notation $c_i = c(z_i)$, for $i \leq t$ and defining

$$J(c_0, \dots, c_t) = \sum_{j=0}^{t-1} \mathbf{1}_{c_{j+1} \neq c_j}\,,$$

we have

$$\mathbb{P}_x^{\mathrm{an}}(X_t = z, \mathcal{C}_t) = \left(1 + O\big(\tfrac{t^2}{n}\big)\right) \frac{1}{n} \sum_{k=0}^{t-1} \sum_{\substack{(c_1,\dots,c_{t-1}) \in [m]^{t-1} \\ J(c_0,\dots,c_t) = k}} \left(\frac{\alpha}{m-1}\right)^k \Big(1-\alpha\Big)^{t-k}.$$

Notice that the quantity expressed by the double sum above can be be interpreted as the probability that the Markov chain with transition matrix as in Eq. (4.7) is in $c_t = c(z)$ at time t. Therefore

$$\mathbb{P}_x^{\mathrm{an}}(X_t = z, \mathcal{C}_t) = \frac{1}{n} Q^t(c(x), c(z)) \left(1 + O\left(\frac{t^2}{n}\right)\right).$$

□

Lemma 4.8 can be strengthened to a statement valid for the quenched law of the SRW $(X_t)_{t\geq 0}$, provided that we consider sufficiently small times, as it will be shown in the next subsection.

Proposition 4.10. *For any $\alpha = \alpha_n \in [0, 1]$, let $1 \ll t \ll \sqrt{n}\log(n)^{-2}$. Then,*

$$\max_{i\leq m}\max_{x\in V}\left|\mathbf{P}_x^G(X_t \in V_i) - Q^t(c(x), i)\right| = o_{\mathbb{P}}(1)\,, \tag{4.10}$$

where, $Q^t(c(x), i)$ is defined as in Eq. (4.9).

Proof. Notice that, since $m \asymp 1$, it suffices to prove (4.10) for a fixed $i \leq m$. We first consider the case in which $\alpha \ll 1$. Let $K = \lfloor\log^2 n\rfloor$ and call $\mathcal{E}$ the event that all the vertices in the graph have out-degree at least $C_1\log(n)$ and $\mathcal{D}$ the event in which each vertex has a tree-excess at most 1 in its out-neighborhood of height 4. Recall that, thanks to Proposition 4.5 and Lemma 4.7, $\mathbb{P}(\mathcal{D}\cap\mathcal{E}) = 1 - o(1)$. Then

$$\mathbb{P}(\mathbf{1}_{\mathcal{E}\cap\mathcal{D}}\mathbf{P}_x^G(X_t \in V_i) \geq Q^t(c(x), i) + \delta) \leq \frac{\mathbb{E}[\mathbf{1}_{\mathcal{E}\cap\mathcal{D}}\mathbf{P}_x^G(X_t \in V_i)^K]}{(Q^t(c(x), i) + \delta)^K}.$$

Consider now K annealed random walks $(X^{(\ell)})_{\ell\leq K}$ starting at some $x \in V$. We let the walks evolve—all starting from x—one after the other, for a time t. Clearly, these walks are not independent, but each is independent of the previous ones *conditionally on the environment discovered so far*. Let us define the family of events $(B_\ell)_{\ell\leq K}$. For each $\ell \leq K$, B_ℓ is the event that:

(i) $X_t^{(j)} \in V_i$ for each $j \leq \ell$;

(ii) all the out-degrees of the vertices visited by the first ℓ walks are at least $C_1\log(n)$ (cf. Prop. 4.5);

(iii) the out-neighborhood of x of height 4 discovered by the first ℓ trajectories has a tree excess at most 1.

With this definitions we have

$$\mathbb{E}[\mathbf{1}_{\mathcal{E}\cap\mathcal{D}}\mathbf{P}_x^G(X_t \in V_i)^K] \leq \mathbb{P}_x^{\mathrm{an}}(B_K) = \mathbb{P}_x^{\mathrm{an}}(B_1)\prod_{\ell=2}^{K}\mathbb{P}_x^{\mathrm{an}}(B_\ell|B_{\ell-1})\,.$$

We start by noting that, given $B_{\ell-1}$, the event B_ℓ is contained in the union of the following events:

(1) For all times $s \leq 4$ the ℓ-th trajectory is always in a vertex already visited by one of the previous walks. Since we are working on the events (ii) and (iii), for any couple (x, y) there are at most 2 paths of length 4 joining x to y and each such path has a weight $\leq (C_1\log(n))^{-4}$. The probability of the event described above is thus upper bounded by

$$8K(C_1\log(n))^{-4} = o(1)\,.$$

(2) The event in (1) does not occur, i.e., there exists some $s \le 4$ such that the walk visits an unvisited vertex, and there exists a time $s' \in (s,t]$ at which the trajectory intersects again one of the previous walks (including itself); this happens with probability less than
$$\frac{Kt^2}{n} = o(1)\,.$$

(3) None of the events above occurs, and at time $s=4$ the walk is out of $V_{c(x)}$. Since $\alpha \ll 1$, this happens with probability at most
$$1-(1-\alpha)^4 = o(1)\,.$$

(4) None of the event above occurs, yet at time t the ℓ-th trajectory is found in V_i; thanks to Lemma 4.8, this happens with probability at most
$$\begin{aligned}\max_{s\le 4}\max_{y\in V_{c(x)}} \mathbb{P}^{\mathrm{an}}_y(X_{t-s}\in V_i, \mathcal{C}_{t-s}) &= \max_{s\le 4} Q^{t-s}(c(x),i)\left(1+O\left(\frac{t^2}{n}\right)\right)\\ &= \max_{s\le 4} Q^{t-s}(c(x),i)+o(1)\\ &= Q^t(c(x),i)+o(1)\,,\end{aligned}$$
where the first identity comes from the fact that, by symmetry, the annealed probability on the l.h.s. is independent of $y\in V_{c(x)}$; the second one follows from the fact that $Q^t(c(x),i)\le 1$ and $t^2 \ll n$; the third one uses that $t \gg 1$.

In conclusion, uniformly over $\ell \le K$ we have $\mathbb{P}^{\mathrm{an}}_x(B_\ell | B_{\ell-1}) = Q^t(c(x),i)+o(1)$. Thanks to the choice of K, we conclude that, being $\delta>0$ fixed,
$$\mathbb{P}(\mathbf{1}_{\mathcal{E}\cap\mathcal{D}}\mathbf{P}^G_x(X_t\in V_i)\ge Q^t(c(x),i)+\delta) \le \left(\frac{Q^t(c(x),i)+o(1)}{Q^t(c(x),i)+\delta}\right)^K = o(n^{-1}).$$
Therefore
$$\begin{aligned}\mathbb{P}\big(\max_{x\in V}\mathbf{P}^G_x(X_t\in V_i)\ge Q^t(c(x),i)+\delta\big) &\le \\ \mathbb{P}(\mathbf{1}_{\mathcal{E}\cap\mathcal{D}}\max_{x\in V}\mathbf{P}^G_x(X_t\in V_i)&\ge Q^t(c(x),i)+\delta)+\mathbb{P}(\mathcal{E}\cup\mathcal{D}) = o(1)\,.\end{aligned}\tag{4.11}$$
To prove a uniform lower bound on $\mathbf{P}^G_x(X_t\in V_i)$, one can consider the events
$$\bar{\mathcal{E}}_{x,i,\delta} = \big\{\mathbf{P}^G_x(X_t\in V_i)\le Q^t(c(x),i)-\delta\big\}$$
and
$$\hat{\mathcal{E}}_{x,j,\delta} = \left\{\mathbf{P}^G_x(X_t\in V_j)\ge Q^t(c(x),j)+\frac{\delta}{m-1}\right\}\,.$$
Clearly, for any $i\le m$, $x\in V_i$ and $\delta>0$, $\cup_{j\ne i}\hat{\mathcal{E}}_{x,j,\delta}\supseteq \bar{\mathcal{E}}_{x,i,\delta}$. Therefore,
$$\begin{aligned}\mathbb{P}\big(\min_{x\in V}\mathbf{P}^G_x(X_t\in V_i)\le Q^t(c(x),i)-\delta\big) &= \mathbb{P}\big(\cup_{x\in V}\bar{\mathcal{E}}_{x,i,\delta}\big)\\ &\le \mathbb{P}\big(\cup_{j\ne i}\cup_{x\in V}\hat{\mathcal{E}}_{x,j,\delta}\big)\\ &\le m\max_{j\le m}\mathbb{P}\big(\cup_{x\in V}\hat{\mathcal{E}}_{x,j,\delta}\big)\,.\end{aligned}\tag{4.12}$$

Since $m \asymp 1$ and since the probability on the r.h.s. of (4.12) coincides, replacing δ with $\frac{\delta}{m-1}$, with the one on the l.h.s. of (4.11), we conclude that

$$\mathbb{P}\big(\min_{x\in V} \mathbf{P}_x^G(X_t \in V_i) \le Q^t(c(x),i) - \delta\big) = o(1)\,. \tag{4.13}$$

Therefore, in the case $\alpha \ll 1$, (4.10) follows from (4.11) and (4.13) and a union bound over $i \le m$.

We are left to consider the case $\alpha \asymp 1$. In this case, $Q^t(j,i) = \frac{1}{m} + o(1)$ for any $t \gg 1$. We can argue as above, but rather than the events (1), (2), (3) and (4), we can consider the events (1), (2) and (4'), where

(4') The events (1) and (2) do not occur, yet at time t the j-th trajectory is found in V_i; thanks to Lemma 4.8, this happens with probability at most

$$\begin{aligned}\max_{s\le 4}\max_{y\in V} \mathbb{P}_y^{\mathrm{an}}(X_t \in V_i, \mathcal{C}_t) &= \max_{s\le 4}\max_{j\le m} Q^{t-s}(j,i)\left(1 + O\left(\frac{t^2}{n}\right)\right)\\ &= \frac{1}{m} + o(1)\,.\end{aligned}$$

This completes the proof. □

4.3 Weakly supercritical regime

In this section and in the following one, we approach the regime $\alpha^{-1} \gg \boldsymbol{t}_{\mathrm{ent}}$.

4.3.1 First jump across two communities

We now consider the first time at which the random walk traverses a rewired edge,

$$\tau_{\mathrm{jump}} = \min\{t > 0 : c(X_t) \neq c(X_{t-1})\}\,. \tag{4.14}$$

Letting $z_0 = x$, and arguing as in the proof of Lemma 4.8 we obtain

$$\begin{aligned}\mathbb{E}[\mathbf{P}_x^G(\tau_{\mathrm{jump}} > t, \mathcal{C}_t)] &= \sum_{\substack{(z_1,\dots,z_t)\in V_{c(x)}^t\\ \text{cycle-free}}} \prod_{i=0}^{t-1} \mathbb{P}(z_i \to z_{i+1})\mathbb{E}\left[\frac{1}{D_{z_i}^+}\middle|\mathbf{1}_{\{z_i\to z_{i+1}\}} = 1\right]\\ &= (1-\alpha)^t\left(1 + O\big(\tfrac{t^2}{n}\big)\right).\end{aligned} \tag{4.15}$$

We now show that, if t is (twice) the entropic time then, w.h.p.—and uniformly over the starting position—the quenched probability to see a jump to another community before t is small. Before stating the proposition, we need a preliminary lemma that serves as a bootstrap for the forthcoming Proposition 4.12.

Lemma 4.11. *If $\alpha \ll 1$, for any constant $a \in \mathbb{N}$,*

$$\max_{x\in V} \mathbf{P}_x^G(\tau_{\mathrm{jump}} \le a) = o_{\mathbb{P}}(1)\,.$$

Proof. Let $\mathcal{E}$ again denote the event that all the vertices in the graph have out-degree at least $C_1 \log(n)$. It is enough to prove that

$$\max_{x\in V} \frac{O_x^+}{D_x^+} = o_{\mathbb{P}}(1)\,. \tag{4.16}$$

To see the validity of the estimate in the latter display, we use Bennett's inequality, which gives, for any fixed $\varepsilon > 0$

$$-\log \mathbb{P}(O_x^+ \geq \varepsilon D_x^+\,,\, \mathcal{E}) \geq -\log \mathbb{P}(O_x^+ \geq \varepsilon C_1 \log(n)) \gtrsim \varepsilon \log(\varepsilon\alpha^{-1}) \log(n)\,.$$

Then (4.16) follows by Proposition 4.5, the fact that $\log(\varepsilon\alpha^{-1}) \gg 1$, and a union bound. □

Proposition 4.12. *Let $\alpha^{-1} \gg \boldsymbol{t}_{\rm ent}$. Then for $T \ll \min\{\alpha^{-1}, \sqrt{n}(\log(n))^{-2}\}$,*

$$\max_{x\in V} \mathbf{P}_x^G(\tau_{\rm jump} \leq T) = o_{\mathbb{P}}(1)\,.$$

Proof. We proceed by the same line of argument as in the proof of Proposition 4.10. Notice that in this case there is no loss of generality in assuming $T \gg 1$, since the (random) map $T \mapsto \mathbf{P}_x^G(\tau_{\rm jump} \leq T)$ is deterministically increasing for any choice of $x \in V$. Let $K = \lfloor \log^2 n \rfloor$ and call $\mathcal{E}$ the event that all the vertices in the graph have out-degree at least $C_1 \log(n)$ and $\mathcal{D}$ the event in which each vertex has a tree-excess at most 1 in its out-neighborhood of height 4. Then

$$\mathbb{P}(\mathbf{1}_{\mathcal{E}\cap\mathcal{D}} \mathbf{P}_x^G(\tau_{\rm jump} \leq T) \geq \delta) \leq \frac{\mathbb{E}[\mathbf{1}_{\mathcal{E}\cap\mathcal{D}} \mathbf{P}_x^G(\tau_{\rm jump} \leq T)^K]}{\delta^K}.$$

Consider now K annealed random walks $(X^{(\ell)})_{\ell\leq K}$ starting at some $x \in V$. Let us define the family of events $(B_\ell)_{\ell\leq K}$. For each $\ell \leq K$, B_ℓ is the event that:

(i) $\inf\{s \geq 1 \mid c(X_s^{(j)}) \neq c(x)\} \leq T$, for all $j \leq \ell$;

(ii) all the out-degrees of the vertices visited by the first ℓ walks are at least $C_1 \log(n)$;

(iii) the out-neighborhood of x of height 4 discovered by the first ℓ trajectories has a tree excess at most 1.

With this definitions we have

$$\mathbb{E}[\mathbf{1}_{\mathcal{E}\cap\mathcal{D}} \mathbf{P}_x^G(\tau_{\rm jump} \leq T)^K] \leq \mathbb{P}_x^{\rm an}(B_K) = \mathbb{P}_x^{\rm an}(B_1) \prod_{\ell=2}^{K} \mathbb{P}_x^{\rm an}(B_\ell | B_{\ell-1})\,.$$

Let us start by noting that given $B_{\ell-1}$, the event B_ℓ is contained in the union of the following four events:

(0) The ℓ-th trajectory jumps to some $y \in V$ such that $c(y) \neq c(x)$ at some time $s \leq 4$ (the probability of this event is $o(1)$ thanks to Lemma 4.11);

(1) For all $s \leq 4$ the ℓ-th trajectory is always in a vertex already visited by one of the previous walks; thanks to (iii), for any y there are at most 2 paths of length 4 joining x to y, moreover, thanks to (ii), this happens with probability at most $8K(C_1 \log(n))^{-4} = o(1)$;

(2) There exists some $s \leq 4$ such that the walk visits an unvisited vertex, and there exists a time $s' \in (s, T]$ at which it intersects again one of the previous walks (including itself); this happens with probability less than $\frac{KT^2}{n} = o(1)$;

(3) None of the event above is verified, yet before time T the ℓ-th trajectory jumps to some $y \in V$ such that $c(y) \neq c(z)$; thanks to Eq. (4.15), this happens with probability at most

$$\mathbb{P}^{\rm an}_x(\tau_{\rm jump} \leq T, \mathcal{C}_T)(1+o(1)) = o(1)\,.$$

In conclusion, we have $\mathbb{P}^{\rm an}_x(B_\ell | B_{\ell-1}) = o(1)$. Thanks to the choice of K, we conclude, for $\delta > 0$ fixed,

$$\mathbb{P}(\mathbf{1}_{\mathcal{E}\cap\mathcal{D}}\mathbf{P}^G_x(\tau_{\rm jump} \leq T) \geq \delta) \leq \left(\frac{o(1)}{\delta}\right)^K = o\left(n^{-c}\right)\,, \qquad \forall c > 0\,.$$

Therefore

$$\mathbb{P}(\max_{x\in V}\mathbf{P}^G_x(\tau_{\rm jump} \leq T) \geq \delta) \leq \mathbb{P}(\mathbf{1}_{\mathcal{E}\cap\mathcal{D}}\max_{x\in V}\mathbf{P}^G_x(\tau_{\rm jump} \leq T) \geq \delta) + o(1) = o(1)\,.$$

□

4.3.2 Local equilibrium: a first timescale

In what follows we will sometimes commit a slight abuse of notation by lifting π_i to a probability measure on the entire vertex set V.

Theorem 4.13. *Let $\alpha^{-1} \gg \boldsymbol{t}_{\rm ent}$. Then, for any fixed $\varepsilon > 0$ and $T \ll \min\{\alpha^{-1}, \sqrt{n}(\log(n))^{-2}\}$,*

$$\max_{i\leq m}\max_{x\in V_i}\max_{t\in[(1+\varepsilon)\boldsymbol{t}_{\rm ent},T]}\|\mathbf{P}^G_x(X_t \in \cdot) - \pi_i\|_{\rm TV} = o_{\mathbb{P}}(1)\,.$$

Proof. Fix $i \leq m$, and notice that there is no loss of generality in assuming $\varepsilon \in (0,1]$. Let $\mathcal{J} = [(1+\varepsilon)\boldsymbol{t}_{\rm ent}, T]$. By the triangle inequality

$$\begin{aligned}
&\max_{t\in\mathcal{J}}\max_{x\in V_i}\|\mathbf{P}^G_x(X_t \in \cdot) - \pi_i\|_{\rm TV}\\
&\quad\leq \max_{t\in\mathcal{J}}\max_{x\in V_i}\|\mathbf{P}^G_x(X_t \in \cdot) - \mathbf{P}^{G_i}_x(X_t \in \cdot)\|_{\rm TV} + \max_{t\in\mathcal{J}}\max_{x\in V_i}\|\mathbf{P}^{G_i}_x(X_t \in \cdot) - \pi_i\|_{\rm TV}\\
&\quad\leq \max_{x\in V_i}\mathbf{P}^G_x(\tau_{\rm jump} \leq T) + \max_{x\in V_i}\|\mathbf{P}^{G_i}_x(X_{(1+\varepsilon)\boldsymbol{t}_{\rm ent}} \in \cdot) - \pi_i\|_{\rm TV}\,,
\end{aligned}$$

where the second inequality can be deduced by coupling the random walk on $(G_i)_{1\leq i\leq m}$ and G in the natural way up to time $\tau_{\rm jump}$. Thanks to Proposition 4.12 and Theorem 4.2, respectively, we may maximize the two terms on the r.h.s. over $i \leq m$ and $x \in V_i$ and get the desired upper bound. □

In conclusion, Theorem 4.13 shows that for $T \ll \min\{\alpha^{-1}, \sqrt{n}\log(n)^{-2}\}$ the dynamics is trapped in the local equilibrium corresponding to the starting community.

4.3.3 Global equilibrium: a second timescale

In the section, we are going to show that each $T \gg \alpha^{-1}$ provides an upper bound on the mixing time of the SRW on our digraph. The claim will be divided into two parts depending on the value of α. In particular, here we will focus on the window $\boldsymbol{t}_{\mathrm{ent}} \ll \alpha^{-1} \ll \sqrt{n}\log(n)^{-2}$, while we postpone to Section 4.4 the discussion of the strongly supercritical regime, where $\sqrt{n}\log(n)^{-2} \lesssim \alpha^{-1} \ll n\lambda\log(n)$. As it will be clear along the proofs, the two regimes require different tools and techniques.

Theorem 4.14. *Let α be such that $\boldsymbol{t}_{\mathrm{ent}} \ll \alpha^{-1} \ll \sqrt{n}\log(n)^{-2}$. If T is such that $T \gg \alpha^{-1}$, then*

$$\max_{x\in V} \left\| \mathbf{P}^G_x(X_T \in \cdot) - \pi \right\|_{\mathrm{TV}} = o_{\mathbb{P}}(1)\,.$$

Moreover, it holds $\left\| \frac{1}{m}\sum_{i=1}^m \pi_i - \pi \right\|_{\mathrm{TV}} = o_{\mathbb{P}}(1)\,.$

Proof. Start by fixing $T \ll n^{\frac{1}{2}}\log(n)^{-2}$. The thesis will hold for general T by monotonicity. Notice also that there is no loss of generality in replacing T by $T + 2\boldsymbol{t}_{\mathrm{ent}}$. We use the deterministic bound

$$\begin{aligned}
\left\| \mathbf{P}^G_x(X_{T+2\boldsymbol{t}_{\mathrm{ent}}} \in \cdot) - \frac{1}{m}\sum_{i=1}^m \pi_i \right\|_{\mathrm{TV}} &= \left\| \sum_{y\in V} P^T(x,y)P^{2\boldsymbol{t}_{\mathrm{ent}}}(y,\cdot) - \frac{1}{m}\sum_{i=1}^m \pi_i \right\|_{\mathrm{TV}} \\
&= \left\| \sum_{i=1}^m \left(\sum_{y\in V_i} P^T(x,y)P^{2\boldsymbol{t}_{\mathrm{ent}}}(y,\cdot) - \frac{\pi_i}{m} \right) \right\|_{\mathrm{TV}} \\
&\le \sum_{i=1}^m \left\| \sum_{y\in V_i} P^T(x,y)P^{2\boldsymbol{t}_{\mathrm{ent}}}(y,\cdot) - \frac{\pi_i}{m} \right\|_{\mathrm{TV}}\,.
\end{aligned} \tag{4.17}$$

Notice that, thanks to Proposition 4.10, we can bound

$$\max_{i\le m}\max_{x\in V} \left| \mathbf{P}^G_x(X_T \in V_i) - \frac{1}{m} \right| = o_{\mathbb{P}}(1). \tag{4.18}$$

Let us fix $i \le m$ and focus on the total variation distance on the r.h.s. of (4.17). We have

$$\begin{aligned}
&\max_{i\le m}\max_{x\in V}\left\|\sum_{y\in V_i}P^T(x,y)P^{2\boldsymbol{t}_{\rm ent}}(y,\cdot)-\frac{\pi_i}{m}\right\|_{\rm TV}\\
&\le \max_{i\le m}\max_{x\in V}\left\|\sum_{y\in V_i}P^T(x,y)\left(P^{2\boldsymbol{t}_{\rm ent}}(y,\cdot)-\pi_i\right)\right\|_{\rm TV}+\max_{i\le m}\max_{x\in V}\left|\mathbf{P}^G_x(X_T\in V_i)-\frac{1}{m}\right| \qquad (4.19)\\
&\le \max_{i\le m}\max_{x\in V}\sum_{y\in V_i}P^T(x,y)\left\|P^{2\boldsymbol{t}_{\rm ent}}(y,\cdot)-\pi_i\right\|_{\rm TV}+o_{\mathbb{P}}(1)\\
&\le \max_{i\le m}\max_{y\in V_i}\left\|P^{2\boldsymbol{t}_{\rm ent}}(y,\cdot)-\pi_i\right\|_{\rm TV}+o_{\mathbb{P}}(1)=o_{\mathbb{P}}(1)\,,
\end{aligned}$$

where both in the second and in the third inequalities we used (4.18), and the last asymptotic bound follows from Theorem 4.13. Plugging (4.19) into (4.17) we deduce that

$$\max_{x\in V}\left\|\mathbf{P}^G_x(X_{T+2\boldsymbol{t}_{\rm ent}}\in\cdot)-\frac{1}{m}\sum_{i=1}^m\pi_i\right\|_{\rm TV}=o_{\mathbb{P}}(1)\,.$$

Then,

$$\begin{aligned}
\left\|\pi-\frac{1}{m}\sum_{i=1}^m\pi_i\right\|_{\rm TV}&\le\sum_{x\in V}\pi(x)\left\|\mathbf{P}^G_x(X_{T+2\boldsymbol{t}_{\rm ent}}\in\cdot)-\frac{1}{m}\sum_{i=1}^m\pi_i\right\|_{\rm TV}\\
&\le\max_{x\in V}\left\|\mathbf{P}^G_x(X_{T+2\boldsymbol{t}_{\rm ent}}\in\cdot)-\frac{1}{m}\sum_{i=1}^m\pi_i\right\|_{\rm TV}=o_{\mathbb{P}}(1)\,.
\end{aligned}$$

This concludes the proof. □

4.4 Strongly supercritical regime

If $\sqrt{n}\log(n)^{-2}\lesssim\alpha^{-1}\ll\lambda n\log(n)$, the relevant timescale, beyond the scale of $\boldsymbol{t}_{\rm ent}$, is $\alpha^{-1}\gtrsim\sqrt{n}\log(n)^{-2}$. We cannot rely on the approximation obtained in Proposition 4.10 to control the random walk behavior of such a timescale. For this *strongly supercritical* regime, we need to generalize the estimates obtained for the *weakly supercritical* regime in Section 4.3 using a different set of tools. In particular, the following two statements provide the analogue of Proposition 4.10 and Theorem 4.14.

Proposition 4.15. *For $\sqrt{n}\log(n)^{-2}\lesssim\alpha^{-1}\ll\lambda n\log(n)$, there exists some $C\gg1$ such that, if $t\le C\alpha^{-1}$,*

$$\max_{i\le m}\max_{x\in V_i}\left|\mathbf{P}^G_x(X_t\in V_i)-Q^t(i,i)\right|=o_{\mathbb{P}}(1)\,,$$

where, $Q^t(c(x),i)$ is defined as in Eq. (4.9).

Theorem 4.16. *For $\sqrt{n}\log(n)^{-2} \lesssim \alpha^{-1} \ll \lambda n \log(n)$, if T is such that $T \gg \alpha^{-1}$, then*

$$\max_{x\in V} \left\| \mathbf{P}_x^G(X_T \in \cdot) - \pi \right\|_{\mathrm{TV}} = o_{\mathbb{P}}(1)\,. \tag{4.20}$$

Moreover, $\left\| \frac{1}{m}\sum_{i=1}^m \pi_i - \pi \right\|_{\mathrm{TV}} = o_{\mathbb{P}}(1)\,.$

As mentioned above, despite the clear analogy with Proposition 4.10 and Theorem 4.14, the proofs of Proposition 4.15 and Theorem 4.16 exploit a different line of argument, based on quasi-stationary distributions, which also shows the emergence of homogenization for such small values of α. In order to facilitate the reading, before entering the details, we provide a brief account of the organization of the rest of this section.

Organization of the section. The rest of the section is divided into five parts. In Section 4.4.1 we introduce the notion of *gates*. In words, a *gate* is a vertex that has an edge that points toward another community. Clearly, in the strongly supercritical regime, gates are rare, since only few vertices have such inter-community connections. With this idea in mind, in Section 4.4.2 we provide a control on the first time the random walk visits the set of gates. In particular, using the framework of *quasi-stationary distributions*, we show that such a first visit is well approximated by an exponential random variable and characterize its expectation. In Section 4.4.3 we use the understanding of the hitting time of the set of gates to couple the random walk with a toy process which enjoys some sort of *renewal* property which makes it simpler to analyze. Finally, in Section 4.4.4 we use this coupling to complete the proof of Proposition 4.15 and Theorem 4.16.

4.4.1 Gates

Fixed a community $i \le m$, the idea is to couple the random walk on G started at some $x \in V_i$ with a simpler process on V_i, up to the first time when the random walk moves to another community. To provide further details, it is necessary to first introduce some additional notation. We we call *gates of* V_i the subset of vertices in V_i having at least a rewired out-edge, i.e.,

$$\mathcal{G}_i := \{y \in V_i \mid O_y^+ > 0\}\,. \tag{4.21}$$

In the regime $\alpha^{-1} \gtrsim \sqrt{n}\log(n)^{-2}$, this set turns out to be small, in the sense that it has a small stationary value, as explained by the next result.

Lemma 4.17. *If* $\alpha^{-1} \gg \lambda \log(n)$,

$$\max_{i\le m} \left| \frac{\pi_i(\mathcal{G}_i)}{\alpha\lambda\log(n)} - 1 \right| = o_{\mathbb{P}}(1)\,.$$

Proof. For $i \le m$, thanks to Theorem 4.4, $\pi_i(\mathcal{G}_i)$ is w.h.p. well approximated by $\sum_{x\in\mathcal{G}_i} \frac{D^-_{x,i}}{n^2 p}$. The latter is a sum of random variables taken on a random set, but it is not difficult to show that it concentrates around its expectation. Indeed, for $k > 0$ it holds

$$\begin{aligned}\mathbb{P}(\big||\mathcal{G}_i| - \mathbb{E}[|\mathcal{G}_i|]\big| > k) \le \frac{\mathbb{V}\mathrm{ar}(|\mathcal{G}_i|)}{k^2} &= \frac{\sum_{x\in V_i} \mathbb{V}\mathrm{ar}(\mathbf{1}_{\{O_x^+>0\}})}{k^2} \\ &= (1+o(1))\frac{n(\lambda\alpha\log(n))(1-\lambda\alpha\log(n))}{k^2},\end{aligned} \tag{4.22}$$

where we used that, for $x \in V_i$, we have $\mathbb{P}(x \in \mathcal{G}_i) = 1-(1-\alpha p)^n = \lambda\alpha\log(n)(1+o(1))$. Choosing k such that $\sqrt{n\lambda\alpha\log(n)} \ll k \ll n\lambda\alpha\log(n)$, we get that

$$\mathbb{P}\left(\big||\mathcal{G}_i| - \lambda n\alpha\log(n)\big| > k\right) = o(1). \tag{4.23}$$

Fixed any $\varepsilon > 0$, using the Chernoff bound, we obtain

$$\begin{aligned}&\mathbb{P}\left(\left\{\Big|\sum_{x\in\mathcal{G}_i} D^-_{x,i} - |\mathcal{G}_i|\lambda\log(n)\Big| > \varepsilon|\mathcal{G}_i|\lambda\log(n)\right\} \cap \left\{\Big||\mathcal{G}_i| - n\alpha\lambda\log(n)\Big| < k\right\}\right) \\ &\le \max_{\delta\in[-\frac{1}{10},\frac{1}{10}]} \mathbb{P}\left(\left|\sum_{y=1}^{n\alpha\lambda\log(n)(1+\delta)} D^-_{y,i} - n\alpha(\lambda\log(n))^2(1+\delta)\right| > (1+\delta)\varepsilon n\alpha(\lambda\log(n))^2\right) \\ &\le 2\exp\left\{-\frac{\varepsilon^2 n\alpha(\lambda\log(n))^2}{4}\right\}.\end{aligned} \tag{4.24}$$

Notice that the event $\{x \in \mathcal{G}_i\}$ does not depend on $D^-_{x,i}$, and hence we can rely on the classical bound for the sums of i.i.d. Bernoulli random variables. Then, choosing $\varepsilon = \frac{1}{\log(n)}$ suffices to make the estimate in (4.24) vanish. The desired result then follows by combining (4.23) and (4.24). □

4.4.2 First visit time to gates

We now introduce the so-called *quasi-stationary distribution*, that is, the long-run distribution of the walk on G_i conditioned to the event of not having hit the set $\mathcal{G}_i$ yet. Let $[P_i]_{\mathcal{G}_i}$ be the sub-Markovian kernel in which the rows and columns indexed by the vertices in $\mathcal{G}_i$ have been removed. Then, called $\mathfrak{l}_i$ the largest eigenvalue of $[P_i]_{\mathcal{G}_i}$, by the Perron-Frobenius theorem there exists a probability distribution $\mu_i^\star$ which is a left eigenvector for $[P_i]_{\mathcal{G}_i}$ associated to the eigenvalue $\mathfrak{l}_i$. In particular, the hitting time of $\mathcal{G}_i$ for the simple random walk on V_i (with kernel P_i) started at $\mu_i^\star$ is *exactly* geometrically distributed with parameter $\mathfrak{l}_i$. This $\mathfrak{l}_i$ can be shown to be < 1 and can also be characterized at first order by the expected hitting time of $\mathcal{G}_i$ starting at π_i. If we denote by $\tau_S := \min\{t \ge 0 : X_t \in S\}$ the hitting time of a set S, these fact are summarized by the following proposition due to Aldous [1982].

Theorem 4.18 (Cf. Prop. A.1 and Lemma A.2 in Quattropani and Sau [2023]). *Let $(W_t)_{t\ge0}$ be a Markov chain on a finite state space Ω with transition matrix Π and unique* *stationary distribution ρ, fully supported on Ω. Let $\partial \in \Omega$ be a* target state. *Then, there exist a unique probability distribution $\mu_\star$ on $\Omega \setminus \partial$ and a unique $\mathfrak{l} \in (0,1)$ such that*

$$\lim_{t\to\infty} \mathbf{P}_\rho\left(W_t = x \mid \tau_\partial > t\right) = \mu_\star(x)\,, \qquad \forall x \in \Omega\setminus\partial\,,$$

and

$$\mathbf{P}_{\mu_\star}\left(\tau_\partial > t\right) = (1-\mathfrak{l})^t\,, \qquad \forall t \geq 0\,.$$

Moreover, observing that $\mathfrak{l} = (\mathbf{E}_{\mu_\star}[\tau_\partial])^{-1}$, it holds

$$\left|\frac{\mathbf{E}_{\mu_\star}[\tau_\partial]}{\mathbf{E}_\rho[\tau_\partial]} - 1\right| = \left|\frac{\mathfrak{l}^{-1}}{\mathbf{E}_\rho[\tau_\partial]} - 1\right| \leq \frac{20}{3}\frac{t_{\mathrm{mix}}(2+\log \mathbf{E}_\rho[\tau_\partial])}{\mathbf{E}_\rho[\tau_\partial]}\,,$$

where

$$t_{\mathrm{mix}} = \inf\left\{t \geq 0 \mid \max_{x\in\Omega} \|\Pi(y,\cdot) - \rho\|_{\mathrm{TV}} \leq (2\mathrm{e})^{-1}\right\}\,.$$

Additionally, consider a sequence of Markov chains with $\Omega_N = [N]$, for $N \in \mathbb{N}$. If

$$\lim_{N\to\infty} T \times \rho(\partial) = 0\,,$$

where

$$T = t_{\mathrm{mix}} \times \log\left(1/\min_{x\in\Omega_N} \rho(x)\right)\,,$$

then, the expected hitting time of ∂, starting at stationarity can be estimated asymptotically by

$$\lim_{N\to\infty} \frac{\mathbf{E}_\rho[\tau_\partial]}{R/\rho(\partial)} = 1\,, \tag{4.25}$$

where

$$R = \sum_{t=0}^{T} \mathbf{P}_\partial(W_t = \partial)\,.$$

In our setting, the *target state* is actually a *target set*, namely $\mathcal{G}_i$. This is not a big issue: indeed, if one considers the Markov kernel $\tilde{P}_i$ on $\tilde{V}_i = (V_i \setminus \mathcal{G}_i) \cup \partial$, where the state ∂ represent the merging of $\mathcal{G}_i$ into a single state and the transitions are set to

$$\tilde{P}_i(x,y) = \begin{cases} P_i(x,y) & x,y \neq \partial \\ \sum_{z\in\mathcal{G}_i} P_i(x,z) & x \neq \partial \text{ and } y = \partial \\ \sum_{z\in\mathcal{G}_i} \frac{\pi_i(z)}{\pi_i(\mathcal{G}_i)} P_i(z,y) & x = \partial \text{ and } y \neq \partial \\ \sum_{z\in\mathcal{G}_i}\sum_{v\in\mathcal{G}_i} \frac{\pi_i(z)}{\pi_i(\mathcal{G}_i)} P_i(z,v) & x = \partial \text{ and } y = \partial \end{cases}\,,$$

then one can realize at once that the restriction of $\tilde{P}_i$ to $V_i \setminus \partial$ coincides with the restriction of P_i to $V_i \setminus \mathcal{G}_i$. Moreover, by direct computation, it is possible to see that, as soon as P_i has a unique stationary distribution π_i, then $\tilde{P}_i$ has a unique stationary distribution $\tilde{\pi}_i$ satisfying

$$\tilde{\pi}_i(x) = \begin{cases} \pi_i(x) & x \neq \partial \\ \pi_i(\mathcal{G}_i) & x = \partial \end{cases}\,.$$

For each $i = 1, \dots, m$, we take the choice $\Pi = \tilde{P}_i$. Replacing $\rho(\partial)$ by $\pi_i(\mathcal{G}_i)$, and R by

$$\tilde{R}_i = 1 + \sum_{t=1}^{\tilde{T}_i} \tilde{P}_i^t(\partial, \partial)\,, \tag{4.26}$$

where

$$\tilde{T}_i = \tilde{t}_{\mathrm{mix},i} \times \log\left(1/\min_{x \in \tilde{V}_i} \tilde{\pi}_i(x)\right),$$

and

$$\tilde{t}_{\mathrm{mix},i} = \inf\left\{t \geq 0 \,\middle|\, \max_{x \in \tilde{V}_i} \|\tilde{P}_i^t(x, \cdot) - \tilde{\pi}_i\|_{\mathrm{TV}} \leq (2\mathrm{e})^{-1}\right\}, \tag{4.27}$$

we may estimate the quantity

$$\mathbf{E}_{\pi_i}[\tau_{\mathcal{G}_i}] = \tilde{\mathbf{E}}_{\tilde{\pi}_i}[\tau_\partial]$$

as in (4.25), and provide the following rewriting of Theorem 4.18, in which we achieve a first-order asymptotic approximation of the rate $\mathfrak{l}$, and which will be proved in the following subsection.

Proposition 4.19. *With high probability, for every $i \leq m$ there exists a probability distribution $\mu_i^\star$, supported on $V_i \setminus \mathcal{G}_i$, and a real number $\mathfrak{l}_i$ such that*

$$\mathbf{P}_{\mu_i^\star}^{G_i}(\tau_{\mathcal{G}_i} > t) = (1 - \mathfrak{l}_i)^t\,,$$

where $\mathfrak{l}_i$ satisfies

$$\mathfrak{l}_i = (1 + o_{\mathbb{P}}(1))\lambda\alpha\log(n)\,. \tag{4.28}$$

In particular

$$\max_{i \leq m} \sup_{t \geq 0} \left|\mathbf{P}_{\mu_i^\star}^{G_i}(\lambda\alpha\log(n) \cdot \tau_{\mathcal{G}_i} > t) - \mathrm{e}^{-t}\right| = o_{\mathbb{P}}(1)\,.$$

Before embarking on the proofs we provide a brief scheme of the upcoming statements. The proof pf Proposition 4.19 is straightforward if we can provide a control on the objects which appeared above. This is done in Lemmata 4.20 and 4.21. To prove the two lemmata, will rely on Lemma 4.22. We will first prove the latter, and then prove Lemmata 4.20 and 4.21.

Lemma 4.20. *Recalling the definition in* (4.27)

$$\mathbb{P}\big(\max_{i \leq m} \tilde{t}_{\mathrm{mix},i} \leq 6\boldsymbol{t}_{\mathrm{ent}} \log(n)\big) = 1 - o(1)\,. \tag{4.29}$$

Lemma 4.21. *Recalling the definition in* (4.26)

$$\max_{i \leq m} \tilde{R}_i = 1 + o_{\mathbb{P}}(1).$$

Proof of Proposition 4.19. Plugging the estimates in Lemmata 4.17, 4.20, and 4.21 into Theorem 4.18 (and recalling Theorem 4.4 and Proposition 4.5), Proposition 4.19 follows at once. □

To conclude, we just need to prove the two lemmata. To this aim and to ease the notation, in what follows we will call

$$\mu_{\mathcal{G}_i}(x) = \frac{\pi_i(x)\mathbf{1}_{x\in\mathcal{G}_i}}{\pi_i(\mathcal{G}_i)}\,, \qquad x \in V_i\,, \tag{4.30}$$

the restriction of π_i to $\mathcal{G}_i$ and we set

$$\mu^{\text{out}}_{\mathcal{G}_i}(x) = \mu_{\mathcal{G}_i} P_i(x)\,, \qquad x \in V_i\,. \tag{4.31}$$

The latter is the distribution after one step starting at a random gate sampled with probability proportional to π_i. With this notation

$$\tilde{R}_i = \sum_{t=0}^{\tilde{T}_i} \mathbf{P}^{G_i}_{\mu_{\mathcal{G}_i}}(\bar{X}_t \in \mathcal{G}_i) = 1 + \sum_{t=1}^{T_i} \mathbf{P}^{G_i}_{\mu_{\mathcal{G}_i}}(\bar{X}_t \in \mathcal{G}_i)\,,$$

where $(\bar{X}_t)_{t\geq 0}$ is the process in which the transition probabilities out of any vertex in $\mathcal{G}_i$ are set to $\mu^{\text{out}}_{\mathcal{G}_i}$, and the other transition probabilities are the same as $\tilde{X}$ (and X). Notice that $\tilde{X}$ is a projection of $\bar{X}$, and clearly,

$$\mathbf{P}^{G_i}_{\mu_{\mathcal{G}_i}}(\bar{X}_t \in \mathcal{G}_i) = \mathbf{P}^{G_i}_{\partial}(\tilde{X}_t = \partial)\,, \qquad \forall t \geq 0\,.$$

The proofs of Lemmata 4.20 and 4.21 rely on the following technical estimate, which will be immediately proved.

Lemma 4.22. *Let*

$$\bar{\tau}_{\mathcal{G}_i} = \inf\left\{t \geq 0 \mid \bar{X}_t \in \mathcal{G}_i\right\} \qquad \textit{and} \qquad \bar{\tau}^+_{\mathcal{G}_i} = \inf\left\{t \geq 1 \mid \bar{X}_t \in \mathcal{G}_i\right\}\,.$$

Then

$$\max_{i\leq m} \mathbf{P}^{G_i}_{\mu_{\mathcal{G}_i}}(\bar{\tau}^+_{\mathcal{G}_i} \leq \log(n)^3) = o_{\mathbb{P}}\big(n^{-1/3}\big)\,. \tag{4.32}$$

Proof of Lemma 4.22. We consider the following partial construction of G_i:

1. For each $x \in V_i$, sample a random variable D^+_x with distribution $\text{Bin}(n-1, p)$.
2. Attach, to each $x \in V_i$, D^+_x tails of arrows.
3. To each tail, attach a $\text{Ber}(\alpha)$ random variable and call *marked* a tail for which such a random variable is 1.

We call Σ the set of all possible realizations generated by the randomness just described. Notice that, in order to end up with a (sub-)graph having the correct distribution, it is enough to select, for each vertex $x \in V_i$ with D^+_x tails the same number of (distinct) vertices in $V_i \setminus \{x\}$. Notice also that the set of gates is fully determined by Σ, since it coincides with the set of vertices having at least one marked tail.

Let $\mathcal{M}$ be the set of probability measures on $\mathcal{G}_i$ defined as

$$\mathcal{M} = \left\{ \mu \in \mathcal{P}(\mathcal{G}_i) \mid \max_{x,y \in \mathcal{G}_i} \frac{\mu(x)}{\mu(y)} \le C \right\} , \tag{4.33}$$

for some bounded $C > 0$ that depends on C_1, C_2 in Proposition 4.5. It holds

$$\mathbb{P}(\mu_{\mathcal{G}_i} \notin \mathcal{M}) = o(1) . \tag{4.34}$$

where $\mu_{\mathcal{G}_i}$ is the probability measure defined in Eq. (4.30). This is an immediate consequence of Theorem 4.4 and Lemma 4.17. Let $\mathcal{F} \subseteq \Sigma$ be the set of realizations such that the event in Proposition 4.5 concerning $(D_x^+)_{x \in V_i}$ is satisfied, and such that $\frac{1}{2} n\alpha\lambda \log(n) \le |\mathcal{G}_i| \le 2n\alpha\lambda \log(n)$. Thanks to Proposition 4.5 and Eq. (4.22) in the proof of Lemma 4.17, it holds

$$\mathbb{P}(\mathcal{F}^{\complement}) = o(1). \tag{4.35}$$

We will later show that

$$\max_{\sigma \in \mathcal{F}} \mathbb{P}\left(\max_{\mu \in \mathcal{M}} \mathbf{P}_{\mu}^{G_i}(\bar{\tau}_{\mathcal{G}_i}^+ \le \log(n)^3) > n^{-1/3} \mid \sigma \right) = o(1) , \tag{4.36}$$

but let us first point out how the desired result can be derived from (4.36):

$$\begin{aligned}
&\mathbb{P}\left(\mathbf{P}_{\mu_{\mathcal{G}_i}}^{G_i}(\bar{\tau}_{\mathcal{G}_i}^+ \le \log(n)^3) > n^{-1/3} \right) \\
&\quad \le \mathbb{P}\left(\mathbf{P}_{\mu_{\mathcal{G}_i}}^{G_i}(\bar{\tau}_{\mathcal{G}_i}^+ \le \log(n)^3) > n^{-1/3}, \mathcal{F} \right) + \mathbb{P}(\mathcal{F}^{\complement}) \\
&\quad \le \mathbb{P}\left(\mathbf{P}_{\mu_{\mathcal{G}_i}}^{G_i}(\bar{\tau}_{\mathcal{G}_i}^+ \le \log(n)^3) > n^{-1/3}, \mathcal{F}, \mu_{\mathcal{G}_i} \in \mathcal{M} \right) + \mathbb{P}(\mu_{\mathcal{G}_i} \notin \mathcal{M}) + \mathbb{P}(\mathcal{F}^{\complement}) \\
&\quad \le \max_{\sigma \in \mathcal{F}} \mathbb{P}\left(\max_{\mu \in \mathcal{M}} \mathbf{P}_{\mu}^{G_i}(\bar{\tau}_{\mathcal{G}_i}^+ \le \log(n)^3) > n^{-1/3} \mid \sigma \right) + \mathbb{P}(\mu_{\mathcal{G}_i} \notin \mathcal{M}) + \mathbb{P}(\mathcal{F}^{\complement}) = o(1) .
\end{aligned}$$

The three terms on the r.h.s. of the latter display vanish due to (4.36), (4.34), and (4.35), respectively. This would complete the proof of (4.32).

We are left to prove (4.36). Observe that if $\mu \in \mathcal{M}$, then

$$\max_{x \in \mathcal{G}_i} \mu(x) \le \min_{y \in \mathcal{G}_i} \mu(y) C \le C / |\mathcal{G}_i|.$$

As a consequence, for any $\sigma \in \mathcal{F}$ and all n large enough, so that $C < \log(n)$,

$$\begin{aligned}
&\mathbb{P}\left(\max_{\mu \in \mathcal{M}} \mathbf{P}_{\mu}^{G_i}(\bar{\tau}_{\mathcal{G}_i}^+ \le \log(n)^3) > n^{-1/3} \mid \sigma \right) \\
&\qquad \le \mathbb{P}\left(\sum_{x \in \mathcal{G}_i} \frac{\mathbf{P}_x^{G_i}(\bar{\tau}_{\mathcal{G}_i}^+ \le \log(n)^3)}{|\mathcal{G}_i|} > \frac{1}{n^{1/3} \log(n)} \mid \sigma \right) .
\end{aligned} \tag{4.37}$$

Write μ_u for the uniform distribution on $\mathcal{G}_i$, and consider the chain $(\bar{\bar{X}})_{t \ge 0}$ in which, when visiting $\mathcal{G}_i$, the chain is instantaneously set at some vertex in $\mathcal{G}_i$ u.a.r.. Clearly,

$$\mathbf{P}_{\mu_u}^{G_i}(\bar{\tau}_{\mathcal{G}_i}^+ \le \log(n)^3) = \mathbf{P}_{\mu_u}^{G_i}(\bar{\bar{\tau}}_{\mathcal{G}_i}^+ \le \log(n)^3) , \tag{4.38}$$

where $\bar{\bar{\tau}}^+_{\mathcal{G}_i}$ is the analogue of $\bar{\tau}^+_{\mathcal{G}_i}$ for the chain $(\bar{\bar{X}}_t)_{t\geq 0}$. By Markov's inequality,

$$\begin{aligned}\max_{\sigma\in\mathcal{F}}\mathbb{P}\left(\mathbf{P}^{G_i}_{\mu_u}(\bar{\bar{\tau}}^+_{\mathcal{G}_i}\leq\log(n)^3)>\frac{1}{n^{1/3}\log(n)}\mid\sigma\right)\\ \leq\log(n)n^{1/3}\max_{\sigma\in\mathcal{F}}\mathbb{E}[\mathbf{P}^{G_i}_{\mu_u}(\bar{\bar{\tau}}^+_{\mathcal{G}_i}\leq\log(n)^3)\mid\sigma]\,.\end{aligned}\tag{4.39}$$

Call $\mathbb{P}^{\rm an}_\sigma$ the measure associated to the annealed walk conditioned to σ and with starting distribution μ_u, which is defined as follows:

(i) Select an element of $\mathcal{G}_i$ u.a.r.;

(ii) select one of its tails u.a.r., match it to a vertex u.a.r. in V_i and move the random walk to such a vertex;

(iii) as soon as the vertex currently visited by the walk is not in $\mathcal{G}_i$ continue in this fashion, but if the selected random tail is already matched, then simply follow the edge while, if it is not matched, match it to a vertex chosen u.a.r. among those that are not currently connected to the present vertex yet, and move the random walk accordingly;

(iv) if instead the vertex visited by the walk at time $r\geq 1$ is some $v\in\mathcal{G}_i$, then select a vertex $w\in\mathcal{G}_i$ with probability μ_u, one of its tail u.a.r., and move the random walk as described above. In this case, set $\bar{\bar{\tau}}^{+,\rm an}_{\mathcal{G}}=r$.

Notice that, for any σ,

$$\mathbb{E}[\mathbf{P}^{G_i}_{\mu_u}(\bar{\bar{\tau}}^+_{\mathcal{G}_i}\leq\log(n)^3)\mid\sigma]=\mathbb{P}^{\rm an}_\sigma(\bar{\bar{\tau}}^{+,\rm an}_{\mathcal{G}}\leq\log(n)^3)\,.\tag{4.40}$$

Let $\mathcal{A}_t$ be the event in which there exists two distinct times $s,s'\in[1,t]$ such that the random walk visits the same vertex at s and s'. Since we are focusing on $\sigma\in\mathcal{F}$, for $t=\log(n)^3$, and large n, it holds

$$\begin{aligned}\mathbb{P}^{\rm an}_\sigma(\bar{\bar{\tau}}^{+,\rm an}_{\mathcal{G}_i}\leq t)&\leq\mathbb{P}^{\rm an}_\sigma(\bar{\bar{\tau}}^{+,\rm an}_{\mathcal{G}_i}\leq t,\mathcal{A}_t^\complement)+\mathbb{P}^{\rm an}_\sigma(\mathcal{A}_t)\\ &\leq t\frac{|\mathcal{G}_i|}{n}+\frac{t^2}{n}\leq\frac{2n\alpha\lambda\log(n)^4+\log(n)^6}{n}=o(n^{-1/2}\log^7(n))\,,\end{aligned}\tag{4.41}$$

and therefore (4.36) follows at once by (4.37), (4.38), (4.39), (4.40) and (4.41). □

We can now prove Lemmata 4.20 and 4.21.

Proof of Lemma 4.20. Consider again the process $(\bar{X}_t)_{t\geq 0}$ in which the transition probabilities out of any vertex in $\mathcal{G}_i$ are given by the distribution $\mu^{\rm out}_{\mathcal{G}_i}$ defined in Eq. (4.31), and denote with $\bar{P}_i$ its transition kernel. Called $\bar{\pi}_i$ the stationary distribution associated to $\bar{P}_i$, and defined

$$\bar{t}_{{\rm mix},i}=\inf\left\{t\geq 0\,\middle|\,\max_{x\in V_i}\|\bar{P}^t_i(x,\cdot)-\bar{\pi}_i\|_{\rm TV}\leq(2{\rm e})^{-1}\right\}\,,$$

it is immediate that $\tilde{t}_{\mathrm{mix},i} \leq \bar{t}_{\mathrm{mix},i}$, since the process $\tilde{X}$ is a projection of $\bar{X}$. Notice also that, thanks to Lemma 4.17,

$$\|\bar{\pi}_i - \pi_i\|_{\mathrm{TV}} \leq \pi_i(\mathcal{G}_i) = o_{\mathbb{P}}(1)\,. \tag{4.42}$$

We are interested in bounding, for $S = 3\boldsymbol{t}_{\mathrm{ent}}\log(n)$,

$$\max_{x\in V_i} \left\|\bar{P}_i^{2S}(x,\cdot) - \bar{\pi}_i\right\|_{\mathrm{TV}}\,.$$

Let now, for $A \subset V_i$ and $x \in V_i$,

$$\begin{aligned}\mathcal{O}_x(A) &:= \left|\mathbf{P}_x^{G_i}(\bar{X}_{2S} \in A, \bar{\tau}_{\mathcal{G}_i} \leq 2S) - \mathbf{P}_x^{G_i}(X^i_{2S} \in A, \tau_{\mathcal{G}_i} \leq 2S)\right| \\ &= \left|\sum_{s=0}^{2S}\sum_{y\in\mathcal{G}_i} \mathbf{P}_x^{G_i}(\bar{\tau}_{\mathcal{G}_i} = s, \bar{X}_s = y)\left[\mathbf{P}^{G_i}_{\mu^{\mathrm{out}}_{\mathcal{G}_i}}(\bar{X}_{2S-s}\in A) - \mathbf{P}_y^{G_i}(X^i_{2S-s}\in A)\right]\right|.\end{aligned}$$

Since $\bar{X}$ can be perfectly coupled, up to the first hitting of the set $\mathcal{G}_i$, with the simple random walk on G_i, that in this proof we explicitly denote by X^i to avoid confusion, we have

$$\begin{aligned}\max_{x\in V_i}\left\|\bar{P}_i^{2S}(x,\cdot) - \bar{\pi}_i\right\|_{\mathrm{TV}} &\leq \max_{x\in V_i}\|P_i^{2S}(x,\cdot) - \pi_i\|_{\mathrm{TV}} + \max_{x\in V_i}\max_{A\subset V_i}\mathcal{O}_x(A) + o_{\mathbb{P}}(1) \\ &= \max_{x\in V_i}\max_{A\subset V_i}\mathcal{O}_x(A) + o_{\mathbb{P}}(1)\,,\end{aligned} \tag{4.43}$$

where we used Corollary 4.6. It holds

$$\mathcal{O}_x(A) \leq \mathcal{O}_x^{\mathrm{small}}(A) + \mathcal{O}_x^{\mathrm{large}}(A)\,, \tag{4.44}$$

where

$$\begin{aligned}\mathcal{O}_x^{\mathrm{small}}(A) &:= \left|\sum_{s=0}^{S}\sum_{y\in\mathcal{G}_i} \mathbf{P}_x^{G_i}(\bar{\tau}_{\mathcal{G}_i} = s, \bar{X}_s = y)\left[\mathbf{P}^{G_i}_{\mu^{\mathrm{out}}_{\mathcal{G}_i}}(\bar{X}_{2S-s}\in A) - \mathbf{P}_y^{G_i}(X^i_{2S-s}\in A)\right]\right|, \\ \mathcal{O}_x^{\mathrm{large}}(A) &:= \left|\sum_{s=S}^{2S}\sum_{y\in\mathcal{G}_i} \mathbf{P}_x^{G_i}(\bar{\tau}_{\mathcal{G}_i} = s, \bar{X}_s = y)\left[\mathbf{P}^{G_i}_{\mu^{\mathrm{out}}_{\mathcal{G}_i}}(\bar{X}_{2S-s}\in A) - \mathbf{P}_y^{G_i}(X^i_{2S-s}\in A)\right]\right|.\end{aligned}$$

On the one hand, for what concerns $\mathcal{O}_x^{\mathrm{large}}(A)$, since $\bar{X}$ is coupled with X^i up to time $\bar{\tau}_{\mathcal{G}_i}$, bounding the quantity between square brackets by 2, using Corollary 4.6 and Lemma 4.17 it follows, for large n,

$$\begin{aligned}\max_{x\in V_i}\max_{A\subset V_i}\mathcal{O}_x^{\mathrm{large}}(A) &\leq 2\max_{x\in V_i}\sum_{s=S}^{2S}\mathbf{P}_x^{G_i}(X^i_s\in\mathcal{G}_i) \\ &\leq 5S\pi_i(\mathcal{G}_i) = o_{\mathbb{P}}(1)\,.\end{aligned} \tag{4.45}$$

On the other hand, $\mathcal{O}^{\rm small}_x(A)$ can be bounded as follows: for any $A\subset V_i$

$$\begin{aligned}\max_{x\in V_i}\mathcal{O}^{\rm small}_x(A)&\le\max_{y\in\mathcal{G}_i}\max_{s\le S}\left|\mathbf{P}^{G_i}_{\mu^{\rm out}_{\mathcal{G}_i}}(\bar X_{2S-s}\in A)-\mathbf{P}^{G_i}_y(X^i_{2S-s}\in A)\right|\\&\le\max_{y\in\mathcal{G}_i}\max_{s\le S}\left|\mathbf{P}^{G_i}_{\mu^{\rm out}_{\mathcal{G}_i}}(\bar X_{2S-s}\in A)-\bar\pi_i(A)\right|+\\&\quad+\max_{y\in\mathcal{G}_i}\max_{s\le S}\left|\mathbf{P}^{G_i}_y(X^i_{2S-s}\in A)-\pi_i(A)\right|+|\pi_i(A)-\bar\pi_i(A)|\\&=\max_{s\le S}\left|\mathbf{P}^{G_i}_{\mu^{\rm out}_{\mathcal{G}_i}}(\bar X_{2S-s}\in A)-\bar\pi_i(A)\right|+o_{\mathbb{P}}(1)\,,\end{aligned}\tag{4.46}$$

where we added and subtracted $\pi_i(A)+\bar\pi_i(A)$, used twice the triangle inequality, Corollary 4.6 and (4.42). Taking the maximum over $A\subset V_i$, we get

$$\begin{aligned}\max_{A\subset V_i}\max_{x\in V_i}\mathcal{O}^{\rm small}_x(A)&\le\max_{s\le S}\left\|\mu^{\rm out}_{\mathcal{G}_i}\bar P_i^{2S-s}-\bar\pi_i\right\|_{\rm TV}+o_{\mathbb{P}}(1)\\&\le\max_{s\le S}\left\|\mu^{\rm out}_{\mathcal{G}_i}P_i^{2S-s}-\pi_i\right\|_{\rm TV}+o_{\mathbb{P}}(1)=o_{\mathbb{P}}(1)\end{aligned}\tag{4.47}$$

where for the last equality we simply applied Theorem 4.2, while for the second inequality, we added and subtracted $\mu^{\rm out}_{\mathcal{G}_i}P_i^{2S-s}-\pi_i$, used the triangle inequality, and used that, thanks to Lemma 4.22,

$$\max_{s\le S}\max_{A\subset V_i}\left|\sum_{y\in V_i}\mu^{\rm out}_{\mathcal{G}_i}(y)\left(\bar P_i^{2S-s}(y,A)-P_i^{2S-s}(y,A)\right)\right|\le\mathbf{P}^{G_i}_{\mu^{\rm out}_{\mathcal{G}_i}}(\bar\tau_{\mathcal{G}_i}<2S)=o_{\mathbb{P}}(1)\,.\tag{4.48}$$

At this point, the desired result, Eq. (4.29), follows by putting together (4.43), (4.44), (4.45), (4.46), (4.47) and (4.48). □

Proof of Lemma 4.21. Follows at once by Lemmata 4.20 and 4.22 and the fact that, by Theorem 4.4 and Proposition 4.5,

$$\log\left(1/\min_{x\in\tilde V_i}\tilde\pi_i\right)=(1+o_{\mathbb{P}}(1))\log(n)\,.$$

□

Remark 4.23. *In the following, it will be useful to consider the transition matrix $\hat P_i$ of a SRW on V_i, where the external out-edges of $\mathcal{G}_i$ (i.e, pointing towards another community) are canceled instead of being rewired. Notice that the very same conclusion of Lemma 4.22, can be obtained replacing P_i, with $\hat P_i$ to define the modified random walk $\bar X$. Indeed, the only point at which the argument differs is the step (iv) in the definition of the annealed walk (right below (4.39)): instead of selecting a tail of $w\in\mathcal{G}_i$ u.a.r., one has to select u.a.r. a* non-marked *tail of $w\in\mathcal{G}_i$. Nevertheless, up to the first return to the set $\mathcal{G}_i$, the two processes coincide.*

Notice also that the property of belonging to $\mathcal{M}$ in (4.33) is stable under small perturbations. Then, Lemma 4.22 holds for any approximation $\check{\mu}$ of $\mu_{\mathcal{G}_i}$ (which is defined in (4.30)) such that $\sup_{x\in\mathcal{G}_i} \check{\mu}(x)/\mu_{\mathcal{G}_i}(x)$ is w.h.p. bounded. Later, our choice of $\check{\mu}$ we will be the restriction to $\mathcal{G}_i$ of the distribution $P_i^s(x,\cdot)$, for some $s \geq \log(n)^2$ and $x \in V_i$. Indeed, Corollary 4.6 ensures that $P_i^s(x,\cdot)$ can be w.h.p. approximated in ℓ^∞-norm by π_i, and then, on this event, its restriction to $\mathcal{G}_i$ is near to $\mu_{\mathcal{G}_i}$ in ℓ^∞-norm.

4.4.3 The coupling

For an arbitrary initial state $x \in V_i$, we will now consider a coupled construction of the simple random walk on G started at x, $(X_t)_{t\geq 0}$, and a toy process $(Y_t)_{t\geq 0}$, which is related to the quasi-stationary behavior of the random walk out of $\mathcal{G}_i$.

Definition 4.24. *We consider the joint Markov processes $(X_t, Y_t)_{t\geq 0}$:*

1. *On the one hand, marginally $(X_t)_{t\geq 0}$ is the simple random walk on G with initial distribution δ_x for some $x \in V_i$;*
2. *on the other hand, marginally $(Y_t)_{t\geq 0}$ is a random walk on G_i, started at the quasi-stationary distribution $\mu_i^\star$, such that, when it hits $\mathcal{G}_i$, at the next step Y is reinitialized at $\mu_i^\star P_i$. Moreover, we enrich such a process by appending a mark, (κ_t, ρ_t), defined as follows: we start by setting $(\kappa_0, \rho_0) = (0,0)$ and, whenever at some time $t > 0$, $Y_t = v$, for some $v \in \mathcal{G}_i$, we set $\kappa_t = \kappa_{t-1}+1$ and toss a coin with probability of success O_v^+/D_v^+: if it is a success, we set $\rho_t = \rho_{t-1} + 1$, otherwise $\rho_t = \rho_{t-1}$.*

We couple the two processes as follows: the coupling is articulated in stages, *and each stage is made by two steps each, (A) and (possibly) (B), where*

- *(A) we use the optimal coupling between $P_i^t(x,\cdot)$ and $\mu_i^\star P_i^t(\cdot)$ up to the first time t such that:*
  - *(i) either $X_t = Y_t$; in this case, we then let the two walks evolve following the same path up to the hitting time of the set $\mathcal{G}_i$.*
  - *(ii) or $X_t \neq Y_t$ and $Y_t \in \mathcal{G}_i$; in this case we declare the coupling as* failed *and continue the construction of the two processes independently.*
  - *(iii) or $X_t \neq Y_t$, and X_t traverses a rewired edge; in this case, we declare the coupling as* failed *and continue the construction of the two processes independently.*
- *(B) In case (i), call $v \in \mathcal{G}_i$ the vertex visited by the two processes and:*
  - *(iv) Toss a coin with success probability O_v^+/D_v^+:*
    - *if it results in a head, we say that the random walks X traverses a rewired edge of v u.a.r., declare the coupling as* successful *and continue the construction of the two processes independently;*

- * *if it results in a tail, then let the random walk X choose one of the internal out-edges of v u.a.r., and let the random walk Y choose a new starting point according to $\mu_i^\star P_i$. After that, repeat stage (A).*

We will call $\hat{\mathbf{P}}^{G_i}_{x,\mu_i^\star}$ the probability measure associated to the coupling just described.

In short, the random iterative coupling with measure $\hat{\mathbf{P}}^{G_i}_{x,\mu_i^\star}$ is successful if X_t and Y_t meet before hitting $\mathcal{G}_i$ at any iteration, until, moving together, they traverse a rewired edge. We now state that this happens with high quenched probability.

Proposition 4.25. *Let $\mathcal{S}$ denote the event that the coupling in Definition 4.24 is successful. Then*

$$\max_{i\le m}\max_{x\in V_i}\hat{\mathbf{P}}^{G_i}_{x,\mu_i^\star}(\mathcal{S})=1-o_{\mathbb{P}}(1)\,.$$

Before presenting the proof, let us point out that on the event $\mathcal{S}$, the time $\tau_{\rm jump}$ at which the random walk does its first inter-community jump coincides with a much simpler random time, namely

$$\tau_\rho=\inf\{t\ge 0\mid \rho_t>0\}\,. \tag{4.49}$$

Thanks to the fact that $(Y_t)_{t\ge0}$ is reinitialized to $\mu_i^\star P_i$ any time it hits $\mathcal{G}_i$, and since $\mu_i^\star P_i(\mathcal{G}_i)=\mathfrak{l}_i$, we have that $(\kappa_t-\kappa_{t-1})_{t\ge1}$ are i.i.d. Bernoulli random variables with parameter $\mathfrak{l}_i$. Therefore, the quantities

$$\sigma_k=\inf\{t\ge0\mid\kappa_t=k\}-\inf\{t\ge0\mid\kappa_t=k-1\}\,,\qquad k\ge1\,,$$

are i.i.d. geometrically distributed with parameter $\mathfrak{l}_i$. Since, thanks to Proposition 4.19, $\mathfrak{l}_i=(1+o_{\mathbb{P}}(1))\alpha\lambda\log(n)$, as an immediate corollary, we have the following estimate, which will be useful later.

Corollary 4.26. *For every $k\ge0$ it holds*

$$\max_{i\le m}\hat{\mathbf{P}}^{G_i}_{\mu_i^\star}\left(\sigma_k<n^{\frac13}\right)\le n^{-\frac17}\,.$$

Proof. It holds

$$\hat{\mathbf{P}}^{G_i}_{\mu_i^\star}\left(\sigma_k<n^{\frac13}\right)=1-(1-\mathfrak{l}_i)^{n^{\frac13}}\sim1-\exp\left(-n^{\frac13}\alpha\lambda\log n\right).$$

Since $\alpha\ll n^{-1/2}\log^2(n)$, then the latter decays as $n^{\frac13}\alpha\log n\ll n^{-\frac17}$. □

Notice that if in stage (B) we had thrown, at each time t that $\kappa_t-\kappa_{t-1}=1$, a coin with success probability $(\lambda\log(n))^{-1}$, then we could immediately deduce that the time of the first success is asymptotically exponentially distributed with rate α. Of course, for the process $(Y_t)_{t\ge0}$, "the probability of success“ is not $(\lambda\log(n))^{-1}$: it depends on the gate visited by the process (through its out-degree and the number of its rewired out-edges). The next proposition shows that, thanks to a sort of *homogenization property*, the toy model just described actually captures the correct picture.

Proposition 4.27. *Recalling the definition of τ_ρ in (4.49)*

$$\max_{i\le m}\sup_{t\ge 0}\left|\hat{\mathbf{P}}^{G_i}_{\mu_i^\star}(\alpha\tau_\rho > t) - \mathrm{e}^{-t}\right| = o_{\mathbb{P}}(1)\,. \tag{4.50}$$

Being the coupling w.h.p. successful by Proposition 4.25, as a consequence of Proposition 4.27 we deduce the following.

Corollary 4.28. *Recalling the definition of (4.14),*

$$\max_{i\le m}\max_{x\in V_i}\sup_{t\ge 0}\left|\mathbf{P}^{G}_{x}(\alpha\tau_{\text{jump}} > t) - \mathrm{e}^{-t}\right| = o_{\mathbb{P}}(1)\,.$$

Proof of Proposition 4.27. Fix $i \le m$. The proof is articulated in five steps:

[1] First, we show that the density of gates in $\mathcal{G}_i$ which have a out-degree different (at first order) than $\lambda\log(n)$ is at most $\log(n)^{-2}$;

[2] Then we show that the same is true for the gates $x \in \mathcal{G}_i$ which have $O_x^+ \ge 2$;

[3] Then we show that the distribution of the first element of $\mathcal{G}_i$ visited by the random walk initialized at $\mu_i^\star$ is essentially uniform;

[4] Call *nice* the subset of gates which do not have the properties in step **[1]** and **[2]**. We show that the first $\log(n)^{3/2}$ visits to $\mathcal{G}_i$ occur at *nice* gates;

[5] We wrap up the argument developed in the previous steps to conclude the validity of (4.50).

Step [1]. Call

$$\mathcal{E}_\varepsilon = \left\{\,\#\left\{x\in\mathcal{G}_i : \left|D_x^+ - \lambda\log(n)\right| \ge \varepsilon\lambda\log(n)\right\} > |\mathcal{G}_i|\log(n)^{-2}\right\}\,.$$

As mentioned, we want to show that $\mathbb{P}(\mathcal{E}_\varepsilon) = o(1)$, for some $\varepsilon = o(1)$. Notice the distribution of the out-degree of x conditioned on the event $\{x \in \mathcal{G}_i\}$ can be written explicitly as follows: for any $j \in [1, n]$

$$\begin{aligned}\mathbb{P}(D_x^+ = j | x\in\mathcal{G}_i) &= \frac{\mathbb{P}(x\in\mathcal{G}_i|D_x^+ = j)\mathbb{P}(D_x^+ = j)}{\mathbb{P}(x\in\mathcal{G}_i)}\\ &= \frac{(1-(1-\alpha)^j)\mathbb{P}(D_x^+ = j)}{\sum_{k=1}^n(1-(1-\alpha)^k)\mathbb{P}(D_x^+ = k)}\\ &= \frac{(1-(1-\alpha)^j)\mathbb{P}(D_x^+ = j)}{1-(1-\alpha p)^n} \le 2\mathbb{P}(D_x^+ = j)\,.\end{aligned}$$

Since by the Chernoff bound it holds

$$\mathbb{P}\big(|D_x^+ - \lambda\log(n)| > \varepsilon\log(n)\big) \le \exp\left(-\tfrac{1}{3}\varepsilon^2\lambda\log(n)\right)\,,$$

we get

$$\mathbb{P}\big(|D_x^+ - \lambda\log(n)| > \varepsilon\log(n) \mid x\in\mathcal{G}_i\big) \le 2\exp\left(-\tfrac{1}{3}\varepsilon^2\lambda\log(n)\right) .$$

Now, notice that, thanks to Eq. (4.23),

$$\mathbb{P}(\mathcal{E}_\varepsilon) = \mathbb{P}\big(\mathcal{E}_\varepsilon \cap \{|\mathcal{G}_i|/(n\alpha\lambda\log(n)) \in [1/2,2]\}\big) + o(1) .$$

On the event $\{|\mathcal{G}_i|/(n\alpha\lambda\log(n)) \in [1/2,2]\}$, we can use the union bound and the fact that $\log\binom{N}{m} = m\log(\frac{N}{m})(1+o(1)) \le 2m\log(\frac{N}{m})$, for $m \ll N$, to get

$$\begin{aligned}
&\mathbb{P}\big(\mathcal{E}_\varepsilon \cap \{|\mathcal{G}_i|/(n\alpha\lambda\log(n)) \in [1/2,2]\}\big)\\
&\quad\le \max_{C\in[\frac12,2]} \binom{2n\alpha\lambda\log(n)}{Cn\alpha\lambda\log(n)^{-1}} \left(2\exp\left(-\tfrac{1}{3}\varepsilon^2\lambda\log(n)\right)\right)^{Cn\alpha\lambda\log(n)^{-1}}\\
&\quad\lesssim \max_{C\in[\frac12,2]} \exp\left(2\frac{Cn\alpha\lambda}{\log(n)}\log\left(\frac{2}{C}\log(n)^2\right)\right)\left(2\exp\left(-\tfrac{1}{3}\varepsilon^2\lambda\log(n)\right)\right)^{\frac{Cn\alpha\lambda}{\log(n)}}\\
&\quad\lesssim \max_{C\in[\frac12,2]} \exp\left(2Cn\alpha\lambda\frac{\log(\log(n)^2)}{\log(n)}\right)2^{\frac{Cn\alpha\lambda}{\log(n)}}\exp\left(-\tfrac{C}{3}\varepsilon^2\lambda^2 n\alpha\right)\\
&\quad\le \exp\left(n\alpha\lambda\left(8\frac{\log\log(n)}{\log(n)} + \frac{2\log(2)}{\log(n)} - \varepsilon^2\frac{6}{\lambda}\right)\right) .
\end{aligned}$$

Therefore, we can choose, for example, $\varepsilon = \log(n)^{-1/3}$, and have

$$\mathbb{P}(\mathcal{E}_\varepsilon) = o(1) . \tag{4.51}$$

Step [2]. Call $\mathcal{G}_{i,\mathrm{bad}} = \{x\in\mathcal{G}_i \mid O_x^+ \ge 2\}$, and consider the event

$$\mathcal{R} = \left\{|\mathcal{G}_{i,\mathrm{bad}}| > |\mathcal{G}_i|\log(n)^{-2}\right\} .$$

To show that the latter has a vanishingly small probability it is enough to realize that

$$\mathbb{E}[|\mathcal{G}_{i,\mathrm{bad}}|] \le n\alpha^2\lambda^2\log(n)^2 .$$

Indeed, by Markov inequality, and recalling $\alpha \ll n^{-1/2}\log(n)^2$,

$$\begin{aligned}
\mathbb{P}(\mathcal{R}) &= \mathbb{P}\left(\mathcal{R}\cap\{|\mathcal{G}_i| \ge \tfrac12\mathbb{E}[|\mathcal{G}_i|]\}\right) + o(1)\\
&\le \frac{2n\alpha^2\lambda^2\log(n)^4}{n\alpha\lambda\log(n)} + o(1) \xrightarrow[n\to+\infty]{} 0 .
\end{aligned} \tag{4.52}$$

Step [3]. Now we focus on the probability distribution on the set of gates, according to which the first gate is visited, when starting at quasi-stationarity. We define

$$\begin{aligned}
\mu_{\mathcal{G}_i}^{\mathrm{in}}(x) &= \frac{\sum_{y\in V_i\setminus\mathcal{G}_i}\mu_i^\star(y)P(y,x)}{\sum_{y\in V_i\setminus\mathcal{G}_i}\mu_i^\star(y)\sum_{z\in\mathcal{G}_i}P(y,z)}\\
&= (1+o_{\mathbb{P}}(1))\frac{\sum_{y\in V_i\setminus\mathcal{G}_i}\mu_i^\star(y)P(y,x)}{\alpha\lambda\log(n)} , \qquad x\in\mathcal{G}_i ,
\end{aligned}$$

where the second line is due to $\mu_i^\star P(\mathcal{G}_i) = \mathfrak{l}_i$ and the $o_{\mathbb{P}}(1)$ term is uniform over $x \in \mathcal{G}_i$. We claim that

$$\max_{x \notin \mathcal{G}_i} \frac{\mu_i^\star(x)}{\pi_i(x)} \leq 1 + o_{\mathbb{P}}(1)\,.$$

To prove it, recalling the definition of S in Corollary 4.6, we follow the same argument as in [Manzo et al., 2021, Eqs. 5.15–5.19]. We have that, for any $x \notin \mathcal{G}_i$,

$$\begin{aligned}
\mu_i^\star(x) &= (1-\mathfrak{l}_i)^{-S} \sum_{y \in V_i \setminus \mathcal{G}_i} \mu_i^\star(y)[P_i]^S_{\mathcal{G}_i}(y,x) \\
&\leq (1-\mathfrak{l}_i)^{-S} \sum_{y \in V_i \setminus \mathcal{G}_i} \mu_i^\star(y) P_i^S(y,x) \\
&= (1+o_{\mathbb{P}}(1))(1-\mathfrak{l}_i)^{-S} \sum_{y \in V_i \setminus \mathcal{G}_i} \mu_i^\star(y)\pi_i(x) \\
&= (1+o_{\mathbb{P}}(1))(1-\mathfrak{l}_i)^{-S}\pi_i(x) \\
&= (1+o_{\mathbb{P}}(1))\pi_i(x)\,,
\end{aligned}$$

where the first line follows by the definition of quasi-stationary distribution; the second line follows by the trivial fact $[P_i]^S_{\mathcal{G}_i}(y,x) \leq P_i^S(y,x)$ for any $x, y \notin \mathcal{G}_i$; the third line follows by Corollary 4.6 and the fact that, thanks to Theorem 4.4 and Proposition 4.5, $\pi_i(x) + o_{\mathbb{P}}(1) = (1+o_{\mathbb{P}}(1))\pi_i(x)$ uniformly in $x \in V_i$; in the fourth line we simply took the sum over $y \in V_i \setminus \mathcal{G}_i$; and in the last line we used that, thanks to Eq. (4.28), $\mathfrak{l}_i S = o_{\mathbb{P}}(1)$. Therefore,

$$\begin{aligned}
\max_{x \in \mathcal{G}_i} \mu^{\rm in}_{\mathcal{G}_i}(x) &\leq (1+o_{\mathbb{P}}(1)) \max_{x \in \mathcal{G}_i} \frac{\sum_{y \in V_i \setminus \mathcal{G}_i} \pi_i(y) P(y,x)}{\alpha\lambda\log(n)} \\
&\leq (1+o_{\mathbb{P}}(1)) \max_{x \in \mathcal{G}_i} \frac{\pi_i(x)}{\alpha\lambda\log(n)} = O_{\mathbb{P}}\left(\frac{1}{|\mathcal{G}_i|}\right)\,,
\end{aligned} \tag{4.53}$$

where the last bound follows from Theorem 4.4 and Lemma 4.17. In other words, the gates visited by the process are essentially uniformly distributed.

Step [4]. Fix $\varepsilon = \log(n)^{-1/3}$ and call

$$\mathcal{G}_{i,\rm nice} = \left\{x \in \mathcal{G}_i \setminus \mathcal{G}_{i,\rm bad} \mid D_x^+ \in [(1-\varepsilon)\lambda\log(n), (1+\varepsilon)\lambda\log(n)]\right\}\,.$$

Notice that, thanks to steps **[1]** and **[2]**, in particular Eqs. (4.52) and (4.51) respectively,

$$\mathbb{P}\left(\frac{|\mathcal{G}_{i,\rm nice}|}{|\mathcal{G}_i|} \geq 1 - 2\log(n)^{-2}\right) = 1 - o(1)\,. \tag{4.54}$$

Call $(B_k)_{k\geq 1}$ the the sequence of vertices in $\mathcal{G}_i$ that are visited by the process $(Y_t)_{t\geq 0}$. Consider the event

$$\mathcal{W} = \{B_k \in \mathcal{G}_{i,\rm nice},\ \forall k \leq \log(n)^{3/2}\}.$$

From step **[3]**, particularly (4.54) and (4.53), it follows

$$\hat{\mathbf{P}}^{G_i}_{\mu_i^\star}(\mathcal{W}) = 1 - o_{\mathbb{P}}(1)\,. \tag{4.55}$$

Step [5]. Now recall the notation $(\kappa_t, \rho_t)_{t\geq 0}$ given in Definition 4.24 and τ_ρ as in Eq. (4.49). Set $s = t/\alpha$. We want to control

$$\hat{\mathbf{P}}^{G_i}_{\mu_i^\star}(\tau_\rho > s) = \hat{\mathbf{P}}^{G_i}_{\mu_i^\star}(\rho_s = 0) = \hat{\mathbf{P}}^{G_i}_{\mu_i^\star}(\rho_s = 0, \mathcal{W}) + \hat{\mathbf{P}}^{G_i}_{\mu_i^\star}(\rho_s = 0, \mathcal{W}^\complement)\,. \tag{4.56}$$

The second term on the r.h.s. of (4.56) is $o_{\mathbb{P}}(1)$ thanks to (4.55). Let us bound the other one. Recalling that $(\kappa_s - \kappa_{s-1})_{s\geq 1}$ are i.i.d. Bernoulli random variables with parameter $\mathfrak{l}_i$, it immediately follows that $\kappa_s \sim \mathrm{Bin}(s, \mathfrak{l}_i)$, so that, as $\mathfrak{l}_i = \lambda\alpha\log(n)(1 + o_{\mathbb{P}}(1)) = o_{\mathbb{P}}(1)$, it holds

$$\hat{\mathbf{Var}}^{G_i}_{\mu_i^\star}(\kappa_s) = s(\mathfrak{l}_i - \mathfrak{l}_i^2) = s\lambda\alpha\log(n)(1 + o_{\mathbb{P}}(1)) = \hat{\mathbf{E}}^{G_i}_{\mu_i^\star}[\kappa_s](1 + o_{\mathbb{P}}(1)).$$

Taking $\delta = (\log\log\log(n))^{-1}$ and $s \in \left[\frac{1}{\alpha C_n}, \frac{C_n}{\alpha}\right]$, for some sequence C_n diverging sufficiently slowly, by Chebychev inequality,

$$\begin{aligned}
\hat{\mathbf{P}}^{G_i}_{\mu_i^\star}\left(\left|\frac{\kappa_s}{s\alpha\lambda\log(n)} - 1\right| \geq \delta\right) &\leq \frac{\hat{\mathbf{E}}^{G_i}_{\mu_i^\star}[\kappa_s]}{\delta^2(\hat{\mathbf{E}}^{G_i}_{\mu_i^\star}[\kappa_s])^2}(1 + o_{\mathbb{P}}(1)) \\
&\leq \frac{(\log\log\log(n))^2 C_n}{\lambda\log(n)}(1 + o_{\mathbb{P}}(1)) = o_{\mathbb{P}}(1)\,.
\end{aligned}$$

In conclusion, the first term on the r.h.s. of (4.56) can be bounded, for s as above, by

$$\begin{aligned}
\hat{\mathbf{P}}^{G_i}_{\mu_i^\star}(\rho_s = 0) &= (1 - o_{\mathbb{P}}(1))\hat{\mathbf{P}}^{G_i}_{\mu_i^\star}\left(\rho_s = 0 \,\middle|\, \mathcal{W} \cap \left\{\left|\frac{\kappa_s}{s\alpha\lambda\log(n)} - 1\right| < \delta\right\}\right) + o_{\mathbb{P}}(1) \\
&= (1 - o_{\mathbb{P}}(1))\left(1 - \frac{1}{(1 + O(\varepsilon))\lambda\log(n)}\right)^{(1+O(\delta))s\alpha\lambda\log(n)} + o_{\mathbb{P}}(1) \\
&= \mathrm{e}^{-s\alpha} + o_{\mathbb{P}}(1)\,,
\end{aligned}$$

where we used that ε and δ are vanishing and, in the second line, we used the conditioning and that, for $s\alpha\log(n) \ll \log(n)^{3/2}$, on the event $\mathcal{W}$, the gates visited by the random walk have degrees in $[(1 - \varepsilon)\lambda\log(n), (1 + \varepsilon)\lambda\log(n)]$. Then, provided that C_n diverges sufficiently slowly, we get

$$\max_{s\in\left[\frac{1}{\alpha C_n}, \frac{C_n}{\alpha}\right]} \left|\hat{\mathbf{P}}^{G_i}_{\mu_i^\star}(\tau_\rho > s) - \mathrm{e}^{-s\alpha}\right| = o_{\mathbb{P}}(1)\,.$$

The proof ends recalling that $s\alpha = t$. □

We are now in a good position to present the proof of Proposition 4.25.

Proof of Proposition 4.25. First of all we observe the following fact: if $\rho_t > 0$ for some $t \geq 0$ then, at time t, we have sufficient information to declare if the coupling is *successful* or *failed*. Therefore, thanks to Proposition 4.27, with probability $1 - o_{\mathbb{P}}(1)$ the coupling consists of less than $\log(n)^{3/2}$ stages. Moreover, thanks to Theorem 4.2 and the subadditivity of the total variation distance, the probability of a meeting before time $\log(n)^2$ is $1 - o_{\mathbb{P}}(n^{-1})$ by Corollary 4.6. This means that the probability that along the coupling there is a stage in which the two processes meet after time $\log(n)^2$ is $o_{\mathbb{P}}(1)$.

Now, each stage the coupling might fail because of two alternative reasons:

1. On the one hand, a stage might produce a failure if (ii) in Definition 4.24 happens: the process Y hits $\mathcal{G}_i$ before meeting the first process under optimal coupling. By Corollary 4.26 the probability that Y hits $\mathcal{G}_i$ before time $\log(n)^2$ is $o_{\mathbb{P}}(n^{-\frac{1}{7}})$. Hence, the probability that the coupling fails due to this kind of event is $o_{\mathbb{P}}(\log(n)^{3/2} n^{-\frac{1}{7}})$.
2. On the other hand, a stage might produce a failure if (iii) in Definition 4.24 happens: the process X might visit $\mathcal{G}_i$ and traverse a rewired edge before time $\log(n)^2$. For what concerns the first stage, in which the starting point of the walk is arbitrary, thanks to Proposition 4.12, the probability of this event is $o_{\mathbb{P}}(1)$. For any successive stage, first notice that after time $\log(n)^2$, by Corollary 4.6, the distribution of the random walk X can be w.h.p. approximated in ℓ^∞-norm by π_i. Thanks to Lemma 4.22, complemented with Remark 4.23, we have that the expected number of visits to $\mathcal{G}_i$ within time $\log(n)^2$ is $o_{\mathbb{P}}(n^{-1/3})$. Hence, the probability that there exists a stage in which $\mathcal{G}_i$ is visited before $\log(n)^2$ is $o_{\mathbb{P}}(\log(n)^{3/2} n^{-\frac{1}{3}})$.

□

4.4.4 Proof of Proposition 4.15 and Theorem 4.16

We are almost ready to conclude the proof of Proposition 4.15, and then to show how Theorem 4.16 can easily be deduced from it using the same set of arguments used to deduce Theorem 4.14 from Proposition 4.10.

We start by stressing that, thanks to Corollary 4.28, we know that the jumping time of the process starting at any $x \in V$, properly scaled, is well approximated by a standard exponential random variable. We now want to show that the jumps occur uniformly at random among communities.

Lemma 4.29. *For $\sqrt{n}\log(n)^{-2} \ll \alpha^{-1} \ll \lambda n \log(n)$*

$$\max_{i\leq m}\max_{j\neq i}\max_{x\in V_i}\left|\mathbf{P}_x^G(X_{\tau_{\text{jump}}} \in V_j) - \frac{1}{m-1}\right| = o_{\mathbb{P}}(1)\,.$$

Proof. Notice that vertices in $\mathcal{G}_{i,\text{nice}}$ have, by definition, a unique rewired edge. To each $y \in \mathcal{G}_{i,\text{nice}}$, one can assign a mark $J(y)$, chosen u.a.r. in $[m] \setminus \{i\}$, representing the community to which the unique rewired edge of y connects after rewiring. We will then partition

$$\mathcal{G}_{i,\text{nice}} = \sqcup_{j\neq i}\mathcal{G}^j_{i,\text{nice}}\,,$$

where $\mathcal{G}^j_{i,\mathrm{nice}}$ is the subset of $\mathcal{G}_{i,\mathrm{nice}}$ having mark j. Since $|\mathcal{G}_{i,\mathrm{nice}}| = \omega_{\mathbb{P}}(1)$, and by the uniform choice,

$$\max_{i\le m}\max_{j\ne i}\left|\frac{|\mathcal{G}^j_{i,\mathrm{nice}}|/|\mathcal{G}_{i,\mathrm{nice}}|}{(m-1)^{-1}}-1\right| = o_{\mathbb{P}}(1)\,. \tag{4.57}$$

Moreover, specializing (4.53) to the case $x\in\mathcal{G}_{i,\mathrm{nice}}$, where D^+_x can be well estimated, one has

$$\max_{x\in\mathcal{G}_{i,\mathrm{nice}}}\mu^{\mathrm{in}}_{\mathcal{G}_i}(x)\le(1+o_{\mathbb{P}}(1))\frac{1}{|\mathcal{G}_i|}=(1+o_{\mathbb{P}}(1))\frac{1}{|\mathcal{G}_{i,\mathrm{nice}}|}\,, \tag{4.58}$$

where in the last step we used (4.54). As a consequence, the hitting measure of $\mathcal{G}_i$ is asymptotically uniform on $\mathcal{G}_{i,\mathrm{nice}}$. In conclusion, the desired result follows by putting together Proposition 4.25, (4.57) and (4.58). □

Thanks to Lemma 4.29 we can couple the evolution of the non-Markovian process $(c(X_t))_{t\ge0}$ with the evolution of a simple random walk on a complete graph with m vertices, with transitions as in (4.7). Using this fact, we can finally prove Proposition 4.15 and Theorem 4.16.

Proof Proposition 4.15. We consider the following iterated version of the coupling in Definition 4.24:

- We start the coupling at some $x\in V_i$ for some $i\le m$.
- If the coupling succeeds and at time τ_{jump} the process is found at some $y\in V_j$ with $j\ne i$, then the coupling is restarted with the initial distribution $\delta_y\otimes\mu^\star_j$.
- Iterate this procedure up to the first iteration at which the corresponding coupling fails.

Call $\widehat{\mathbf{P}}^G_x$ the measure associated to this coupling, and fix an arbitrary sequence of integers $C\equiv C_n\gg1$. Call $\mathcal{E}^{\mathrm{succ}}_C$ the event in which the coupling succeeds up to the C^2-th iteration. Then, thanks to Proposition 4.25

$$\min_{x\in V}\widehat{\mathbf{P}}^G_x(\mathcal{E}^{\mathrm{succ}}_C)=(1-o_{\mathbb{P}}(1))^{C^2}\,. \tag{4.59}$$

Call now $\mathcal{E}^{\mathrm{iter}}_C$ the event in which by time $t=C\alpha^{-1}$ there are at most C^2 iterations. Then,

$$\min_{x\in V}\widehat{\mathbf{P}}^G_x(\mathcal{E}^{\mathrm{iter}}_C\cap\mathcal{E}^{\mathrm{succ}}_C)=\min_{x\in V}\widehat{\mathbf{P}}^G_x(\mathcal{E}^{\mathrm{succ}}_C)\widehat{\mathbf{P}}^G_x(\mathcal{E}^{\mathrm{iter}}_C\mid\mathcal{E}^{\mathrm{succ}}_C)=(1-o_{\mathbb{P}}(1))^{2C^2}\,, \tag{4.60}$$

where in the second equality we used (4.59) and Corollary 4.28 on a union bound on the C^2 iterations of the coupling, which ensures that—conditionally on the event $\mathcal{E}^{\mathrm{succ}}_C$—the length of each iteration is asymptotically geometrically distributed with parameter α. As a consequence, the $o_{\mathbb{P}}(1)$ in (4.60), is the maximal one among the errors in (4.59) and Corollary 4.28. In conclusion, as soon as $C\gg1$ is chosen to diverge sufficiently slowly, the iterative version of the coupling will be successful w.h.p. up to time $t=C\alpha^{-1}$. In particular, up to that time, the number of iterations will be w.h.p. at most C^2 and the length of such iterations will be w.h.p. coupled to an independent geometric random variable of rate α. Moreover, thanks to Lemma 4.29, the inter-community jump ending each iteration is approximately uniformly distributed among the other communities. Therefore, if $(\tau^{(\ell)}_{\rm jump})_{\ell\le C^2}$ denotes the sequence of the lengths of the first C^2 iterations, we can couple the sequence $\big(c(X_{\tau^{(\ell)}_{\rm jump}})\big)_{\ell\le C^2}$ with the trajectory (of length C^2) of the process with transition matrix Q at a total-variation cost bounded by $1-C^2o_{\mathbb{P}}(1)$. The latter goes to 1 as soon as C diverges sufficiently slowly. At this point the desired result follows at once by noting that if, for any $t\ge 0$, we sample i.i.d. geometric random variables of parameter α up to the first time in which their cumulative sum is above t (hence, sampling the jump times), and then we sample the path of a simple random walk on the complete graph, we can entirely reconstruct the trajectory of process with transition matrix Q. □

Proof of Theorem 4.16. The proof of this fact follows the same line of argument as in Theorem 4.14. In our setting, Eq. (4.18) follows from Proposition 4.15 for any $T\ll C\alpha^{-1}$, where $C\gg 1$ is as given in Proposition 4.15. The remainder of the argument proceeds through the same steps, leading to the validity of Eq. (4.20) for any $\alpha^{-1}\ll T\ll\sqrt{C}\alpha^{-1}$. Finally, by the monotonicity of the distance to equilibrium, the result holds for any $T\gg\alpha^{-1}$. □

4.5 Critical and subcritical regime

In this section, we analyze the mixing behavior of the random walk in the regime $\alpha^{-1}\lesssim \boldsymbol{t}_{\rm ent}$. This range of α includes both the subcritical and critical cases, described in Eqs. (4.3) and (4.4). The proofs of these results are adaptations of the techniques developed in Bordenave et al. [2018] and Chapter 3 to the current setting with multiple communities. Since it will be useful later, in what follows we take $\alpha^{-1}=C\boldsymbol{t}_{\rm ent}$, for $C>0$. We again stress that Eq. (4.3) will hold only for $C\to 0$.

4.5.1 Concentration results for random walk paths

In this section we recall some of the results given in the previous chapter, and restate them in a different fashion. The first random object we need is the following. For each oriented path $\mathfrak{p}=(x_0,\dots,x_t)$ in G, let the quenched probability *mass* of $\mathfrak{p}$ be

$$\mathbf{m}(\mathfrak{p}):=\prod_{s=0}^{t-1}P(x_s,x_{s+1})\,.$$

We state the following result, which can be written in the shape of a Law of Large Numbers for the variables $\log(P(X_s,X_{s+1}))$.

Theorem 4.30 (Quenched LLN)**.** *Let $t = \Theta(\boldsymbol{t}_{\mathrm{ent}})$ and $\theta \in (0,1)$ be such that there exists $\rho > 0$, $\rho \neq 1$, satisfying $\log\theta = \rho \mathrm{H} t(1+o(1))$. Then*

$$\max_{x\in V} \left| \mathbf{P}_x^G\left(\mathbf{m}(X_0, X_1, \dots, X_t) > \theta\right) - \mathbf{1}_{\{\rho>1\}} \right| \xrightarrow[n\to+\infty]{\mathbb{P}} 0.$$

Proof. The proof follows by a straightforward adaptation of the proof of Theorem 3.21 from Chapter 3 which, in turn, is an adaptation of the proof of [Bordenave et al., 2019, Theorem 4]. Indeed, all vertex out-degrees have the same law, $\mathrm{Bin}(n,p)$, and the proof in the above mentioned references does not rely on the details of the graph structure but only on the out-degree distribution and the fact that the random walk is w.h.p. non-backtracking on the timescale $\log(n)$. □

This result has the same shape of Eq. (4.3) and is the core of the cutoff result. In fact, it is easy to observe that $\theta = \mathrm{e}^{-\mathrm{H}\boldsymbol{t}_{\mathrm{ent}}} = n^{-1}$ provides a threshold for the concentration of the mass of the paths with length $\boldsymbol{t}_{\mathrm{ent}}$. That specific choice of θ is not covered by the theorem above, but in Chapter 3 it is possible to find a quenched CLT refinement, valid in this critical regime, to describe the cutoff window under a suitable hypothesis.

The second ingredient is provided by the following family of paths, which allows us to capture typical features of SRW paths.

Definition 4.31 (Nice paths)**.** *Let $\varepsilon \in (0,1)$, and*

$$\eta := 2\varepsilon \boldsymbol{t}_{\mathrm{ent}}, \qquad \sigma := (1-\varepsilon)\boldsymbol{t}_{\mathrm{ent}}, \qquad t := \sigma + \eta + 1 = (1+\varepsilon)\boldsymbol{t}_{\mathrm{ent}} + 1.$$

We say that a path $\mathfrak{p} = (x, x_1, x_2, \dots, x_{t-1}, y)$ of length t from x to y is nice *if:*

(i) the entire path is such that $\mathbf{m}(\mathfrak{p}) \le \frac{1}{n\log^3 n}$;

(ii) the sub-path $(x_{\sigma+1}, \dots, x_{t-1}, y)$ is the unique path in G of length at most η from x_σ to y;

(iii) the sub-path $(x, x_1, \dots, x_\sigma)$ is contained in the random tree $\mathcal{T}_x(\sigma)$ constructed as follows:

Fix a realization of G and a root node $x \in V$. Let $\mathcal{G}^0 = \mathcal{T}^0 := \{x\}$. Then, for $\ell \ge 1$:

(1) Let $\mathcal{E}^\ell := \{(v,y) : v \in \mathcal{G}^{\ell-1}, y \in \mathcal{B}_v^+(1) \setminus \mathcal{G}^{\ell-1}\}$, be the set of edges which have not been visited by the first $\ell-1$ iterations, and with tails in $\mathcal{G}^{\ell-1}$.

(2) Choose $e = (v,y) \in \mathcal{E}^\ell$ such that, if $\mathfrak{p}_{x,v}$ is the unique path from x to v in $\mathcal{G}^{\ell-1}$, $\mathbf{m}(\mathfrak{p}_{x,v})(D_v^+)^{-1}$ is not below $\mathrm{e}^{-(1+\varepsilon)\mathrm{H}\sigma}$ and it is maximal among $(v,y) \in \mathcal{E}^\ell$, and v is at distance at most $\sigma - 1$ from x (use a deterministic criterion to break ties).
If such edge does not exist, stop the procedure and set $\mathcal{T}_x(\sigma) := \mathcal{T}^{\ell-1}$;

(3) Generate $\mathcal{G}^\ell$ by adding e to $\mathcal{G}^{\ell-1}$;

(4) Generate $\mathcal{T}^\ell$ by adding e to $\mathcal{T}^{\ell-1}$ if it results in a tree, otherwise $\mathcal{T}^\ell = \mathcal{T}^{\ell-1}$.

Finally, for every fixed $x, y \in V$, we define $P^t_{\rm nice}(x,y)$ to be the probability that the simple random walk started at x arrives at y at time t after having followed a nice path.

As mentioned above, nice paths w.h.p. host typical trajectories of the SRW:

Proposition 4.32. *Fix $\beta > 1$ and $t = \beta \boldsymbol{t}_{\rm ent}$. Then, uniformly in a starting position in $V^\star_\varepsilon$, defined in Eq. (4.8), the SRW w.h.p. follows nice paths. More precisely,*

$$\max_{x \in V^\star_\varepsilon} \Big(1 - \sum_{y \in V} P^t_{\rm nice}(x,y)\Big) = o_{\mathbb{P}}(1).$$

Proof. Thanks to Lemma 4.7, the requirements (ii) and (iii) of Definition 4.31 are w.h.p. satisfied. Then, letting $\theta = (n \log^3 n)^{-1}$,

$$\max_{x \in V^\star_\varepsilon} \Big(1 - \sum_{y \in V} P^t_{\rm nice}(x,y)\Big) \le \max_{x \in V^\star_\varepsilon} \mathbf{P}^G_x \left(\mathbf{m}(X_0, X_1, \dots, X_t) > \theta\right) + o_{\mathbb{P}}(1),$$

and the l.h.s. (namely, the cost of requirement (i)) vanishes by Theorem 4.30. □

4.5.2 Bounds on the total variation profile

With the latter tools at hand and with the help of Proposition 4.33 below, which is an adaptation of Proposition 3.26 from Chapter 3, it is possible to bound effectively the distance to equilibrium.

Proposition 4.33. *Let $\alpha^{-1} \lesssim \boldsymbol{t}_{\rm ent}$, $x \in V$, and $t = \sigma + \eta + 1$ as in Definition 4.31. Then,*

$$\mathbb{P}\left(P^t_{\rm nice}(x,y) \le (1+\delta)\nu_{c(x)}(y) + \frac{\delta}{n},\ \forall x, \in V^\star_\varepsilon, \forall y \in V, \right) = 1 - o(1), \qquad \forall \delta > 0,$$

where $Q^{\sigma+1}(\cdot,\cdot)$ is defined in Eq. (4.9) and, for $i \le m$,

$$\nu_i(y) := \frac{1}{n} \sum_{w \in V} Q^{\sigma+1}(i, c(w)) P^\eta(w,y). \tag{4.61}$$

Proof. Let $x \in V^\star_\varepsilon$, as defined in Eq. (4.8), and let $\mathcal{F}$ denote the partial environment generated by the tree $\mathcal{T}_x(\sigma)$ and $\mathcal{B}^-_y(\eta)$, the in-neighborhood of y of depth η. We split $P^t_{\rm nice}(x,y)$ in m different addends. For $i = 1, \dots, m$, let

$$P^t_{{\rm nice},i}(x,y) = \sum_{z \in V^+_{\mathcal{F}}(i)} \sum_{v \in V^-_{\mathcal{F}}} \mathbf{m}(\mathfrak{p}_{x,z}) \frac{1}{D^+_z} \mathbf{m}(\mathfrak{p}_{v,y}) \mathbf{1}_{\{z \to v\}} \mathbf{1}_{\{\mathfrak{p} \text{ is a nice path}\}},$$

where $V^+_{\mathcal{F}}(i)$ is the set of vertices at depth σ in $\mathcal{T}_x(\sigma) \cap V_i$, and $V^-_{\mathcal{F}}$ is the set of vertices in $\partial\mathcal{B}^-_y(\eta)$ such that there exists a unique path of length η to y. Then, it holds $P^t_{\rm nice}(x,y) = \sum_{i \le m} P^t_{{\rm nice},i}(x,y)$. Setting $\mathbb{P}_{\mathcal{F}}(\cdot) := \mathbb{P}(\cdot \,|\mathcal{F})$, we have that, for $z, v \in V$,

$$\mathbb{E}\left[\frac{\mathbf{1}_{\{z \to v\}}}{D^+_z} \,|\mathcal{F}\right] = \mathbb{E}\left[\frac{\mathbf{1}_{\{z \to v\}}}{D^+_z} \,|\mathcal{F}, \mathbf{1}_{\{z \to v\}} = 1\right] \mathbb{P}\left((z,v) \in E\right) = \frac{p_{zv}}{np}(1 + o(1)).$$

Then,

$$\begin{aligned}\mathbb{E}[P^t_{\mathrm{nice},i}(x,y)\,|\,\mathcal{F}\,] &\le \sum_{v\in V^-_{\mathcal{F}}}\sum_{z\in V^+_{\mathcal{F}}(i)} \mathbf{m}(\mathfrak{p}_{x,z})\mathbb{E}\left[\frac{\mathbf{1}_{\{z\to v\}}}{D^+_z}\,|\mathcal{F}\right]\mathbf{m}(\mathfrak{p}_{v,y})\\ &\le \sum_{v\in V^-_{\mathcal{F}}}\sum_{z\in V^+_{\mathcal{F}}(i)} \mathbf{m}(\mathfrak{p}_{x,z})\frac{p_{zv}}{np}\mathbf{m}(\mathfrak{p}_{v,y})(1+o(1)).\end{aligned}$$

Since the random walk performs w.h.p. a nice path with support concentrated on the vertices of the tree $\mathcal{T}_x(\sigma)$ (Proposition 4.32), and thanks to Proposition 4.10, for $i\le m$, w.h.p. it holds

$$\left|\sum_{z\in V^+_{\mathcal{F}}(i)}\mathbf{m}(\mathfrak{p}_{x,z}) - Q^\sigma(c(x),i)\right| = o(1).$$

Then, observing that $\frac{p_{zv}}{p} = Q(c(z),c(v))$, it holds

$$\begin{aligned}\mathbb{E}[P^t_{\mathrm{nice},i}(x,y)\,|\,\mathcal{F}\,] &\le Q^\sigma(c(x),i)\sum_{v\in V^-_{\mathcal{F}}}\frac{Q(i,c(v))}{n}\mathbf{m}(\mathfrak{p}_{v,y})(1+o(1))\\ &\le Q^\sigma(c(x),i)\sum_{w\in V}\frac{Q(i,c(w))}{n}P^\eta(w,y)(1+o(1)).\end{aligned}$$

Summing over $i\le m$, we conclude

$$\begin{aligned}\mathbb{E}[P^t_{\mathrm{nice}}(x,y)\,|\,\mathcal{F}\,] &\le \sum_{w\in V}\frac{Q^{\sigma+1}(c(x),c(w))}{n}P^\eta(w,y)(1+o(1)) \\ &= \nu_{c(x)}(y)(1+o(1)).\end{aligned} \tag{4.62}$$

The proof then continues as in Proposition 3.26, employing Eq. (4.62), a suitable Bernstein's concentration inequality, and averaging over the partial environment $\mathcal{F}$. □

We now proceed with our analysis, pursuing an upper bound and a lower bound for the distance to stationarity at times t greater than $\beta\boldsymbol{t}_{\mathrm{ent}}$ for $\beta>1$, and smaller than $\beta\boldsymbol{t}_{\mathrm{ent}}$ for $\beta<1$, respectively.

Upper bound. We start stating the following straightforward lemma.

Lemma 4.34. *Let $\alpha^{-1}=C\boldsymbol{t}_{\mathrm{ent}}$. Then for each $j=1,\dots,m$, it holds $\pi(V_j)=\frac{1}{m}(1+o_{\mathbb{P}}(1))$.*

Proof. By definition of stationarity, for each $j=1,\dots,m$, and for $t=(\log(n))^2$,

$$\left|\pi(V_j)-\frac{1}{m}\right| \le \sum_{x\in V}\pi(x)\left|\mathbf{P}^G_x(X_t\in V_j)-\frac{1}{m}\right| \le \max_{x\in V}\left|\mathbf{P}^G_x(X_t\in V_j)-\frac{1}{m}\right|,$$

which is $o_{\mathbb{P}}(1)$ by Proposition 4.10. □

Choose now a time $t \geq \beta \boldsymbol{t}_{\text{ent}}$, for $\beta > 1$. Applying Proposition 4.33, and later Proposition 4.32, for every $\delta > 0$ it holds

$$\begin{aligned}\max_{x\in V_\varepsilon^\star}\|P^t(x,\cdot)-\nu_{c(x)}\|_{\text{TV}} &\leq \max_{x\in V_\varepsilon^\star}\sum_{y\in V}\left[(1+\delta)\nu_{c(x)}(y)+\frac{\delta}{n}-P^t_{\text{nice}}(x,y)\right]\\ &\leq \max_{x\in V_\varepsilon^\star}\left(1-\sum_{y\in V}P^t_{\text{nice}}(x,y)\right)+2\delta = 2\delta + o_{\mathbb{P}}(1).\end{aligned} \tag{4.63}$$

Then, thanks to Lemma 4.7 and Eq. (4.63), for $\ell = 3\log\log(n)$,

$$\max_{x\in V}\|P^{t+\ell}(x,\cdot)-\nu_{c(x)}\|_{\text{TV}} \leq \mathbf{P}^G_x(X_\ell\notin V_\varepsilon^\star)+\max_{x\in V_\varepsilon^\star}\|P^t(x,\cdot)-\nu_{c(x)}\|_{\text{TV}} = o_{\mathbb{P}}(1). \tag{4.64}$$

Since, for every $w\in V$, it holds $\sum_{i=1}^m Q^{\sigma+1}(i,c(w))=1$, we can define

$$\nu := \frac{1}{m}\sum_{i=1}^m \nu_i = \frac{1}{mn}\sum_{w\in V}P^\eta(w,\cdot). \tag{4.65}$$

Employing Lemma 4.34 and later Eq. (4.64), it holds

$$\begin{aligned}\|\nu-\pi\|_{\text{TV}} &= \left\|\frac{1}{m}\sum_{i=1}^m\nu_i-\sum_{z\in V}\pi(z)P^{(1+\varepsilon)\boldsymbol{t}_{\text{ent}}}(z,\cdot)\right\|_{\text{TV}}\\ &= \left\|\sum_{i=1}^m\sum_{z\in V_i}\pi(z)\left[\nu_i-P^{(1+\varepsilon)\boldsymbol{t}_{\text{ent}}}(z,\cdot)\right]\right\|_{\text{TV}}+o_{\mathbb{P}}(1)\\ &\leq \max_{i\leq m}\max_{z\in V_i}\left\|P^{(1+\varepsilon)\boldsymbol{t}_{\text{ent}}}(z,\cdot)-\nu_{c(z)}\right\|_{\text{TV}} = o_{\mathbb{P}}(1).\end{aligned} \tag{4.66}$$

This means that ν constitutes a good proxy for π. Then, by the triangle inequality, and thanks to Eqs. (4.64) and (4.66), we can conclude the following upper bound

$$\begin{aligned}\max_{x\in V}\left\|P^{t+\ell}(x,\cdot)-\pi\right\|_{\text{TV}} &\leq \max_{x\in V}\left\|P^{t+\ell}(x,\cdot)-\nu_{c(x)}\right\|_{\text{TV}}+\max_{i\leq m}\|\nu_i-\nu\|_{\text{TV}}+\|\nu-\pi\|_{\text{TV}}\\ &\leq \max_{i\leq m}\|\nu_i-\nu\|_{\text{TV}}+o_{\mathbb{P}}(1).\end{aligned} \tag{4.67}$$

Lower bound. Let $t \leq \beta\boldsymbol{t}_{\text{ent}}$, for $\beta < 1$. We are going to show that the law of X_t at a time $t \leq \beta\boldsymbol{t}_{\text{ent}}$, is w.h.p. concentrated on a set with cardinality $o(n)$. To this end, let $\theta = n^{-\beta(2-\beta)}$ and let

$$S_x := \{y\in V:\ \text{there exists a path } \mathfrak{p} \text{ of length } t \text{ such that } \mathbf{m}(\mathfrak{p})\geq\theta\}.$$

We have $|S_x| \le \theta^{-1} = n^{\beta(2-\beta)} = o(\frac{n}{\log(n)})$, since, for $\beta < 1$, we have $\beta(2-\beta) < 1$. Moreover, it holds $-\frac{\log\theta}{\mathrm{H}t} = 2 - \beta > 1$, and we conclude the following lower bound:

$$\begin{aligned} \min_{x\in V}\left\|\mathbf{P}_x^G(X_t\in\cdot)-\pi\right\|_{\mathrm{TV}} &\ge \min_{x\in V}\left[\mathbf{P}_x^G(X_t\in S_x)-\pi(S_x)\right] \\ &\ge \min_{x\in V}\mathbf{P}_x^G\left(\mathbf{m}(X_0,\dots,X_t)\ge n^{-\beta(2-\beta)}\right)-\max_{x\in V}\pi(S_x) \\ &= 1-o_{\mathbb{P}}(1), \end{aligned} \tag{4.68}$$

where the last inequality holds combining Theorem 4.30 with the observation that

$$\max_{x\in V}\pi(S_x) = o_{\mathbb{P}}(1)\,,$$

thanks to Theorem 4.4 and Proposition 4.5.

4.6 Proof of the main results

4.6.1 Supercritical regime

In this regime, the proof of Theorem 4.3 is split into two parts, depending on the chosen timescale. We first analyze the case $t \asymp \boldsymbol{t}_{\mathrm{ent}}$, proving Eq. (4.5), and then move to the case $t \asymp \alpha^{-1}$, proving Eq. (4.6).

Relaxation to a local equilibrium. Let us first assume $\boldsymbol{t}_{\mathrm{ent}} \ll \alpha^{-1} \ll \sqrt{n}\log(n)^{-2}$. We will sometimes commit a slight abuse of notation by lifting π_i to a probability measure on the entire vertex set V. Let $t = \beta\boldsymbol{t}_{\mathrm{ent}}$ for some $\beta < 1$. By Theorem 4.2,

$$\min_{i\le m}\min_{x\in V_i}\|\mathbf{P}_x^{G_i}(X_t^i\in\cdot)-\pi_i\|_{\mathrm{TV}} = 1-o_{\mathbb{P}}(1)\,, \tag{4.69}$$

In particular, for every $\delta > 0$ and $x \in V_i$, it must hold

$$\frac12\sum_{y\in V}\left|\mathbf{P}_x^{G_i}(X_t^i=y)-\frac1m\pi_i(y)\right| \ge 1-\frac{m-1}{2m}-\delta+o_{\mathbb{P}}(1),$$

otherwise it would be

$$\begin{aligned} \left\|\mathbf{P}_x^{G_i}(X_t^i\in\cdot)-\pi_i\right\|_{\mathrm{TV}} &\le \frac12\sum_{y\in V}\left|\mathbf{P}_x^{G_i}(X_t^i=y)-\frac1m\pi_i(y)\right|+\frac12\cdot\frac{m-1}{m}\sum_{y\in V}\pi_i(y) \\ &< 1-\frac{m-1}{2m}-\delta+\frac{m-1}{2m}+o_{\mathbb{P}}(1) = 1-\delta+o_{\mathbb{P}}(1), \end{aligned}$$

which is in contradiction with (4.69). As a consequence, using the characterization of π given in Theorem 4.14, and using that, by Proposition 4.12,

$$\begin{aligned} \left\|\mathbf{P}_x^G(X_t\in\cdot)-\mathbf{P}_x^{G_i}(X_t^i\in\cdot)\right\|_{\mathrm{TV}} &\le \max_{i\le m}\max_{x\in V_i}\check{\mathbf{P}}_x(X_t\ne X_t^i) \\ &\le \max_{i\le m}\max_{x\in V_i}\mathbf{P}_x^G(\tau_{\mathrm{jump}}\le t) = o_{\mathbb{P}}(1), \end{aligned}$$

we have that, for every $\delta > 0$ and $x \in V_i$,

$$\begin{aligned}\left\|\mathbf{P}_x^G(X_t \in \cdot) - \pi\right\|_{\mathrm{TV}} &= \left\|\mathbf{P}_x^{G_i}(X_t^i \in \cdot) - \frac{1}{m}\sum_{j=1}^m \pi_j\right\|_{\mathrm{TV}} + o_{\mathbb{P}}(1) \\ &= \frac{1}{2}\sum_{y\in V}\left|\mathbf{P}_x^{G_i}(X_t^i = y) - \frac{1}{m}\pi_i(y)\right| + \sum_{j\neq i}\frac{1}{2m}\sum_{y\in V}\pi_j(y) + o_{\mathbb{P}}(1) \\ &\geq 1 - \frac{m-1}{2m} - \delta + \frac{m-1}{2m} + o_{\mathbb{P}}(1) = 1 - \delta + o_{\mathbb{P}}(1).\end{aligned} \tag{4.70}$$

Let now $\beta > 1$. By definition of total variation distance

$$\begin{aligned}\min_{i\leq m}\min_{x\in V_i}\|\mathbf{P}_x^G(X_t \in \cdot) - \pi\|_{\mathrm{TV}} &\geq \min_{i\leq m}\min_{x\in V_i}\left|\mathbf{P}_x^G(X_t \in V_i) - \pi(V_i)\right| \\ &= \frac{m-1}{m}\left(1 - \frac{m}{m-1}\alpha\right)^{t-1} + o_{\mathbb{P}}(1)\,,\end{aligned} \tag{4.71}$$

where we have used Proposition 4.10 and the characterization of π given in Theorem 4.14. By our choice $t = \beta\boldsymbol{t}_{\mathrm{ent}} \ll \alpha^{-1}$, one gets

$$\min_{i\leq m}\min_{x\in V_i}\|\mathbf{P}_x^G(X_t \in \cdot) - \pi\|_{\mathrm{TV}} \geq \frac{m-1}{m} - o_{\mathbb{P}}(1)\,. \tag{4.72}$$

For what concerns the upper bound, for $\beta > 1$, let us fix some $\gamma \geq 0$, possibly depending on n, and $\varepsilon > 0$ such that $(1+\varepsilon)\boldsymbol{t}_{\mathrm{ent}} + \gamma \leq \beta\boldsymbol{t}_{\mathrm{ent}}$. By splitting over the vertex on which the SRW sits at time $(1+\varepsilon)\boldsymbol{t}_{\mathrm{ent}}$, and over the community of such vertex one gets, for $x \in V$,

$$P^{(1+\varepsilon)\boldsymbol{t}_{\mathrm{ent}}+\gamma}(x,\cdot) = \sum_{y\in V} P^\gamma(x,y)P^{(1+\varepsilon)\boldsymbol{t}_{\mathrm{ent}}}(y,\cdot) = \sum_{i=1}^m\sum_{y\in V_i} P^\gamma(x,y)P^{(1+\varepsilon)\boldsymbol{t}_{\mathrm{ent}}}(y,\cdot)\,.$$

Using Theorem 4.13, we can write

$$\max_{x\in V}\left\|P^{(1+\varepsilon)\boldsymbol{t}_{\mathrm{ent}}+\gamma}(x,\cdot) - \sum_{i=1}^m P^\gamma(x,V_i)\pi_i\right\|_{\mathrm{TV}} = o_{\mathbb{P}}(1).$$

Let us now focus on the case $\boldsymbol{t}_{\mathrm{ent}} \ll \alpha^{-1} \ll \sqrt{n}\log(n)^{-2}$. By Proposition 4.10 we get

$$\max_{x\in V}\left\|P^{(1+\varepsilon)\boldsymbol{t}_{\mathrm{ent}}+\gamma}(x,\cdot) - \sum_{i=1}^m Q^\gamma(c(x),i)\pi_i\right\|_{\mathrm{TV}} = o_{\mathbb{P}}(1).$$

Thanks to Theorem 4.14 and recalling the definition in (4.9), we then obtain

$$\begin{aligned}\max_{x\in V}\left\|P^{(1+\varepsilon)\boldsymbol{t}_{\mathrm{ent}}+\gamma}(x,\cdot) - \pi\right\|_{\mathrm{TV}} &= \frac{1}{2}\sum_{i=1}^m\left|Q^\gamma(c(x),i) - \frac{1}{m}\right| + o_{\mathbb{P}}(1) \\ &= \frac{(m-1)(1-\frac{m}{m-1}\alpha)^\gamma}{m} + o_{\mathbb{P}}(1).\end{aligned} \tag{4.73}$$

Taking $\gamma \ll \alpha^{-1}$ and by monotonicity, we conclude that

$$\max_{x\in V} \left\| P^{\beta \boldsymbol{t}_{\rm ent}}(x,\cdot) - \pi \right\|_{\rm TV} \le \frac{m-1}{m} + o_{\mathbb{P}}(1). \tag{4.74}$$

Then, putting together Eqs. (4.70), (4.72), and (4.74), Eq. (4.5) follows. Similarly, in the case $\sqrt{n}\log(n)^{-2} \ll \alpha^{-1} \ll \lambda n \log(n)$, instead of Proposition 4.10 and Theorem 4.14, we use Proposition 4.15 and Theorem 4.16, leading to the same result.

Convergence to the global equilibrium. Let us first assume $\boldsymbol{t}_{\rm ent} \ll \alpha^{-1} \ll \sqrt{n}\log(n)^{-2}$. Using Eq. (4.71), with $t = \beta\alpha^{-1}$, for $\beta > 0$,

$$\min_{i\le m}\min_{x\in V_i} \|\mathbf{P}^G_x(X_t \in \cdot) - \pi\|_{\rm TV} \ge \frac{m-1}{m} \mathrm{e}^{-\frac{\beta m}{m-1}} + o_{\mathbb{P}}(1)\,.$$

For the upper bound, consider now, in Eq. (4.73), $\gamma = t - (1+\varepsilon)\boldsymbol{t}_{\rm ent} \sim \beta\alpha^{-1}$. Then, again by monotonicity, we have

$$\max_{x\in V} \|\mathbf{P}^G_x(X_t \in \cdot) - \pi\|_{\rm TV} \le \frac{m-1}{m} \mathrm{e}^{-\frac{\beta m}{m-1}} + o_{\mathbb{P}}(1)\,,$$

concluding the proof. Similarly, in the case $\sqrt{n}\log(n)^{-2} \ll \alpha^{-1} \ll \lambda n \log(n)$, instead of Proposition 4.10 and Theorem 4.14, we use Proposition 4.15 and Theorem 4.16 , leading to the same result. This proves Eq. (4.6). □

4.6.2 Subcritical regime

The upper bound in Eq. (4.3) (which is non-trivial only for $\beta > 1$) can be obtained observing that in the subcritical case $\alpha^{-1} \ll \boldsymbol{t}_{\rm ent}$, it holds

$$Q^\sigma(i,\cdot) = \frac{1}{m}(1+o(1)), \qquad \forall i \le m,$$

Then, recalling the definitions in Eqs. (4.61) and (4.65), it follows

$$\max_{i\le m} \|\nu_i - \nu\|_{\rm TV} = o_{\mathbb{P}}(1).$$

This, plugged in Eq. (4.67), concludes the upper bound. The lower bound in Eq. (4.3) (which is non-trivial only for $\beta < 1$) is precisely Eq. (4.68). □

4.6.3 Critical regime

We fix $C > 0$ and $t = \beta\boldsymbol{t}_{\rm ent}$, for some $\beta > 0$, $\beta \ne 1$. We want to prove Eq. (4.4). For $\beta < 1$, the bound is, again, precisely given by Eq. (4.68). Then, we consider $\beta > 1$. By a trivial bound and Proposition 4.10,

$$\begin{aligned}\min_{x\in V} \left\| P^{\beta \boldsymbol{t}_{\rm ent}}(x,\cdot) - \pi \right\|_{\rm TV} &\ge \min_{x\in V} \left| P^{\beta \boldsymbol{t}_{\rm ent}}(x, V_{c(x)}) - \pi(V_{c(x)}) \right| \\ &= \tfrac{m-1}{m}\left(1 - \tfrac{m}{m-1}\alpha\right)^{\beta \boldsymbol{t}_{\rm ent}} + o_{\mathbb{P}}(1) \\ &= \tfrac{m-1}{m}\mathrm{e}^{-\frac{\beta}{C}\frac{m}{m-1}}(1+o(1)) + o_{\mathbb{P}}(1).\end{aligned}$$

In the what follows we will prove the following result, which closes the discussion for $\beta > 1$.

Lemma 4.35. *For every* $0 < \varepsilon < \beta - 1$,

$$\max_{x\in V} \left\| P^{\beta \boldsymbol{t}_{\rm ent}}(x,\cdot) - \pi \right\|_{\rm TV} \le \tfrac{m-1}{m} \mathrm{e}^{-\frac{\beta}{C}\frac{m}{m-1}}(1+o(1)) + \frac{6m\varepsilon}{C} + o_{\mathbb{P}}(1).$$

Since $\varepsilon > 0$ can be taken arbitrarily small, we get Eq. (4.4). □

To prove Lemma 4.35, it is useful to consider, for every $i \le m$, the measures

$$\pi_{V_i}(x) = \frac{\pi(x)\mathbf{1}_{\{x\in V_i\}}}{\pi(V_i)}\,, \qquad x \in V.$$

Lemma 4.36. *Let* $\alpha^{-1} = C\boldsymbol{t}_{\rm ent}$*, for* $C > 0$*. Then*

$$\mathbb{P}\left(\max_{i\le m} \left\| \frac{1}{n}\sum_{y\in V_i} P^{\varepsilon \boldsymbol{t}_{\rm ent}}(y,\cdot) - \pi_{V_i} \right\|_{\rm TV} \le \frac{4m\varepsilon}{C} \right) = 1 - o(1).$$

Proof of Lemma 4.36. Using the inequality $|a+b| \ge |a| - |b|$, for $a, b \in \mathbb{R}$, we have

$$\begin{aligned}
&\left\| \frac{1}{mn}\sum_{y\in V} P^{\eta}(y,\cdot) - \frac{1}{m}\sum_{i=1}^{m}\pi_{V_i} \right\|_{\rm TV} \\
&\quad= \frac{1}{2}\sum_{j=1}^{m}\sum_{z\in V_j}\frac{1}{m}\left| \frac{1}{n}\sum_{y\in V_j}P^{\eta}(y,z) - \sum_{i=1}^{m}\pi_{V_i}(z) + \frac{1}{n}\sum_{y\in V_j^{\complement}}P^{\eta}(y,z) \right| \\
&\quad\ge \frac{1}{2}\sum_{j=1}^{m}\sum_{z\in V_j}\frac{1}{m}\left[\left| \frac{1}{n}\sum_{y\in V_j}P^{\eta}(y,z) - \pi_{V_j}(z) \right| - \frac{1}{n}\sum_{y\in V_j^{\complement}}P^{\eta}(y,z) \right] \\
&\quad= \frac{1}{2}\sum_{j=1}^{m}\frac{1}{m}\left[\sum_{z\in V_j}\left| \frac{1}{n}\sum_{y\in V_j}P^{\eta}(y,z) - \pi_{V_j}(z) \right| - \frac{1}{n}\sum_{y\in V_j^{\complement}}P^{\eta}(y,V_j) \right].
\end{aligned} \tag{4.75}$$

On the other hand, for every $j \le m$ it holds

$$\begin{aligned}
\left\| \frac{1}{n}\sum_{y\in V_j}P^{\eta}(y,\cdot) - \pi_{V_j} \right\|_{\rm TV} &= \frac{1}{2}\sum_{z\in V}\left| \frac{1}{n}\sum_{y\in V_j}P^{\eta}(y,z) - \pi_{V_j}(z) \right| \\
&= \frac{1}{2}\left[\sum_{z\in V_j}\left| \frac{1}{n}\sum_{y\in V_j}P^{\eta}(y,z) - \pi_{V_j}(z) \right| + \frac{1}{n}\sum_{y\in V_j}P^{\eta}(y,V_j^{\complement}) \right].
\end{aligned} \tag{4.76}$$

Putting Eqs. (4.75) and (4.76) together, it follows

$$\sum_{j=1}^{m} \frac{1}{m} \left\| \frac{1}{n} \sum_{y \in V_i} P^\eta(y,\cdot) - \pi_{V_i} \right\|_{\mathrm{TV}} \leq \left\| \frac{1}{mn} \sum_{y \in V} P^\eta(y,\cdot) - \frac{1}{m} \sum_{i=1}^{m} \pi_{V_i} \right\|_{\mathrm{TV}} + \Phi(\varepsilon), \quad (4.77)$$

where

$$\Phi(\varepsilon) := \frac{1}{2m} \sum_{j=1}^{m} \frac{1}{n} \Big(\sum_{y \in V_j^\complement} P^\eta(y, V_j) + \sum_{y \in V_j} P^\eta(y, V_j^\complement) \Big) = \frac{1}{m} \sum_{j=1}^{m} \frac{1}{n} \sum_{y \in V_j} P^\eta(y, V_j^\complement).$$

Invoking Proposition 4.10, we get

$$\left| \Phi(\varepsilon) - \frac{1}{m} \sum_{j=1}^{m} Q^\eta(j, [m] \setminus \{j\}) \right| = o_{\mathbb{P}}(1).$$

Using that, for $\boldsymbol{t}_{\mathrm{ent}} = \alpha^{-1}/C$, and for each $j \leq m$,

$$\begin{aligned} Q^\eta\left(j, [m] \setminus \{j\}\right) &= \frac{m-1}{m} \left(1 - \left(1 - \tfrac{m}{m-1}\alpha\right)^\eta\right) \\ &= \frac{m-1}{m} \left(1 - \mathrm{e}^{-\frac{m}{m-1}\frac{2\varepsilon}{C}}(1+o(1))\right) \leq \frac{2\varepsilon}{C}(1+o(1)), \end{aligned} \quad (4.78)$$

As a consequence, w.h.p. $\Phi(\varepsilon) \leq \frac{3\varepsilon}{C}$. Recalling that $\pi = \frac{1}{m}\sum_{j=1}^{m} \pi_{V_j}(1 + o_{\mathbb{P}}(1))$, thanks to Lemma 4.34, from Eq. (4.77) we conclude that w.h.p. it holds

$$\frac{1}{m} \sum_{j=1}^{m} \left\| \sum_{y \in V_j} \frac{1}{n} P^\eta(y,\cdot) - \pi_{V_j} \right\|_{\mathrm{TV}} \leq \left\| \frac{1}{mn} \sum_{y \in V} P^\eta(y,\cdot) - \pi \right\|_{\mathrm{TV}} + \frac{3\varepsilon}{C},$$

which is at most $\frac{3\varepsilon}{C} + o_{\mathbb{P}}(1)$, thanks to Eq. (4.66). Then, w.h.p., each term of the sum is bounded by $\frac{4m\varepsilon}{C}$. □

Proof of Lemma 4.35. We consider a time $t = \beta \boldsymbol{t}_{\mathrm{ent}}$, where $\varepsilon > 0$ is chosen such that $\beta > 1 + \varepsilon$. For $x \in V$, it holds

$$\left\| P^t(x,\cdot) - \pi \right\|_{\mathrm{TV}} \leq \left\| P^t(x,\cdot) - \sum_{j=1}^{m} Q^t(c(x), j) \pi_{V_j} \right\|_{\mathrm{TV}} + \left\| \sum_{j=1}^{m} Q^t(c(x), j) \pi_{V_j} - \pi \right\|_{\mathrm{TV}}.$$

Being $\pi = \sum_{j=1}^{m} \pi(V_j) \pi_{V_j}$, taking the maximum over $x \in V$, the second summand provides

$$\begin{aligned} \max_{x \in V} \left\| \sum_{j=1}^{m} Q^t(c(x), j)\pi_{V_j} - \sum_{j=1}^{m} \pi(V_j)\pi_{V_j} \right\|_{\mathrm{TV}} &= \max_{x \in V} \frac{1}{2} \sum_{j=1}^{m} \left| Q^t(c(x), j) - \pi(V_j) \right| \\ \leq \max_{x \in V} \frac{1}{2} \sum_{j=1}^{m} \left| Q^t(c(x), j) - \frac{1}{m} \right| + o_{\mathbb{P}}(1) &= \frac{m-1}{m} \left(1 - \frac{m}{m-1}\alpha\right)^t + o_{\mathbb{P}}(1), \end{aligned} \quad (4.79)$$

where we have employed Lemma 4.34 and the triangle inequality in the second line. This will provide the leading term. We now consider the first summand. By Proposition 4.10, it holds

$$\max_{x\in V}\max_{j\le m}\left|P^{(\beta-1-\varepsilon)\boldsymbol{t}_{\mathrm{ent}}}(x,V_j)-Q^{(\beta-1-\varepsilon)\boldsymbol{t}_{\mathrm{ent}}}(c(x),j)\right|=o_{\mathbb{P}}(1).$$

Then

$$\begin{aligned}
&\max_{x\in V}\left\|P^{\beta\boldsymbol{t}_{\mathrm{ent}}}(x,\cdot)-\sum_{j=1}^m Q^{\beta\boldsymbol{t}_{\mathrm{ent}}}(c(x),j)\pi_{V_j}\right\|_{TV}\\
&\le\max_{x\in V}\left\|\sum_{i=1}^m\sum_{z\in V_i}P^{(\beta-1-\varepsilon)\boldsymbol{t}_{\mathrm{ent}}}(x,z)\left[P^{(1+\varepsilon)\boldsymbol{t}_{\mathrm{ent}}}(z,\cdot)-\sum_{j=1}^m Q^{(1+\varepsilon)\boldsymbol{t}_{\mathrm{ent}}}(i,j)\pi_{V_j}\right]\right\|_{\mathrm{TV}}+o_{\mathbb{P}}(1)\\
&\le\max\max_{z\in V_i}\left\|P^{(1+\varepsilon)\boldsymbol{t}_{\mathrm{ent}}}(z,\cdot)-\sum_{j=1}^m Q^{(1+\varepsilon)\boldsymbol{t}_{\mathrm{ent}}}(i,j)\pi_{V_j}\right\|_{\mathrm{TV}}+o_{\mathbb{P}}(1)\\
&\le\max_{i\le m}\left\|\sum_{k=1}^m\frac{Q^{\sigma+1}(i,k)}{n}\sum_{y\in V_k}P^{\eta}(y,\cdot)-\sum_{j=1}^m Q^{(1+\varepsilon)\boldsymbol{t}_{\mathrm{ent}}}(i,j)\pi_{V_j}\right\|_{\mathrm{TV}}+o_{\mathbb{P}}(1)
\end{aligned}\tag{4.80}$$

where the third inequality follows from Eq. (4.64) (recall that, according to Definition 4.31, $(1+\varepsilon)\boldsymbol{t}_{\mathrm{ent}}=\sigma+\eta+1$) Using the semigroup property of $Q(\cdot,\cdot)$, we can bound the last expression in Eq. (4.80) by

$$\begin{aligned}
&\max_{i\le m}\sum_{k=1}^m Q^{\sigma+1}(i,k)\left\|\frac{1}{n}\sum_{y\in V_k}P^{\eta}(y,\cdot)-\sum_{j=1}^m Q^{\eta}(k,j)\pi_{V_j}\right\|_{\mathrm{TV}}+o_{\mathbb{P}}(1)\\
&\le\max_{k\le m}\left(\left\|\frac{1}{n}\sum_{y\in V_k}P^{\eta}(y,\cdot)-\pi_{V_k}\right\|_{\mathrm{TV}}+\left\|\pi_{V_k}-\sum_{j=1}^m Q^{\eta}(k,j)\pi_{V_j}\right\|_{\mathrm{TV}}\right)+o_{\mathbb{P}}(1),
\end{aligned}$$

Now w.h.p. we can bound the first summand in the last display with $\frac{4m\varepsilon}{C}$ by Lemma 4.36. The second summand has the same magnitude: reasoning in the spirit of Eqs. (4.79) and (4.78), it holds

$$\begin{aligned}
\max_{k\le m}\left\|\pi_{V_k}-\sum_{j=1}^m Q^{\eta}(k,j)\pi_{V_j}\right\|_{\mathrm{TV}}&=\max_{k\le m}\frac{1}{2}\left|1-Q^{\eta}(k,k)\right|+\frac{1}{2}\sum_{j\ne k}Q^{\eta}(k,j)\\
&=\max_{k\le m}\frac{m-1}{m}\left(1-\left(1-\tfrac{m}{m-1}\alpha\right)^{\eta}\right)\le\frac{2\varepsilon}{C}(1+o(1)).
\end{aligned}$$

□

Bibliography

E. Abbe. Community detection and stochastic block models: recent developments. *Journal of Machine Learning Research*, 18(177):1–86, 2018.

R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–97, 2002.

R. Albert and A.-L. Barabási. Emergence of scaling in random networks. *Science*, 286 (5439):509–512, 1999.

D. Aldous. Markov chains with almost exponential hitting times. *Stochastic Process. Appl.*, 13(3):305–310, 1982.

D. Aldous. Random walks on finite groups and rapidly mixing Markov chains. In *Seminar on probability, XVII*, volume 986 of *Lecture Notes in Math.*, pages 243–297. Springer, Berlin, 1983.

D. Aldous and P. Diaconis. Shuffling cards and stopping times. *American Mathematical Monthly*, 93(5):333–348, 1986.

G. W. Anderson, A. Guionnet, and O. Zeitouni. *An Introduction to Random Matrices.* Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2009.

L. Andreis, W. König, and R. I. A. Patterson. A large-deviations principle for all the cluster sizes of a sparse Erdős-Rényi graph. *Random Structures Algorithms*, 59(4): 522–553, 2021.

L. Andreis, W. König, H. Langhammer, and R. I. A. Patterson. A large-deviations principle for all the components in a sparse inhomogeneous random graph. *Probab. Theory Related Fields*, 186(1-2):521–620, 2023.

L. Avena, H. Güldaş, F. den Hollander, and R. van der Hofstad. Random walks on dynamic configuration models: a trichotomy. *Stochastic Processes Appl.*, 129(9):3360–3375, 2019.

L. Avena, H. Güldaş, F. den Hollander, R. van der Hostad, and O. Nagy. Linking the mixing times of random walks on static and dynamic random graphs. *Stochastic Processes Appl.*, 153:145–182, 2022.

L. Avena, F. Capannoli, R. S. Hazra, and M. Quattropani. Meeting, coalescence and consensus time on random directed graphs. *Annals of Applied Probability*, 34(5):4940–4997, 2024.

J. Baik, G. Ben Arous, and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.*, 33(5):1643–1697, 2005.

A. Basak and M. Rudelson. The circular law for sparse non-Hermitian matrices. *Ann. Probab.*, 47(4):2359–2416, 2019.

F. L. Bauer and C. T. Fike. Norms and exclusion theorems. *Numer. Math.*, 2:137–141, 1960.

A. Ben-Hamou. A threshold for cutoff in two-community random graphs. *Ann. Appl. Probab.*, 30(4):1824–1846, 2020.

A. Ben-Hamou and J. Salez. Cutoff for nonbacktracking random walks on sparse random graphs. *Ann. Probab.*, 45(3):1752–1770, 2017.

A. Ben-Hamou, E. Lubetzky, and Y. Peres. Comparing mixing times on sparse random graphs. *Ann. Inst. Henri Poincaré Probab. Stat.*, 55:1116–1130, 2019.

I. Bena, M. Droz, and A. Lipowski. Statistical mechanics of equilibrium and nonequilibrium phase transitions: the yang–lee formalism. *International Journal of Modern Physics B*, 19(29):4269–4329, 2005.

F. Benaych-Georges and R. R. Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.*, 227(1):494–521, 2011.

I. Benjamini and O. Schramm. Recurrence of distributional limits of finite planar graphs. *Electron. J. Probab.*, 6:no. 23, 13, 2001.

I. Benjamini, G. Kozma, and N. Wormald. The mixing time of the giant component of a random graph. *Random Structures Algorithms*, 45(3):383–407, 2014.

N. Berestycki, E. Lubetzky, Y. Peres, and A. Sly. Random walks on the random graph. *Ann. Probab.*, 46(1):456–490, 2018.

S. Bhamidi, G. Bresler, and A. Sly. Mixing time of exponential random graphs. *Ann. Appl. Probab.*, 21(6):2146–2170, 2011.

S. Bhamidi, J. Hannig, C. Y. Lee, and J. Nolen. The importance sampling technique for understanding rare events in Erdős–Rényi random graphs. *Electron. J. Probab.*, 20(107):1–30, 2015.

S. Bhamidi, D. Nam, O. Nguyen, and A. Sly. Survival and extinction of epidemics on random graphs with general degree. *Ann. Probab.*, 49(1):244–286, 2021.

A. Bianchi and G. Passuello. Mixing cutoff for simple random walks on the Chung-Lu digraph. *Random Structures Algorithms*, 66(1):1–20, 2025.

A. Bianchi, F. Collet, and E. Magnanini. The GHS and other inequalities for the two-star model. *ALEA Lat. Am. J. Probab. Math. Stat.*, 19(2):1679–1695, 2022.

A. Bianchi, F. Collet, and E. Magnanini. Limit theorems for Exponential Random Graphs. *Ann. Appl. Probab.*, 34(5):4863–4898, 2024.

A. Bianchi, G. Passuello, and M. Quattropani. Mixing trichotomy for random walks on directed stochastic block models. *arXiv preprint arXiv:2504.06851*, 2025.

P. Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, second edition, 1986.

B. Bollobás. *Random Graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2nd edition, 2001.

C. Bordenave. A new proof of Friedman's second eigenvalue theorem and its extension to random lifts. *Ann. Sci. Éc. Norm. Supér. (4)*, 53(6):1393–1439, 2020.

C. Bordenave and D. Chafaï. Around the circular law. *Probability Surveys*, 9:1–89, 2012.

C. Bordenave and H. Lacoin. Cutoff at the entropic time for random walks on covered expander graphs. *J. Inst. Math. Jussieu*, 21(5):1571–1616, 2022.

C. Bordenave, P. Caputo, and D. Chafaï. Circular law theorem for random Markov matrices. *Probab. Theory Related Fields*, 152(3-4):751–779, 2012.

C. Bordenave, P. Caputo, and D. Chafaï. Spectrum of Markov generators on sparse random graphs. *Comm. Pure Appl. Math.*, 67(4):621–669, 2014.

C. Bordenave, P. Caputo, and J. Salez. Random walk on sparse random digraphs. *Probab. Theory Relat. Fields*, 170(3):933–960, 2018.

C. Bordenave, P. Caputo, and J. Salez. Cutoff at the "entropic time" for sparse Markov chains. *Probab. Theory Relat. Fields*, 173(1):261–292, 2019.

C. Borgs, J. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs I. Subgraph frequencies, metric properties and testing. *Adv. Math.*, 219(6):1801–1851, 2008.

C. Borgs, J. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. Convergent sequences of dense graphs II. Multiway cuts and statistical physics. *Ann. Math.*, 176(1):151–219, 2012.

A. Bovier, S. Marello, and E. Pulvirenti. Metastability for the dilute Curie-Weiss model with Glauber dynamics. *Electron. J. Probab.*, 26:Paper No. 47, 38, 2021.

T. Brailovskaya and R. van Handel. Universality and sharp matrix concentration inequalities. *Geom. Funct. Anal.*, 34(6):1734–1838, 2024.

X. S. Cai, P. Caputo, G. Perarnau, and M. Quattropani. Rankings in directed configuration models with heavy tailed in-degrees. *Ann. Appl. Probab.*, 33(6B):5613–5667, 2023.

A. Caimo and I. Gollini. Recent advances in exponential random graph modelling. *Mathematical Proceedings of the Royal Irish Academy*, 123A(1):1–12, 2023.

F. Capannoli. Evolution of discordant edges in the voter model on random sparse digraphs. *Electron. J. Probab.*, 30:Paper No. 6, 24, 2025.

P. Caputo and M. Quattropani. Stationary distribution and cover time of sparse directed configuration models. *Probab. Theory Relat. Fields*, 178:1011–1066, 2020.

P. Caputo and M. Quattropani. Mixing time of Pagerank surfers on sparse random digraphs. *Random Structures Algorithms*, 59:376–406, 2021a.

P. Caputo and M. Quattropani. Mixing time trichotomy in regenerating dynamic digraphs. *Stochastic Processes Appl.*, 137:222–251, 2021b.

P. Caputo, M. Quattropani, and F. Sau. Cutoff for the averaging process on the hypercube and complete bipartite graphs. *Electron. J. Probab.*, 28:Paper No. 100, 31, 2023.

P. Caputo, M. Quattropani, and F. Sau. Repeated Block Averages: entropic time and mixing profiles. *arXiv preprint arXiv:2407.16656*, 2024.

A. Chakrabarty, S. Chakraborty, and R. S. Hazra. Eigenvalues outside the bulk of inhomogeneous Erdős-Rényi random graphs. *J. Stat. Phys.*, 181(5):1746–1780, 2020.

A. Chakrabarty, R. S. Hazra, F. den Hollander, and M. Sfragara. Spectra of adjacency and Laplacian matrices of inhomogeneous Erdős-Rényi random graphs. *Random Matrices Theory Appl.*, 10(1):Paper No. 2150009, 34, 2021.

A. Chakrabarty, R. S. Hazra, F. den Hollander, and M. Sfragara. Large deviation principle for the maximal eigenvalue of inhomogeneous Erdős-Rényi random graphs. *J. Theoret. Probab.*, 35(4):2413–2441, 2022.

S. Chatterjee. An introduction to large deviations for random graphs. *Bull. Amer. Math. Soc. (N.S.)*, 53(4):617–642, 2016.

S. Chatterjee. Spectral gap of nonreversible Markov chains. *Ann. Appl. Probab.*, 35(4): 2644–2677, 2025.

S. Chatterjee and P. Dey. Applications of Stein's method for concentration inequalities. *Ann. Probab.*, 38(6):2443–2485, 2010.

S. Chatterjee and P. Diaconis. Estimating and understanding exponential random graph models. *Ann. Stat.*, 41(5):2428–2461, 2013.

S. Chatterjee and R. Durrett. Contact processes on random graphs with power law degree distributions have critical value 0. *Ann. Probab.*, 37(6):2332–2356, 2009.

S. Chatterjee and S. R. S. Varadhan. The large deviation principle for the Erdős–Rényi random graph. *European J. Combin.*, 32(7):1000–1017, 2011.

S. Chatterjee, P. Diaconis, A. Sly, and L. Zhang. A phase transition for repeated averages. *Ann. Probab.*, 50(1):1–17, 2022.

N. Chen, N. Litvak, and M. Olvera-Cravioto. Pagerank in scale-free random graphs. In *Algorithms and models for the web graph*, volume 8882 of *Lecture Notes in Comput. Sci.*, pages 120–131. Springer, Cham, 2014.

N. Chen, N. Litvak, and M. Olvera-Cravioto. Generalized Pagerank on directed configuration networks. *Random Structures Algorithms*, 51(2):237–274, 2017.

F. Chung and L. Lu. Connected components in random graphs with given expected degree sequences. *Ann. Comb.*, 6(2):125–145, 2002a.

F. Chung and L. Lu. The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences of the United States of America*, 99(25):15879–15882, 2002b.

F. Chung and L. Lu. Connected components in random graphs with given expected degree sequences. *Annals of Combinatorics*, 6(2):125–145, 2002c.

F. Chung and L. Lu. The average distance in a random graph with given expected degrees. *Internet Mathematics*, 1(1):91–113, 2003.

F. Chung and L. Lu. *Complex Graphs and Networks*, volume 107 of *CBMS Regional Conference Series in Mathematics*. Conference Board of the Mathematical Sciences, Washington, DC, 2006a.

F. Chung and L. Lu. The volume of the giant component of a random graph with given expected degrees. *SIAM Journal on Discrete Mathematics*, 20(2):395–411, 2006b.

A. Cipriani, R. S. Hazra, N. Malhotra, and M. Salvi. The spectrum of dense kernel-based random graphs. *arXiv preprint arXiv:2506.15159*, 2025.

G. Conchon-Kerjan. Cutoff for random lifts of weighted graphs. *Ann. Probab.*, 50(1): 304–338, 2022.

C. Cooper and A. Frieze. Stationary distribution and cover time of random walks on random di-graphs. *J. Comb. Theory Ser. B*, 102(2):329–362, 2012.

C. Cooper, A. Frieze, and T. Radzik. Multiple random walks in random regular graphs. *SIAM J. Discrete Math.*, 23(4):1738–1761, 2009/10.

S. Coste. The spectral gap of sparse random digraphs. *Ann. Inst. Henri Poincaré Probab. Stat.*, 57(2):644–684, 2021.

A. Dembo and A. Montanari. Ising models on locally tree-like graphs. *Ann. Appl. Probab.*, 20(2):565–592, 2010.

P. Diaconis. The cutoff phenomenon in finite Markov chains. *Proc. Natl. Acad. Sci. USA*, 93(4):1659–1664, 1996.

P. Diaconis and L. Saloff-Coste. Logarithmic sobolev inequalities for finite Markov chains. *Annals of Applied Probability*, 6(3):695–750, 1996.

P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete*, 57(2):159–179, 1981.

J. Ding, E. Lubetzky, and Y. Peres. Total variation cutoff in birth-and-death chains. *Probab. Theory Related Fields*, 146(1-2):61–85, 2010.

S. Dommers. Metastability of the Ising model on random regular graphs at zero temperature. *Probab. Theory Related Fields*, 167(1-2):305–324, 2017.

S. Dommers, F. den Hollander, O. Jovanovski, and F. R. Nardi. Metastability for Glauber dynamics on random graphs. *Ann. Appl. Probab.*, 27(4):2130–2158, 2017.

B. Dubail. Cutoff for mixtures of permuted Markov chains: reversible case. *arXiv preprint arXiv:2401.03937*, 2024a.

B. Dubail. Cutoff for mixtures of permuted Markov chains: general case. *arXiv preprint arXiv:2402.03415*, 2024b.

I. Dumitriu and Y. Zhu. Extreme singular values of inhomogeneous sparse random rectangular matrices. *Bernoulli*, 30(4), Nov. 2024.

R. Durrett. *Random graph dynamics*, volume 20 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2007.

P. Eichelsbacher and M. Löwe. Stein's method for dependent random variables occurring in statistical mechanics. *Electron. J. Probab.*, 15(30):962–988, 2010.

R. S. Ellis. *Entropy, Large Deviations, and Statistical Mechanics*. Springer, New York, 1985.

R. S. Ellis and C. M. Newman. Limit theorems for sums of dependent random variables occurring in statistical mechanics. *Z. Wahrsch. Verw. Gebiete*, 44(2):117–139, 1978.

P. Erdős and A. Rényi. On the evolution of random graphs. *Bull. Inst. Internat. Statist.*, 38:343–347, 1961.

L. Erdős and H.-T. Yau. *A dynamical approach to random matrix theory*, volume 28 of *Courant Lecture Notes in Mathematics*. Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2017.

L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Spectral statistics of Erdős–Rényi graphs I: Local semicircle law. *Ann. Probab.*, 41(3B):2279–2375, 2013.

P. Erdős and A. Rényi. On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:17–61, 1960.

X. Fang, S.-H. Liu, Q.-M. Shao, and Y.-K. Zhao. Normal approximation for exponential random graphs. *Probability Theory and Related Fields*, pages 1–40, 2025a.

X. Fang, S.-H. Liu, Z. Su, and X. Wang. Conditional central limit theorems for exponential random graphs. *arXiv preprint arXiv:2506.15159*, 2025b.

J. Fernley and M. Ortgiese. Voter models on subcritical scale-free random graphs. *Random Structures Algorithms*, 62(2):376–429, 2023.

A. Fischer, G. Reinert, and W. Xu. Stein's method of moment estimators for local dependency exponential random graph models. *arXiv:2503.13191*, 2025. Preprint, March 17 2025.

N. Fountoulakis and B. Reed. The evolution of the mixing rate of a simple random walk on the giant component of a random graph. *Random Structures Algorithms*, 33(1): 68–86, 2008.

D. A. Freedman. On tail probabilities for martingales. *Ann. Probab.*, 3(1):100–118, 1975.

S. Friedli and Y. Velenik. *Statistical Mechanics of Lattice Systems: A Concrete Mathematical Introduction*. Cambridge University Press, Cambridge, 2017.

A. Frieze and M. Karoński. *Introduction to Random Graphs*, volume 84 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 2016.

Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1:233–241, 1981.

A. Garavaglia, R. van der Hofstad, and N. Litvak. Local weak convergence for Pagerank. *Ann. Appl. Probab.*, 30(1):40–79, 2020.

C. Giardinà, C. Giberti, and E. Magnanini. Approximating the cumulant generating function of triangles in the Erdős-Rényi random graph. *J. Stat. Phys.*, 182(2):Paper No. 23, 1–22, 2021.

F. Götze and A. Tikhomirov. The circular law for random matrices. *Ann. Probab.*, 38 (4):1444–1491, 2010.

F. Götze and A. Tikhomirov. On the largest and the smallest singular value of sparse rectangular random matrices. *Electron. J. Probab.*, 28:Paper No. 27, 18, 2023.

F. Götze, H. Kösters, and A. Tikhomirov. Asymptotic spectra of matrix-valued functions of independent random matrices and free probability. *Random Matrices Theory Appl.*, 4(2):1550005, 85, 2015.

R. B. Griffiths, C. A. Hurst, and S. Sherman. Concavity of magnetization of an Ising ferromagnet in a positive external field. *J. Math. Phys.*, 11(3):790–795, 1970.

H. Guggenheimer, A. Edelman, and C. Johnson. A simple estimate of the condition number of a linear system. *Linear Algebra and Its Applications*, pages 2–5, 1995. doi: 10.1080/07468342.1995.11973657.

J. Harris. *An Introduction to Exponential Random Graph Modeling*. Quantitative Applications in the Social Sciences. SAGE Publications, 2013.

J. Hermon, S. Li, D. Yao, and L. Zhang. Mean field behavior during the big bang regime for coalescing random walks. *Ann. Probab.*, 50(5):1813–1884, 2022.

J. Hermon, A. Šarković, and P. Sousi. Cutoff for random walk on random graphs with a community structure. *Ann. Appl. Probab.*, 35(3):2080–2127, 2025.

M. Hildebrand. Random walks on random simple graphs. *Random Structures Algorithms*, 8(4):301–318, 1996.

R. van der Hofstad. *Random Graphs and Complex Networks*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2016.

R. van der Hofstad. *Random Graphs and Complex Networks: Volume 2*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2024.

R. van der Hofstad. Stochastic processes on random graphs, 2025. URL `https://rhofstad.win.tue.nl/Saint_Flour_SPoRG_Grundlehre.pdf`. Monograph.

F. den Hollander and O. Jovanovski. Glauber dynamics on the Erdős-Rényi random graph. In *In and out of equilibrium 3. Celebrating Vladas Sidoravicius*, volume 77 of *Progr. Probab.*, pages 519–589. Birkhäuser/Springer, Cham, 2021.

F. den Hollander, M. Mandjes, A. Roccaverde, and N. Starreveld. Ensemble equivalence for dense graphs. *Electron. J. Probab.*, 23:12–26, 2018.

R. Kenyon and M. Yin. On the asymptotics of constrained exponential random graphs. *J. Appl. Probab.*, 54(1):165–180, 2017.

A. Klenke. *Probability theory—a comprehensive course*. Universitext. Springer, Cham, third edition, 2020.

H. Kösters and A. Tikhomirov. Limiting spectral distributions of sums of products of non-Hermitian random matrices. *Probab. Math. Statist.*, 38(2):359–384, 2018. [On table of contents: Vol. 33 (2013)].

H. Lacoin. The cutoff profile for the simple exclusion process on the circle. *Ann. Probab.*, 44(5):3399–3430, 2016a.

H. Lacoin. Mixing time and cutoff for the adjacent transposition shuffle and the simple exclusion. *Ann. Probab.*, 44(2):1426–1487, 2016b.

T. D. Lee and C. N. Yang. Statistical theory of equations of state and phase transitions. i. theory of condensation. *Phys. Rev. (2)*, 87:404–409, 1952.

D. Levin and Y. Peres. *Markov Chains and Mixing Times*. American Mathematical Society, Providence, RI, 2nd edition, 2017. With contributions by Elizabeth L. Wilmer, and chapter on "Coupling from the past" by James G. Propp and David B. Wilson.

T. M. Liggett. *Interacting particle systems*, volume 276 of *Grundlehren der mathematischen Wissenschaften (Fundamental Principles of Mathematical Sciences)*. Springer-Verlag, New York, 1985.

T. M. Liggett. *Stochastic interacting systems: contact, voter and exclusion processes*, volume 324 of *Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1999.

A. E. Litvak, A. Lytova, K. Tikhomirov, N. Tomczak-Jaegermann, and P. Youssef. Circular law for sparse random regular digraphs. *J. Eur. Math. Soc. (JEMS)*, 23(2): 467–501, 2021.

L. Lovász. *Large Networks and Graph Limits*. American Mathematical Society, Providence, 2012.

L. Lovász and B. Szegedy. Szemerédi's lemma for the analyst. *Geom. Funct. Anal.*, 17 (1):252–270, 2007.

L. Lovász. *Large Networks and Graph Limits*. American Mathematical Society, 2012.

L. Lu. *Probabilistic methods in massive graphs and Internet computing*. PhD thesis, University of California, San Diego, 2002. URL `http://math.ucsd.edu/~llu/thesis.pdf`.

E. Lubetzky and Y. Peres. Cutoff on all Ramanujan graphs. *Geom. Funct. Anal.*, 26: 1190–1216, 2016.

E. Lubetzky and A. Sly. Cutoff phenomena for random walks on random regular graphs. *Duke Math. J.*, 153(3):475–510, 2010.

E. Lubetzky and A. Sly. Cutoff for the Ising model on the lattice. *Invent. Math.*, 191 (3):719–755, 2013.

E. Lubetzky and A. Sly. Cutoff for general spin systems with arbitrary boundary conditions. *Comm. Pure Appl. Math.*, 67(6):982–1027, 2014.

D. Lusher, J. Koskinen, and G. Robins. *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. Structural Analysis in the Social Sciences. Cambridge University Press, 2012.

E. Magnanini and G. Passuello. Statistics for the triangle density in ERGM and its mean-field approximation. *J. Stat. Phys.*, 192(4):Paper No. 54, 31, 2025a.

E. Magnanini and G. Passuello. A standard CLT for triangles in a class of ERGs. *arXiv preprint arXiv:2507.17484*, 2025b.

F. Manzo, M. Quattropani, and E. Scoppola. A probabilistic proof of Cooper & Frieze's "first visit time lemma". *ALEA Lat. Am. J. Probab. Math. Stat.*, 18(2):1739–1758, 2021.

M. L. Mehta. *Random matrices and the statistical theory of energy levels.* Academic Press, New York-London, 1967.

M. L. Mehta. *Random matrices.* Academic Press, Inc., Boston, MA, second edition, 1991.

J. A. Mingo and R. Speicher. *Free probability and random matrices*, volume 35 of *Fields Institute Monographs.* Springer, New York; Fields Institute for Research in Mathematical Sciences, Toronto, ON, 2017.

T. Mountford, D. Valesin, and Q. Yao. Metastable densities for the contact process on power law random graphs. *Electron. J. Probab.*, 18:No. 103, 36, 2013.

J.-C. Mourrat and D. Valesin. Phase transition of the contact process on random regular graphs. *Electron. J. Probab.*, 21:Paper No. 31, 17, 2016.

S. Mukherjee and Y. Xu. Statistics of the two star ERGM. *Bernoulli*, 29(1):24–51, 2023.

J. Neeman, C. Radin, and L. Sadun. Typical large graphs with given edge and triangle densities. *Probability Theory and Related Fields*, 186(3-4):1167–1223, 2023.

M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45 (2):167–256, 2003. electronic.

M. E. J. Newman. *Networks: An Introduction.* Oxford University Press, 2010.

R. I. Oliveira. On the coalescence time of reversible random walks. *Trans. Amer. Math. Soc.*, 364(4):2109–2128, 2012.

L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking : Bringing order to the web. In *The Web Conference*, 1999. URL `https://api.semanticscholar.org/CorpusID:1508503`.

F. Pedrotti and J. Salez. A new cutoff criterion for non-negatively curved chains. *arXiv preprint arXiv:2501.13079*, 2025.

A. Pizzo, D. Renfrew, and A. Soshnikov. On finite rank deformations of Wigner matrices. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(1):64–94, 2013.

M. Quattropani. Mixing trichotomy for an Ehrenfest urn with impurities. *Electron. Commun. Probab.*, 29:13, 2024.

M. Quattropani and F. Sau. On the meeting of random walks on random dfa. *Stochastic Process. Appl.*, 166:104225, 2023.

C. Radin and L. Sadun. Optimal graphons in the edge-2star model. *arXiv preprint arXiv:2305.00333*, 2023.

C. Radin and M. Yin. Phase transitions in exponential random graphs. *Ann. Appl. Probab.*, 23(6):2458–2471, 2013.

G. Reinert and N. Ross. Approximating stationary distributions of fast mixing Glauber dynamics, with applications to exponential random graphs. *The Annals of Applied Probability*, 29(5):3201–3229, 2019.

M. Rudelson and K. Tikhomirov. The sparse circular law under minimal assumptions. *Geom. Funct. Anal.*, 29(2):561–637, 2019.

A. Sah, J. Sahasrabudhe, and M. Sawhney. The limiting spectral law for sparse iid matrices. *arXiv preprint arXiv:2310.17635*, 2025.

J. Salez. Cutoff for non-negatively curved Markov chains. *J. Eur. Math. Soc. (JEMS)*, 26(11):4375–4392, 2024a.

J. Salez. The varentropy criterion is sharp on expanders. *Ann. H. Lebesgue*, 7:239–250, 2024b.

P. Sousi and S. Thomas. Cutoff for random walk on dynamical Erdős-Rényi graph. *Ann. Inst. Henri Poincaré Probab. Stat.*, 56(4):2745–2773, 2020.

C. Stein. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2*, pages 583–603. University of California Press, 1972.

D. Strauss. On a general class of models for interaction. *SIAM Rev.*, 28(4):513–527, 1986.

T. Tao. Outliers in the spectrum of iid matrices with bounded rank perturbations. *Probab. Theory Related Fields*, 155(1-2):231–263, 2013.

T. Tao and V. Vu. Random matrices: the circular law. *Commun. Contemp. Math.*, 10 (2):261–307, 2008.

T. Tao and V. Vu. Random matrices: universality of ESDs and the circular law. *Ann. Probab.*, 38(5):2023–2065, 2010. With an appendix by Manjunath Krishnapur.

C. A. Tracy and H. Widom. Level-spacing distributions and the airy kernel. *Physics Letters B*, 305(1):115–118, 1993.

C. A. Tracy and H. Widom. On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.*, 177(3):727–754, 1996.

R. S. Varga. *Geršgorin and his circles*, volume 36 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2004.

D. Vial and V. Subramanian. Restart perturbations for reversible Markov chains: trichotomy and pre-cutoff equivalence. *Random Structures Algorithms*, 66(3):Paper No. e70007, 27, 2025.

E. P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, 62(3):548–564, 1955.

V. Winstein. Quantitative central limit theorems for exponential random graphs. *arXiv preprint arXiv:2507.10531*, 2025.

W. Xu and G. Reinert. A stein goodness-of-test for exponential random graph models. In A. Banerjee and K. Fukumizu, editors, *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS)*, Proceedings of Machine Learning Research, vol. 130, pages 415–423. PMLR, April 2021.

M. Yin. Critical phenomena in exponential random graphs. *J. Stat. Phys.*, 153(6): 1008–1021, 2013.

O. Zeitouni. Random walks in random environment. In *Lectures on probability theory and statistics*, volume 1837 of *Lecture Notes in Math.*, pages 189–312. Springer, Berlin, 2004.

Acknowledgments

At the end of this doctoral experience, there are several people I wish to thank.

I would like to thank my supervisor Alessandra Bianchi for her guidance through the steps of this PhD. She introduced me to the field of random networks, providing, over these three years, an invaluable contribution to my professional and personal development, for which I will always be grateful.

I sincerely thank my co-supervisor Matteo Quattropani, for his attentive, patient, and steady presence and for always having the right theorem (and word) at hand. I feel fortunate to know that Italian universities host such talented and valuable people.

I am thankful to Alexandre Gaudillière for the stimulating discussions and his kindness during my research visit at Aix-Marseille Université, in France. It has been an extremely enriching experience.

I am grateful to Rajat Hazra for his supervision during my stay at the Leiden Mathematical Institute, in the Netherlands. I am glad that I will have such a capable and thoughtful collaborator nearby in the coming months.

I feel privileged to be part of a stimulating and friendly academic community such as the one of Italian probabilists. A special thanks goes to Elena Magnanini for discussions and our collaboration and Francesca Collet for her support.
I also thank Beatriz, Brune, Bruno, Carlo, Elisa, Emanuele, Federica, Federico, Félix, Francesco, Lars, Mario, Riccardo, Simone, Tommaso, Vanessa and many other people I met during conferences and various mathematical events over the last three years.

Turning to the community of Torre Archimede, in addition to thanking the probability group, I would like to thank Paola Mannucci, Daniela Tonon, Giulia Treu, the Equal Opportunities Committee, and everyone in the department who, alongside their professionalism, contributes to making Torre Archimede a welcoming environment. This is among the reasons why, after eight years, I will consider this department as a home.

I would like to thank the colleagues of the doctoral school, and especially the ones of the XXXVIII cycle, for mutual support, chats, discussions, and a lot of good moments. In particular, let me thank Martina for introducing us to after-lunch crossword sessions, Marco for his effort in bringing us together in a pleasant (and occasionally alcoholic) social routine, and Tommaso for many insightful conversations and his friendship. These are memories I will always treasure.

Let me conclude with a few acknowledgments that go beyond the academic setting. Thanks to Chiara for showing me that strong bonds can withstand time and distance. Thanks to Anna for being a trusted companion in this beautiful and random world. Thanks to Pietro, a close friend and choir mate, with whom I have shared many experiences and adventures. Thanks to Diletta, who has recently become a colleague, but whose gentle presence and depth have always accompanied me in the past eight years.

Thanks to my parents, siblings, and to the relatives who have shared with me, from Earth or Heaven, a part of this journey. A special thanks goes to Caterina, Michele, Angela and Daniele for having become, over the last five years, my second family.
Finally, my deepest gratitude goes to Silvia. You bring light into my life and teach me every day the meaning of the word *love*.