Wireless random-access networks and spectra of random graphs

Sfragara, M.

Wireless Random-Access Networks and Spectra of Random Graphs

Matteo Sfragara

Cover: images by NASA/NOAA, design by Matteo Blandford.
Chapter title pages: images by NASA/NOAA.

The research in this thesis was supported by the Netherlands Organisation for Scientific Research though the NWO Gravitation Programme NETWORKS, Grant Number 024.002.003.

Wireless Random-Access Networks and Spectra of Random Graphs

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op woensdag 28 oktober 2020
klokke 10.00 uur

door

Matteo Sfragara
geboren te Pescara
in 1991

Promotores:

Prof. dr. W.T.F. den Hollander
Prof. dr. S.C. Borst (Technische Universiteit Eindhoven)
Dr. F.R. Nardi (Università degli Studi di Firenze, Technische Universiteit Eindhoven)

Promotiecommissie:

Prof. dr. E.R. Eliel
Prof. dr. S.J. Edixhoven
Prof. dr. F. Baccelli (École Normale Supérieure, Paris)
Prof. dr. M.R.H. Mandjes (Universiteit van Amsterdam)
Dr. L. Miclo (Université Toulouse III - Paul Sabatier)

Contents

CHAPTER 1
Introduction

The present thesis consists of two parts: Part I focuses on metastability properties of queue-based random-access protocols for wireless networks; Part II focuses on spectra of inhomogeneous Erdős-Rényi random graphs.

§1.1 Introduction to Part I

In the first part of this thesis we study mathematical models that address fundamental challenges in wireless networks. We describe the collective behavior of devices sharing a wireless medium and we analyze various distributed protocols to improve the performance of the network.

In Section 1.1.1 we give an overview of wireless networks and describe the Carrier-Sense Multiple-Access (CSMA) protocol, which is the main object of study of the first part of the thesis. This is a distributed algorithm that involves randomness to prevent the devices to transmit simultaneously and their signals to interfere with each other. In Section 1.1.2 we show how these random-access models can be viewed as interacting particle systems on graphs. The interference between signals in the network is captured by a hard-core interaction model on an appropriate interference graph. In Section 1.1.3 we introduce the problem of metastability for general systems and, in particular, for wireless networks. We describe how these hard-core interaction models exhibit metastability: when the activation rates become large, the system tends to stabilize in configurations with the maximum number of active nodes, with extremely slow transitions between them. In Section 1.1.4 we focus on random-access models where the activation rates depend on the queues at the nodes. Not much is known about these queue-dependent models (internal models), since most of the literature deals with models with fixed activation rates (external models). The joint activity state together with the joint queue length process evolve as a time-homogeneous Markov process, whose stationary distribution is challenging to analyze. In Section 1.1.5 we introduce the mathematical model, we define the notions of state of a node, queue length at a node and transition time, and we state the basic assumptions on the activation rates. In Section 1.1.6 we give an outline of Chapters 2–4: in Chapter 2 we study complete bipartite networks; in Chapter 3 we generalize to arbitrary bipartite networks; in Chapter 4 we consider dynamic bipartite networks in which the interference graph changes over time.

§1.1.1 Wireless networks and random-access protocols

Wireless communication plays a significant role in our everyday life and has become an integral part of most of our online activities. It consists of the transmission of data or information, without any conductor, from one device (transmitter) to another (receiver) through radio frequency and radio signals. The data packets are transmitted across the devices, over a few meters to hundreds of kilometers. Depending on the distance of communication, the range of data and the type of devices involved, we distinguish between different types of wireless communication technologies: radio and television broadcasting, satellite communication, cellular communication, global positioning system, Wi-Fi, bluetooth, radio frequency identification.

Since wireless signals typically propagate in all directions, they are often overheard by non-intended receivers, and data packets may not be processed correctly if there

are many conflicting signals on the same frequency channel. We say that a *collision* occurs if nearby ongoing conflicting transmissions interfere with each other. In order to reduce collisions and improve the performance, the network requires a medium access control mechanism. Many such mechanisms have been proposed and analyzed, aiming to detect collisions when they occur or to avoid them before they occur. There are two main classes of collision avoidance medium access algorithms, consisting of centralized algorithms and distributed algorithms.

- **Centralized algorithms.** A global control entity has perfect information of all the interference constraints and coordinates all the transmissions by prescribing a certain scheduling to the devices in the network. In short, all the devices connect to a central server, which is the acting agent for all the communications.

- **Distributed algorithms.** The devices decide autonomously when to start a transmission using only local information. Most of these algorithms involve randomness to avoid simultaneous transmissions and share the medium in the most efficient way. Thanks to their low implementation complexity, randomized algorithms have become a popular mechanism for distributed-medium access control. They are also called *random-access algorithms*.

The main idea behind random-access algorithms is to associate with each device a random clock, independently of all the other devices. The clock determines when the device attempts to access the medium in order to transmit. These algorithms can be described in a simple way and only require local information. However, their macroscopic behavior in large networks tends to be very complex: the network performance critically depends on the global spatial characteristics and the geometry of the network (see [2], [3]). Indeed, nearby devices are typically prevented from simultaneous transmission in order to prevent them to interfere and to disturb each other's signals.

One of the first random-access protocols to study wireless networks was developed in the 1970's and is called ALOHA (see [1], [90]). It requires that every device remains inactive for a random amount of time after every attempt of transmission, so that devices do not start transmitting at the same time. This back-off mechanism was developed to avoid simultaneous activity of nearby devices and to reduce the chances of collisions. The *Carrier-Sense Multiple-Access (CSMA) algorithm* is a collision avoidance protocol that refines the ALOHA protocol by combining the random back-off mechanism with interference sensing (see [71]). It is a carrier-sense (CS) protocol, since the devices first sense the shared medium and only start a packet transmission if no ongoing transmission activity from interfering devices is detected. They attempt to transmit after a random back-off time, but if they sense activity of interfering devices, then they freeze their back-off timer until the transmission medium is sensed idle again. It is a multiple-access (MA) protocol, since several devices can transmit by accessing the same medium alternately. The CSMA protocol provides *collision avoidance*, since it tries to ensure that devices do not start a transmission at the same time in order to prevent collisions. CSMA algorithms are popular in distributed random-access networks and various versions are currently implemented in IEEE 802.11 Wi-Fi networks.

In this thesis we consider different stochastic models for CSMA algorithms in order to investigate the effects of different network geometries and how the spatial configuration of the transmitter-receiver pairs affects the network performance.

§1.1.2 Networks as interacting particle systems

Random-access networks with CSMA protocols can be modeled as *interacting particle systems* on graphs with hard-core interaction (see [104]). The undirected graph, which we refer to as the *interference graph*, describes the conflicting transmissions of the devices due to interference. Each transmitter-receiver pair is represented by a particle, which is active when data packets are being transmitted and inactive otherwise. The interference graph encodes the spatial characteristics and the structure of the network, since neighboring particles are not allowed to be active simultaneously. Each particle is equipped with a random clock and the clocks are all independent of each other. When one of these random clocks ticks, the corresponding particle changes its state. If the particle is active, then it deactivates, while if the particle is inactive, then it activates only if all its neighboring particles are inactive. Data packets arrive to each particle independently and form a queue in the buffer while waiting to be transmitted.

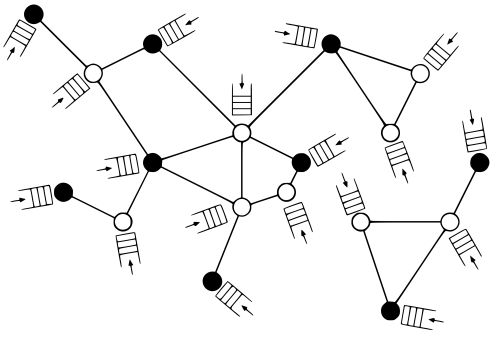


Figure 1.1: The interference graph of a random-access network. Each node represents a communication link between a transmitter and a receiver (active nodes are in black, while inactive nodes are in white). Packets arrive to the transmitter and form a queue. The wireless signals interfere with each other when within a certain interference radius.

Consider the interference graph $G = (N, E)$, where the set of nodes N labels the transmitter-receiver pairs and the set of edges E indicates which nodes interfere. We denote by $X(t) = (X_1(t), \dots, X_N(t)) \in \mathcal{X}$ the joint activity state at time t, with state space

$$\mathcal{X} = \big\{x \in \{0,1\}^N \colon\ x_i x_j = 0\ \ \forall\, (i,j) \in E\big\}, \tag{1.1}$$

where $x_i = 0$ means that node i is inactive and $x_i = 1$ that it is active.

We assume that nodes activate and deactivate according to i.i.d. Poisson clocks. Hence $(X(t))_{t\geq 0}$ evolves as a Markov process with state space $\mathcal{X}$. We consider scenarios in which nodes deactivate at unit rate and become more aggressive over time

when trying to activate, with higher clock rates for activation compared to deactivation. This is relevant for networks in high-load regimes, where the system tends to stabilize in configurations with a maximum number of active nodes, to which we refer as *dominant states*. In a high-load regime those states become extremely rigid, in the sense that we expect extremely slow transitions between them, causing starvation for the currently inactive nodes.

The above-described model has been thoroughly studied in the case where the activation rate at each node i is fixed at some value r_i, $i = 1, \ldots, N$. In that case the joint activity process $(X(t))_{t\geq 0}$ behaves as a reversible time-homogeneous Markov process for any interference graph G, and has a product-form stationary distribution

$$\lim_{t\to\infty} \mathbb{P}\{X(t) = x\} = Z_{\mathcal{X}}^{-1}(r_1, \ldots, r_N) \prod_{i=1}^{N} r_i^{x_i}, \qquad x \in \mathcal{X}, \tag{1.2}$$

with $Z_{\mathcal{X}}(r_1, \ldots, r_N)$ denoting the normalization constant. This model was introduced in the 1980's to analyze the throughput performance of distributed resource sharing and random-access schemes in packet radio networks, in particular, the CSMA protocol (see [9], [10], [68], [69], [87], [103]). The model was rediscovered and further examined twenty years later in the context of IEEE 802.11 WiFi networks (see [46], [48], [75], [102]). If we restrict to $r_i \equiv r$ for all $i = 1, \ldots, N$, then the product-form distribution in (1.2) simplifies to

$$\lim_{t\to\infty} \mathbb{P}\{X(t) = x\} = Z_{\mathcal{X}}^{-1}(r)\, r^{\sum_{i=1}^{N} x_i}, \qquad x \in \mathcal{X}, \tag{1.3}$$

with $Z_{\mathcal{X}}(r) \equiv Z_{\mathcal{X}}(r, \ldots, r)$. From an interacting-particle-systems perspective, the distribution in (1.3) may be recognized as the Gibbs measure of a hard-core interaction model induced by the graph G.

§1.1.3 Metastability in wireless networks

Metastability is a phenomenon where a physical, chemical or biological system, under the influence of a noisy dynamics, moves between different regions of its state space on different time scales (see [22]).

A metastable state is a quasi-equilibrium that persists on a short time scale, but relaxes to an equilibrium on a long time scale, called a stable state. A metastable state represents a configuration where the energy of the system has a local minimum. A stable state represents a configuration where the energy of the system has a global minimum. When the system is subjected to a small noise, we may ask how the transition time depends on the depths of the energy valley around the metastable state and the shape of the bottleneck separating the metastable state from the stable state.

In the past decades there has been intensive study of metastability for interacting particle systems on lattices (see [4], [5], [24], [26], [27], [32], [35], [36], [53], [58], [60], [61], [62], [63], [72], [73], [81], [82], [84]). Different approaches have been proposed, including the path-wise approach (see [33], [34], [49], [50], [77], [83], [86]) and the

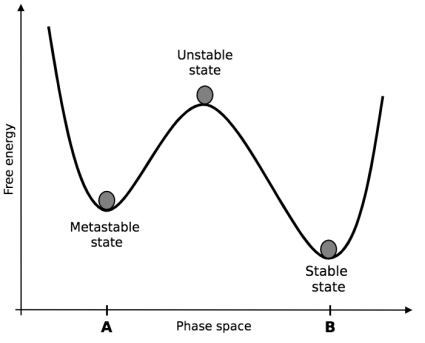


Figure 1.2: The evolution of the system from a metastable state A (local minimum), through an unstable state, to a stable state B (global minimum).

potential-theoretic approach (see [6], [19], [20], [21], [23], [37]). Various asymptotic regimes have been investigated, including metastability at infinite volume and very low temperature (see [25], [30], [31], [38], [39], [70], [78], [79]), and with vanishig magnetic field (see [52], [91]). Recently, attention has turned to the study of metastability for interacting particle systems on graphs, where the lack of periodicity makes the analysis much more difficult (see [43], [44], [57], [66]). The cases where the graph is random are particularly challenging, since the metastable crossover depends on the specific realization of the graph.

Hard-core interaction models are known to exhibit metastability effects. For certain graphs it takes an exceedingly long time for the activity process to reach a stable state when starting from a metastable state. In particular, in a regime where the activation rates become large, the stationary distribution of the joint activity process concentrates on states where the maximum number of nodes is active, with extremely slow transitions between them.

Slow transitions between activity states are not immediately apparent from the stationary distribution. In fact, even when in stationarity each node is active during the same fraction of time, it may well be that over finite time intervals certain nodes are basically barred from activity, while other nodes are transmitting essentially all the time. In other words, the stationary distribution is not directly informative of the transition times between activity states that govern the performance in terms of equitable transmission opportunities for the various users during finite time windows. While the aggregate throughput may improve as the activation rates become large, individual nodes may experience prolonged periods of starvation, possibly interspersed with long sequences of transmissions in rapid succession, resulting in severe build-up of queues and long delays. Indeed, the latter issues have been empirically observed in IEEE 802.11 Wi-Fi networks, and have also been investigated through the lens of the above-described model (see [18], [45], [47], [51], [97], [104]).

Metastability properties are not only of conceptual interest, they are also of great

practical significance. They provide a powerful mathematical paradigm to analyze the likelihood for such unfairness and starvation issues to persist over long time periods. Gaining a deeper understanding of metastability properties and slow transitions is thus instrumental in analyzing starvation behavior in wireless networks, and ultimately of vital importance for designing mechanisms to counter these effects and improve the overall performance as experienced by users.

In this thesis, we study the metastable behavior of a stochastic system of particles with hard-core interactions in a high-density regime. We exploit the particle description of the network to investigate the long transition times between dominant activity states, as well as the temporal starvation and delay issues that these cause.

We consider extreme network topologies as prototypical scenarios, namely, bipartite graphs. This assumption provides mathematical tractability and serves as a stepping stone towards more general network topologies. Consider a bipartite interference graph $G = (N, E) = (U \sqcup V, E)$, where the node set N can be partitioned into two nonempty sets U and V, while E represent the edges describing the interference: two nodes interfere only if one belongs to U and the other belongs to V. In the literature the interference graph is almost always assumed to be static. Our work in this thesis is among the first to explore also the extension to dynamic settings, in which the edges appear and disappear over time.

We denote by $u \in \mathcal{X}$ and $v \in \mathcal{X}$ the joint activity states where all the nodes in either U or V are active, respectively. The activation rates are assumed to be much larger than the deactivation rates, and we also assume a slight imbalance between the activation rates in the two parts of the graph. Starting from state u in which the weak part is all active (metastable state), the system takes a long time before it reaches state v in which the strong part is all active (stable state). This transition represents a global switch in the network. We are interested in studying the distribution of the time until state v is reached when the system starts from state u at time $t = 0$, i.e.,

$$\mathcal{T}_G = \inf\big\{t \geq 0 \colon\ X(t) = v,\, X(0) = u\big\}. \tag{1.4}$$

We refer to the time it takes to go from u to v as *transition time*.

§1.1.4 Queue-based activation protocols

In order to avoid slowdown via metastability, which hinders the performance of the network, protocols are designed for activation and deactivation of the nodes that take into account the current load of the nodes. For instance, to avoid queue overflow, an inactive node with a long queue may want to attempt activation more vigorously, whereas an active node with an empty queue, or with overcrowded neighbors, may want to deactivate and hand over the transmission medium to its neighbors. We assume that packets arrive at the nodes as independent Poisson processes and have independent exponentially distributed sizes. When a packet arrives at a node, it joins the queue at that node and the queue length undergoes an instantaneous jump equal to the size of the arriving packet. The queue decreases at a constant rate (as long as it is positive) when the node is active. Large deviation techniques have been developed for various models in order to control the queue behavior over time (see [76], [95]).

In this thesis we focus on *queue-based* random-access protocols where the activation rates depend on the current queues at the nodes. Specifically, the activation rate is an increasing function of the queue length of the node itself, and possibly a decreasing function of the queue lengths of its neighbors, so as to provide greater transmission opportunities to nodes with longer queues. As a result, these rates vary over time as queues build up or drain when packets arrive or are transmitted. We refer to these models as *internal models*, in the sense that the activation rates are functions of the queue lengths at the various nodes. We denote by $Q_i(t)$ the queue length at node i at time t and by $Q(t) = (Q_1(t), \ldots, Q_N(t)) \in \mathbb{R}^N_{\geq 0}$ the joint queue length vector at time t. In internal models the activation rates are of the form

$$r_i(t) = \begin{cases} g_U(Q_i(t)), & i \in U, \\ g_V(Q_i(t)), & i \in V, \end{cases} \tag{1.5}$$

where the functions $q \mapsto g_U(q)$ and $q \mapsto g_V(q)$ are non-decreasing and such that $\lim_{q\to\infty} g_U(q) = \infty$, $\lim_{q\to\infty} g_V(q) = \infty$ and $g_U(q) = g_V(q) = 0$ when $q < 0$.

Note that $(X(t), Q(t))_{t\geq 0}$ evolves as a time-homogeneous Markov process with state space $\mathcal{X} \times \mathbb{R}^N_{\geq 0}$, since the transition rates depend on time only via the current state of the vector. The queue state not only depends on the history of the packet arrival process (which causes upward jumps in the queue lengths), but also on the past evolution of the activity process itself (through the gradual reduction of the queue lengths during activity periods). The state-dependent nature of the activation rates raises interesting and challenging problems from a methodological perspective, whose solution requires novel concepts in order to handle the two-way interaction between activity states and queue states. The stationary distribution of the Markov process in general does not admit a closed form expression and even the basic throughput characteristics and stability conditions are not known. Indeed, it is not simple to describe explicitly the stability condition for general network topologies (see [100]) and only structural representations or asymptotic results are known (see [29], [74]). In order to study activation rates based on queue lengths, powerful algorithms have been proposed and it has been shown that these achieve throughput optimality under mild assumptions (see [64], [65], [88], [94]).

Most of the literature refers to models where the activation rates are fixed parameters and the underlying Markov process is time-homogeneous (see [59], [83], [105], [106]). In this setting it has been shown that the transition time from the metastable to the stable state is approximately exponential on the scale of its mean. The main idea is to consider the return times to the metastable state of the discrete-time embedded Markov chain as regeneration times. At each regeneration time a Bernoulli trial is conducted. The trial is successful if the stable state is reached before a return to the metastable state occurs, while it is unsuccessful otherwise. In the asymptotic regime, the success probability of each trial is small and the expected length of a single trial is negligible compared to the expected transition time. It is known that the first success time of a large number of trials, each having a small probability of success, is approximately exponentially distributed (see [49], [67]).

Attention has also been paid to models where the activation rates change with time. The underlying Markov process is therefore time-inhomogeneous, and it has

been shown that, under appropriate conditions, the transition time is approximately exponential with a non-constant rate (see [14]). The above-described approach is still fruitful, but the success probability and the length of each trial now depend on its starting time. We call these models *external models*, in the sense that the activation rates are deterministic functions of time. In external models, the activation rates are of the form

$$r_i(t) = \begin{cases} h_U(t), & i \in U, \\ h_V(t), & i \in V, \end{cases} \tag{1.6}$$

for suitable functions $t \mapsto h_U(t)$ and $t \mapsto h_V(t)$. It is interesting to note that the metastable behavior of exit times of simulated annealing in a time-inhomogeneous setting has some similarities with the one of external models: a trichotomy was observed, similar to the one we will discuss in this thesis (see [80]).

Recently, various models for random-access networks with queue-based protocols have been investigated (see [17], [28], [41], [42], [85], [93]). Breakthrough work has shown that, for suitable activation rate functions, these protocols achieve maximum stability, i.e., provide stable queues whenever feasible at all (see [55], [64], [88], [94]). Thus, these policies are capable of matching the optimal throughput performance of centralized scheduling strategies, while requiring less computation and operating in a distributed fashion. On the downside, the very activation rate functions required for ensuring maximum stability tend to result in long queues and poor delay performance (see [18], [54]). This has sparked an interest in understanding, and possibly improving, the delay performance of queue-based random-access schemes. Analyzing metastability properties for the joint activity process is a crucial endeavor in this regard.

In this thesis we specifically examine the transition time of the joint activity process in an asymptotic regime where the initial queue lengths $Q_i(0)$, and hence the activation rates $r_i(Q(t))$, $i = 1, \ldots, N$, become large in a suitable sense.

§1.1.5 Mathematical model

Consider the bipartite graph $G = (U \sqcup V, E)$ and recall that a node in the network can be either active or inactive.

Definition 1.1.1 (State of a node).
The *state of node* i at time t is described by a Bernoulli random variable $X_i(t) \in \{0, 1\}$, defined as

$$X_i(t) = \begin{cases} 0, & \text{if } i \text{ is inactive at time } t, \\ 1, & \text{if } i \text{ is active at time } t. \end{cases} \tag{1.7}$$

The joint activity state $X(t)$ at time t is an element of the set $\mathcal{X}$: the feasible configurations of the network correspond to the collection of independent sets of G. Recall that $u \in \mathcal{X}$ ($v \in \mathcal{X}$) represents the joint activity state where all the nodes in U are active (inactive) and all the nodes in V are inactive (active). The main object of interest in this thesis is the transition time between u and v. We write $\mathbb{P}_u$ and $\mathbb{E}_u$ to denote probability and expectation on path space given that the initial joint activity state is u.

Definition 1.1.2 (Transition time).
The *transition time* $\mathcal{T}_G^Q$ of the graph G given initial queue lengths Q is defined as

$$\mathcal{T}_G^Q = \inf\big\{t \geq 0\colon\ X(t) = v,\, X(0) = u\big\}. \tag{1.8}$$

In other words, $\mathcal{T}_G^Q$ is the time it takes to reach v starting from u. We sometimes write $\mathcal{T}_G$ and omit the dependence on Q when this dependence is clear from the context.

An active node i deactivates according to a deactivation Poisson clock with rate 1: when the clock ticks the node deactivates. Vice versa, an inactive node i attempts to activate at the ticks of an activation Poisson clock with rate $r_i(t)$: an attempt at time t is successful when no neighbors of i are active at time $t-$. The activation rate of i depends on its current queue length $Q_i(t)$ and satisfies (1.5).

Definition 1.1.3 (Queue length at a node).
Let $t \mapsto Q_i^+(t)$ be the input process describing packets arriving according to a Poisson process $t \mapsto N(t)$ with rate λt and having i.i.d. exponential service times of parameter μ, $Y_j \simeq \mathrm{Exp}(\mu)$, $j \in \mathbb{N}$. Let $t \mapsto Q_i^-(t)$ be the output process representing the cumulative amount of work that is processed in the time interval $[0,t]$ at rate c, i.e., $cT_i(t) = c\int_0^t X_i(s)ds$. Define

$$\Delta_i(t) = Q_i^+(t) - Q_i^-(t) = \sum_{j=0}^{N_i(t)} Y_{ij} - cT_i(t), \tag{1.9}$$

and let $s^* = s^*(t)$ be the value where $\sup_{s\in[0,t]}[\Delta_i(t) - \Delta_i(s)]$ is reached, namely, $[\Delta_i(t) - \Delta_i(s^*-)]$. Let $Q_i(t) \in \mathbb{R}_{\geq 0}$ denote the *queue length* at node i at time t. Then

$$Q_i(t) = \max\big\{Q_i(0) + \Delta_i(t),\, \Delta_i(t) - \Delta_i(s^*-)\big\}, \tag{1.10}$$

where $Q_i(0)$ is the initial queue length. The maximum is achieved by the first term when $Q_i(0) \geq -\Delta_i(s^*-)$ (the queue length never sojourns at 0), and by the second term when $Q_i(0) < -\Delta_i(s^*-)$ (the queue length sojourns at 0 at time s^*-). In order to ensure that the queue length remains non-negative, a node deactivates when its queue length hits zero.

The *initial queue lengths* are assumed to be

$$Q_i(0) = \begin{cases} \gamma_U r, & i \in U, \\ \gamma_V r, & i \in V, \end{cases} \tag{1.11}$$

where $\gamma_U \geq \gamma_V > 0$, and r is a parameter that tends to infinity. Thus, the initial queue lengths are of order r, i.e., $Q_i(0) \asymp r$, and the ones at the nodes in U are larger than the ones at the nodes in V. Note that the transition time tends to infinity with r, since the larger the initial queue lengths are, the longer it takes for the transition to occur. We study different models in the limit as the queue lengths become large, and so we are interested in asymptotic results for the transition time as $r \to \infty$.

For each node i, the *input process* $t \mapsto Q_i^+(t) = \sum_{j=0}^{N_i(t)} Y_{ij}$ is a compound Poisson process. In the time interval $[0,t]$ packets arrive at node i according to a Poisson

process $t \mapsto N_i(t)$ with rate λ_U or λ_V, depending on whether the node is in U or V. Moreover, each packet j brings the information of its service time: the service time Y_{ij} of the j-th packet at node i is exponentially distributed with parameter μ. Hence the expected value of $Q_i^+(t)$ for a node in U is the product of the expected value $\mathbb{E}[N_i(t)] = \lambda_U t$ and the expected value $\mathbb{E}[Y_j] = 1/\mu$, i.e., $\mathbb{E}[Q_i^+(t)] = (\lambda_U/\mu)t = \rho_U t$. Analogously, for a node in V we have $\mathbb{E}[Q_i^+(t)] = \rho_V t$. The quantities ρ_U and ρ_V denote the common traffic intensity of the nodes in U and V, respectively. We assume that all the service times are i.i.d. random variables, and are independent of the Poisson process $t \mapsto N_i(t)$.

For each node i, the *output process* is $t \mapsto Q_i^-(t) = cT_i(t) = c\int_0^t X_i(u)\,du$, where the activity process $t \mapsto T_i(t)$ represents the cumulative amount of active time of node i in the time interval $[0,t]$. This is not independent of the input process. Intuitively, the average queue length increases when the node is inactive and decreases when the node is active, which means that packets are being served at a rate c larger than their arrival rate, i.e., $c > \rho_U, \rho_V > 0$. Since all nodes in V are initially inactive, for some time the queue length of these nodes in V is not affected by their output process. However, as soon as a node in V activates, we have to consider its output process as well.

The choice of functions g_U, g_V in (1.5) determines the transition time of the network, since the activation rates of the nodes depend on them.

Definition 1.1.4 (Assumptions on the activation rates).
We assume that the activation functions g_U, g_V fall in the class

$$\mathcal{G} = \Big\{ g\colon\ \mathbb{R} \to \mathbb{R}_{\geq 0}\colon\ g \text{ non-decreasing and continuous,} \\ g(x) = 0 \text{ for } x \in \mathbb{R}_{\leq 0},\ \lim_{x\to\infty} g(x) = \infty \Big\}. \tag{1.12}$$

Moreover, we assume nodes in V to be more aggressive than nodes in U, i.e.,

$$\lim_{x\to\infty} \frac{g_V(x)}{g_U(x)} = \infty, \tag{1.13}$$

so that the transition from u to v can be viewed as the crossover from a metastable state to a stable state.

§1.1.6 Outline of Part I: Chapters 2–4

The three chapters of Part I of this thesis are based on three papers on queue-based random-access protocols for wireless networks.

In Chapter 2 we focus on *complete bipartite interference graphs*, which are useful for modeling dense networks and which provide a worst-case perspective. While there is admittedly no specific physical reason for focusing on complete bipartite graphs, this assumption provides mathematical tractability and serves as a stepping stone towards more general network topologies. The main goal is to compare the transition time of the internal model with that of the external model in which the activation rates depend on the current mean queue length. We define two perturbed models with

externally driven activation rates that sandwich the queue lengths of the internal model and its transition time. We show with the help of coupling that with high probability the mean transition time and its distribution for the internal model are asymptotically the same as for the external model. The chapter is based on [12].

In Chapter 3 we turn our attention to *arbitrary bipartite interference graphs*, for which not necessarily all nodes in U interfere with all nodes in V. In this setting the problem turns out to be considerably more challenging. In order to achieve the full transition, the network goes through a succession of subtransitions, in which a certain succession of complete bipartite subgraphs achieve a metastable crossover and, in doing so, effectively remove themselves from the network. This succession depends in a delicate manner on the full structure of the graph. We formulate a greedy algorithm to analyze the most likely transition paths between dominant states. By combining the results for complete bipartite graphs with a detailed analysis of the algorithm, we are able to determine the mean transition time and its distribution along each transition path. The chapter is based on [13].

In Chapter 4 we study a dynamic version of the random-access protocols to model wireless networks with user mobility. With an explorative intention, we analyze *dynamic bipartite interference graphs* where the interference between nodes changes over time: Poisson i.i.d. clocks are attached to the edges, which can appear and disappear from the graph when their clock ticks. Our approach is based on the intuition that a node in V can activate either when its neighbors are simultaneously inactive or when the edges connecting it with its neighbors disappear. Interpolation between these two situations gives rise to different scenarios and interesting behavior. We identify how the mean transition time depends on the speed of the dynamics. The chapter is based on [92].

§1.2 Introduction to Part II

In the second part of this thesis we study spectral properties of random graphs, in particular, of inhomogeneous Erdős-Rényi random graphs. Random graphs have many applications in the modeling of complex physical, biological and social networks. In order to understand the structure of these networks, we consider the adjacency and Laplacian matrices associated to the underlying graphs and study their eigenvalues.

In Section 1.2.1 we give a brief overview of random matrices and motivate our interest in analyzing their spectra. We introduce Wigner matrices, the Wigner semicircle law and the universality principle. In Section 1.2.2 we define the adjacency and Laplacian matrices of a graph together with their empirical spectral distribution. For random matrices, the eigenvalues are random variables and the empirical spectral distribution is a random probability measure. In Section 1.2.3 we consider the standard Erdős-Rényi random graph and discuss the main regimes of behavior depending on the connection probabilities. We next consider the inhomogeneous Erdős-Rényi random graph, for which we investigate both the limiting spectral distribution and the large-deviation behavior of the largest eigenvalue. In order to introduce these two problems, we discuss known results for standard and inhomogeneous Erdős-Rényi random graphs. In Section 1.2.4 we give a brief introduction to free probability theory, which can be seen as the analogue of classical probablility for non-commutative random variables. Its connection to random matrix theory allows us to indentify the limiting spectral distribution for certain classes of random matrices. In Section 1.2.5 we give a brief intoduction to graphon theory, used to study limits of dense graph sequences. Graphons also provide crucial tools to study large deviations for dense random graphs. In Section 1.2.6 we give an outline of Chapters 5–6: in Chapter 5 we study the empirical spectral distribution and its limiting behavior for the adjacency and Laplacian matrices in the non-dense non-sparse regime; in Chapter 6 we study large deviations for the largest eigenvalue of the adjacency matrix in the dense regime and analyze its rate function in detail.

§1.2.1 Random matrices

The study of *random matrices*, in particular, the properties of their eigenvalues, emerged from applications. Random matrices appeared for the first time in 1928, when Wishart (see [195]) used them in statistics and data analysis. Later, in the 1950s, the natural question regarding their eigenvalue statistics was raised in the pioneering work of Wigner (see [194]). While studying statistical models for nuclear physics, he noticed from experimental data that gaps in energy levels of large nuclei tend to follow the same statistics independently the material. We now know from quantum mechanics that these energy levels correspond to the eigenvalues of a self-adjoint Hamiltonian operator, but the correct form of this operator was not known at the time. Wigner's idea was to model the complex Hamiltonian by a random matrix with independent entries. He ignored all the physical details of the system except the symmetry: he modeled systems with time reversal symmetry by real symmetric random matrices, and systems without time reversal symmetry by complex

Hermitian random matrices. Surprisingly, this simplification reproduced the correct gap statistics, suggesting the existence of a profound *universality principle*.

Wigner matrices are symmetric Hermitian random matrices whose elements are i.i.d. random variables with mean 0 and variance 1. Wigner showed that the empirical spectral distribution converges almost surely to the *semicircle law* he had initially discovered for random matrices with Gaussian elements (see [194]). The semicirle law, now called Wigner semicircle law, has density

$$\rho_{\mathrm{sc}}(x) = \frac{1}{2\pi}\sqrt{4-x^2}, \qquad x \in [-2,2]. \tag{1.14}$$

The i.i.d. requirement and the constant variance condition are not essential for proving the semicircle law. Indeed, also generalized Wigner matrices, where the variances of the elements can be different and each column of the variance profile is stochastic, obey the semicircle law under various conditions (see [109], [157], [163]).

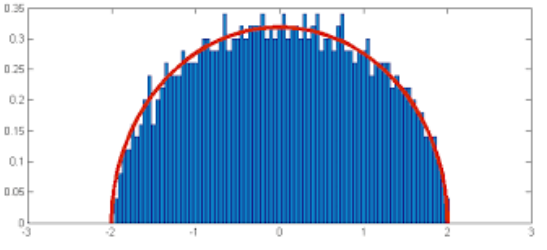


Figure 1.3: Wigner semicircle law.

The Wigner-Dyson-Gaudin-Mehta conjecture states that the local spectral statistics of Wigner matrices exhibit universality: they only depend on the symmetry class of the ensemble and not on the distribution of the matrix elements. In particular, the local spectral statistics are the same as the ones of matrices with Gaussians elements, for which there are explicit formulas. In the meantime this conjecture has been solved for all symmetry classes (see [127], [153], [154], [155], [156], [157], [190]). The universality phenomenon has been recently established also for other models, such as generalized Wigner matrices (see [157]), Wigner-type matrices ([107]), and adjacency matrices of Erdős-Rényi random graphs (see [151], [152], [168], [191]).

Motivated by physical applications, in the 1960s a mathematical theory of the spectrum of random matrices was developed and links with various branches of mathematics, including classical analysis and number theory, were established (see [148], [149], [162], [181]). Over the years, it has become clear that models related to random matrices play an important role in several areas of mathematics. Nowadays, random matrix theory is a central topic in probability and statistical physics, with many connections to combinatorics, numerical analysis, statistics and computer science.

In this thesis we study properties of the eigenvalues of random matrices arising from random graphs. Since a random matrix has random entries, its eigenvalues are random variables. We aim at understanding of the distribution of the eigenvalues

from knowledge of the distribution of the entries. Random numbers and random vectors are known to exhibit universal patterns, such as the law of large numbers and the central limit theorem. It is of great interest to understand their analogues in the non-commutative setting and to identify the behavior of eigenvalues of large random matrices.

§1.2.2 The adjacency and Laplacian matrices

We begin with some basic definitions. Consider a finite simple graph $G = (V, E)$ on N vertices. The *adjacency matrix* $A_N = A(G)$ associated to G is defined as the $\{0, 1\}$-valued $N \times N$ matrix whose elements indicate whether a given pair of vertices is adjacent or not in the graph, i.e., is connected by an edge:

$$A_N(i,j) = \begin{cases} 1, & i \sim j\,, \\ 0, & \text{else.} \end{cases} \tag{1.15}$$

The diagonal elements of the matrix are all zero, since edges from a vertex to itself (loops) are not allowed in simple graphs. Note that A_N is symmetric. Hence it has N real eigenvalues, which can be ordered as

$$\lambda_1(A_N) \geq \ldots \geq \lambda_N(A_N). \tag{1.16}$$

The eigenvalues of the adjacency matrix have various applications in graph theory: they carry information about topological features of the graph, such as connectivity and subgraph counts (see [139], [142]).

The *Laplacian matrix* $\Delta_N = \Delta_N(G)$ associated to G is the $N \times N$ matrix defined as

$$\Delta_N(i,j) = \begin{cases} -\sum_{k=1}^{N} A_N(i,k), & i = j\,, \\ A_N(i,j), & i \neq j\,. \end{cases} \tag{1.17}$$

Note that also Δ_N is symmetric. Hence it also has N real eigenvalues, which can be ordered as

$$\lambda_1(\Delta_N) \geq \ldots \geq \lambda_N(\Delta_N). \tag{1.18}$$

The eigenvalues of the Laplacian matrix carry information about random walks on the graph and allow us to analyze approximation algorithms (see [139]).

For $x \in \mathbb{R}$, let δ_x denote the Dirac measure at x. The *empirical spectral distribution (ESD)* of an $N \times N$ symmetric matrix M is the probability distribution that puts mass $1/N$ at each of the N eigenvalues of M, i.e.,

$$\mathrm{ESD}(M) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i(M)}. \tag{1.19}$$

Hence, the empirical spectral distributions of the adjacency matrix A_N and the Laplacian matrix Δ_N associated to the graph G are defined as

$$\mathrm{ESD}(A_N) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i(A_N)} \qquad \mathrm{ESD}(\Delta_N) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i(\Delta_N)}. \tag{1.20}$$

The empirical spectral distribution is a graph invariant and encodes important information about G. It is therefore one of the main objects of interest in spectral graph theory (see [139]). From perturbation theory for matrices, it is known that the eigenvalues are continuous functions of the elements of the matrix. The empirical spectral distributions of the adjacency and Laplacian matrices are both random probability distributions on $\mathbb{R}$.

§1.2.3 Spectra of Erdős-Rényi random graphs

Spectral graph theory studies the properties of eigenvalues and eigenvectors of the associated adjacency and Laplacian matrices. In the past 20 years many results have been derived about spectra of random matrices associated with random graphs (see [115], [119], [122], [144], [146], [150], [159], [170], [171], [172], [175], [191], [196]).

The standard *Erdős-Rényi random graph* model $G(N,p)$, first introduced by Erdős and Rényi (see [158]), is the most basic random graph model. It consists in a graph on N vertices formed by connecting each pair of vertices i and j with probability $p = p(N)$, independently of each other. Note that, up to symmetry, the adjacency matrix of $G(N,p)$ consists of i.i.d. Bernoulli random variables. Each element of the matrix is 1 with probability p and 0 with probability $1-p$, independently of the other elements.

Depending on the asymptotic behavior of the connection probability $p(N)$ as $N \to \infty$, we distinguish between the following regimes.

(I) **Dense regime**, $p(N) \equiv p \in (0,1)$. The average degree diverges linearly.

(II) **Non-dense non-sparse regime**, $p(N) \to 0$ and $Np(N) \to \infty$. The average degree diverges slower than linearly.

(III) **Sparse regime**, $p(N) \to 0$ and $Np(N) \to a \in (0,1)$. The degree distribution is asymptotically Poisson with parameter a.

(IV) **Sub-sparse regime**, $p(N) \to 0$, $Np(N) \to 0$ and $N^2p(N) \to \infty$. Most vertices have degree 0, but the total number of edges diverges.

(V) **Ultra-sparse regime**, $p(N) \to 0$ and $N^2p(N) \to b \in (0,1)$. The total number of edges is asymptotically Poisson with parameter $\frac{1}{2}b$.

The regimes (I) and (II) are often denoted in the literature as dilute regime or non-sparse regime. Note that the regime where $N^2p(N) \to 0$ is not interesting because all the edges will be missing with high probability.

In this thesis we focus on a generalization of standard Erdős-Rényi random graphs. Namely, we consider *inhomogeneous Erdős-Rényi random graphs*, where each pair of vertices i and j is connected with probability $p_{ij} = p_{ij}(N)$, independently of each other. Many popular graph models arise as special cases of inhomogeneous Erdős-Rényi random graphs, such as random graphs with given expected degrees (see [141]), stochastic block models (see [166]) and W-random graphs (see [123], [177]).

We address two different problems. We first study the limiting spectral distributions of the adjacency and Laplacian matrices in the non-dense non-sparse regime (II). We next study the large deviation principle for the largest eigenvalue of the adjacency matrix in the dense regime (I).

Limiting spectral distribution

One of the challenges in the study of spectra of random graphs is to investigate the convergence of the empirical spectral distributions to a *limiting spectral distribution* for the adjacency and Laplacian matrices as the size of the graph becomes large.

Various results have been proved for standard Erdős-Rényi random graphs. In the non-sparse regime, the adjacency matrix falls into the Wigner class and its empirical spectral distribution converges (after appropriate scaling and centering) to a semicircle law (see [128], [146], [191]). In the sparse regime, the adjacency matrix can be viewed as a singular Wigner ensemble, since the distribution of its elements is highly concentrated around 0. Its analysis is more challenging than in the non-sparse regime. The empirical spectral distribution of the Laplacian matrix converges (again after appropriate scaling) to a free additive convolution of a Gaussian and a semicircle law (see [128], [146], [170]). Both spectra are well understood.

Our goal is to extend these results to inhomogeneous Erdős-Rényi random graphs. Recently, some properties of the empirical spectral distribution of adjacency matrices have been derived via the theory of graphons (see [196]).

Largest eigenvalue

Another interesting challenge in the study of spectra of random graphs is to analyze the behavior of the *largest eigenvalue* of the adjacency matrix.

For standard Erdős-Rényi random graphs it has been shown that, in the dense regime, the largest eigenvalue asymptotically has a normal distribution (see [160]) and satisfies a weak law of large numbers (see [173]). Moreover, it is asymptotically equivalent to the maximum of the maximal mean degree d and the square root of the largest degree (see [173]). In the sparse regime, the behavior of the largest eigenvalue of inhomogeneous Erdős-Rényi random graphs exhibits a crossover at $d \asymp \log N$ (which corresponds to the crossover from disconnected to connected graphs). When $d \ll \log N$ there is a sharp increase in the density of eigenvalues towards the centre of the spectrum, while when $d \gg \log N$ the extreme eigenvalues converge to the edge of the support of the asymptotic eigenvalue distribution (see [116], [117]).

Large deviations for the largest eigenvalue have also been intensely studied. Large deviation theory for random matrices started with the study of large deviations for the empirical spectral distribution of β-ensembles with a quadratic potential (see [111]). The rate was shown to be the square of the number of vertices, and the rate function was shown to be given by a non-commutative notion of entropy. The largest eigenvalue for such ensembles was also studied (see [110]). More recently, large deviations for the empirical spectral distribution of random matrices with non-Gaussian tails were derived (see [121]), and the largest eigenvalue was studied (see [112], [113]). The

adjacency matrix of an Erdős-Rényi random graph does not fall in this regime, and hence different techniques are needed.

For dense Erdős-Rényi random graphs, the breakthrough work of Chatterjee and Varadan (see [138]) introduced a general framework for large deviation principles via Szemerédi's regularity lemma (see [188]) and the theory of graphons (see [125], [177], [178]). It expresses the structure of the random graph conditional on a large deviation in terms of a variational problem involving graphons. The framework was initially set up for subgraph densities, but the results extend to so-called graph parameters, including the operator norm of graphons, which is the extension of the spectral norm (largest eigenvalue) to the space of graphons. Consequently, the large deviation rate function for the upper and lower tails of the largest eigenvalue, and the behavior of the graph conditional on large deviations, can be described in detail (see [137]). The original question from Chatterjee and Varadan was the following. "Fix $0 < p < r < 1$ and take $G \sim G(N, p)$ conditioned to have at least as many triangles as is typical for $G(N, r)$. Is G close in cut-distance to a typical $G(N, r)$?". The region of (p, r) where the answer is positive is called replica symmetric phase and has recently been identified. Analagous results have been derived also in the setting where the largest eigenvalue of $G \sim G(N, p)$ is conditioned to exceed the typical value of the largest eigenvalue of $G(N, r)$ (see [179]).

Recently, a large deviation principle for uniform dense random graphs with a given degree sequence has been established via the above-described framework (see [145]). Dense inhomogeneous Erdős-Rényi random graphs fall in this class. Hence, a large deviation principle holds and general results on large deviations for the largest eigenvalue follow. Our goal is to study the large deviation principle for the largest eigenvalue and to analyze the associated rate function in detail.

§1.2.4 Free probability

In 1983 Voiculescu introduced *free probability theory* in the context of operator algebras in order to address the isomorphism problem of free group factors (see [192]). The theory reached a new level when he discovered connections to random matrix theory (see [193]). The tools developed in operator algebras and free probability theory can now be applied to many classes of random matrices, in particular, to indentify the limiting spectral distribution (see [182]). Since random matrices are also widely used in applied fields, such as wireless communications or statistics, free probability has become quite common. Moreover, it has close connections to combinatorics, representations of symmetric groups, large deviations and quantum information theory.

Definition 1.2.1 (Non-commutative probability space).
We say that the pair $(\mathcal{A}, \phi)$ is a *non-commutative probability space* if $\mathcal{A}$ is a unital algebra and ϕ is a linear functional $\phi : \mathcal{A} \to \mathbb{C}$ with $\phi(1) = 1$.

Let I be an index set. We call (non-commutative) random variables the elements of $\mathcal{A}$, we call moments of a random variable $a \in \mathcal{A}$ the numbers $\phi(a^n)$, $n \in \mathbb{N}$, and we call joint distribution of the random variables $a_1, \ldots, a_k \in \mathcal{A}$ the collection of all mixed moments $\phi(a_{i(1)} \cdots a_{i(l)})$, where $l \in \mathbb{N}$, $i(1), \ldots, i(l) \in \{1, \ldots, k\}$.

Definition 1.2.2 (Free independence).
For each $i \in I$, let $\mathcal{A}_i \subset \mathcal{A}$ be unital subalgebras of $\mathcal{A}$. The subalgebras $(\mathcal{A}_i)_{i\in I}$ are said to be *free* or *freely independent* if $\phi(a_1 \cdots a_k) = 0$ whenever:

(i) $a_j \in A_{i(j)}, i(j) \in I$ and $\phi(a_j) = 0$, for all $j = 1, \ldots, k$, with $k \in \mathbb{N}$;

(ii) $i(1) \neq i(2), i(2) \neq i(3), \ldots, i(k-1) \neq i(k)$, i.e., neighboring elements belong to different subalgebras.

For $i \in I$, let $x_i \in \mathcal{A}$. The random variables $(x_i)_{i\in I}$ are said to be free or freely independent if their generated unital subalgebras $(\mathcal{A}_i)_{i\in I}$ are free, where $\mathcal{A}_i$ is the unital subalgebra of $\mathcal{A}$ generated by x_i.

Note that freeness between two random variables x and y is a rule for calculating the mixed moments in x and y from the moments of x and the moments of y. Freeness can be seen as a non-commutative analogue of the classical probabilistic concept of independence for random variables, which is why it is called free independence.

§1.2.5 Graphons

The analysis of large networks is one of the main challenges in modern graph theory. It is important to have proper definitions of convergence for graph sequences in order to identify limiting objects. A solution to this problem is provided by *graphon theory*, introduced in 2006 by Lovász and Szegedy, which defines graphons as limits of dense graph sequences (see [177]). Graphons characterize the convergence of graph sequences with the help of graph homomorphism densities (see [125], [126]).

Definition 1.2.3 (Graphon).
A *graphon* is a symmetric Lebesgue-measurable function from the unit square to the unit interval. More precisely, the set of graphons $\mathcal{W}$ is defined as

$$\mathcal{W} = \big\{h\colon\ [0,1]^2 \to [0,1]\colon\ h(x,y) = h(y,x)\ \forall\,(x,y) \in [0,1]^2\big\}. \tag{1.21}$$

On the set of graphons it is possible to define a metric in the following way.

Definition 1.2.4 (Cut-metric).
Let $\mathcal{M}$ be the set of Lebesgue measure-preserving bijective maps $\phi\colon\ [0,1] \mapsto [0,1]$. The *cut-distance* between two graphons $h_1, h_2 \in \mathcal{W}$ is defined by

$$d_\square(h_1,h_2) = \sup_{S,T\subseteq[0,1]} \left| \int_{S\times T} (h_1(x,y) - h_2(x,y))\,dx\,dy \right|, \tag{1.22}$$

where S, T run over all measurable subsets of $[0,1]$. The *cut-metric* $\delta_\square$ is defined by

$$\delta_\square(h_1,h_2) = \inf_{\phi\in\mathcal{M}} d_\square(h_1, h_2^\phi), \tag{1.23}$$

where $h_2^\phi(x,y) = h_2(\phi(x),\phi(y))$.

The cut-metric defines an equivalence relation $\sim$ on the space of graphons $\mathcal{W}$ by declaring $h_1 \sim h_2$ if and only if $\delta_\square(h_1, h_2) = 0$, and leads to the quotient space $\widetilde{\mathcal{W}} = \mathcal{W}/_\sim$. For $h \in \mathcal{W}$ we write $\widetilde{h}$ to denote the equivalence class of h in $\widetilde{\mathcal{W}}$. The pair $(\widetilde{\mathcal{W}}, \delta_\square)$ is a compact metric space (see [176]).

There is a natural way to embed a simple graph in the space of graphons. Consider a graph G on N vertices and contruct the associated graphon h^G in the following way. Divide the unit square $[0,1]^2$ into N^2 equal boxes of equal size and assign to each box the value of the corresponding element of the adjacency matrix. More precisely,

$$h^G(x,y) = \begin{cases} 1, & \text{if there is an edge between vertex } \lceil Nx \rceil \text{ and vertex } \lceil Ny \rceil, \\ 0, & \text{else,} \end{cases} \tag{1.24}$$

where $\lceil x \rceil$ denotes the smallest integer larger than or equal to x.

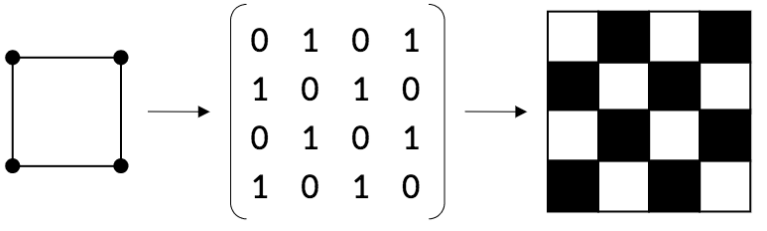


Figure 1.4: Graphon representation of a graph.

Graphon theory is not only connected to graph theory, but also to measure theory, probability theory and functional analysis. Recently, graphon theory has been generalized to include sparse graph sequences (see [123], [124], [161], [174]).

§1.2.6 Outline of Part II: Chapters 5–6

The two chapters of Part II of this thesis are based on two papers on spectral properties of inhomogeneous Erdős-Rényi random graphs.

In Chapter 5 we consider inhomogeneous Erdős-Rényi random graphs in the non-dense non-sparse regime, where the degrees of the vertices diverge sublinearly with the size of the graph. We are interested in the limiting behavior of the *empirical spectral distributions* of the adjacency and Laplacian matrices. We identify their scaling limit. When the connection probabilities have a multiplicative structure, we are able to give an explicit description of the scaling limits using tools from free probability theory. Inhomogeneous Erdős-Rényi random graphs with a multiplicative structure for the connection probabilities arise naturally in different contexts. For instance, they have been shown to play a crucial role in the identification of the limiting spectral distribution of the adjacency matrix of the configuration model (see [144]). The chapter is based on [132].

In Chapter 6 we focus on the behavior of the *largest eigenvalue* of the adjacency matrix in the dense regime, where the degrees of the vertices are proportional to the

size of the graph. Using the framework of Chatterjee and Varadan and the theory of graphons, we prove a large deviation principle for dense inhomogeneous Erdős-Rényi random graphs. We derive a large deviation principle for the largest eigenvalue and analyze the associated rate function in detail. When the connection probabilities have a multiplicative structure, we are able to identify its scaling properties. The chapter is based on [133].

PART I

QUEUE-BASED RANDOM-ACCESS PROTOCOLS FOR WIRELESS NETWORKS

CHAPTER 2

Complete bipartite interference graphs

This chapter is based on:
S.C. Borst, F. den Hollander, F.R. Nardi, M. Sfragara. *Transition time asymptotics of queue-based activation protocols for random-access neworks.* Stochastic Processes and Their Applications, 2020.

Abstract

We consider networks where each node represents a server with a queue. An active node deactivates at unit rate. An inactive node activates at a rate that depends on its queue length, provided none of its neighbors is active. For complete bipartite networks, in the limit as the queues become large, we compute the mean transition time between the two states where one half of the network is active and the other half is inactive. We show that the law of the transition time divided by its mean exhibits a trichotomy, depending on the activation rate functions.

§2.1 Introduction and main results

In Section 2.1.1 we describe the setting and the mathematical model of interest in this chapter. In Section 2.1.2 we state our main results. In Section 2.1.3 we offer a brief discussion of these results and give an outline of the remainder of the chapter.

§2.1.1 Setting

We refer to Section 1.1.5 for a general introduction to the mathematical model. In this section we refine it with some extra notions we will need in the chapter.

Consider a complete bipartite graph G: the node set can be partitioned into two nonempty sets U and V such that the bond set is the product of U and V, i.e., two nodes interfere if and only if one belongs to U and the other belongs to V (see Figure 2.1 for an example). Thus, the collection of all independent sets of G consists of all the subsets of U and all the subsets of V.

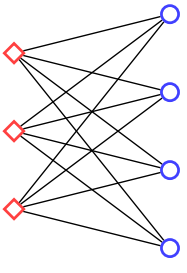


Figure 2.1: A complete bipartite graph with $|U| = 3$ and $|V| = 4$. At time $t = 0$, square-shaped nodes are active and circle-shaped nodes are inactive.

We assume the activation rates to satisfy Definition 1.1.4. Moreover, we focus on the following.

Definition 2.1.1 (Assumption on the activation rates).
We assume *polynomial activation functions* for nodes in U of the form

$$g_U(x) \sim Bx^\beta, \qquad x \to \infty, \tag{2.1}$$

with $B, \beta \in (0, \infty)$. We will discuss more general functions g_U in Remark 2.4.1. We do not require any further assumption on the functions g_V: it turns out that the asymptotic distribution of the transition time is independent of g_V.

Next, we define our two main objects of interest.

Definition 2.1.2 (Pre-transition and transition time).
The *pre-transition time* τ_G is defined as the first time a node in V activates starting from u, i.e.,

$$\tau_G = \inf\{t > 0 \colon\ X_i(t) = 1 \ \exists\, i \in V,\, X(0) = u\}. \tag{2.2}$$

The *transition time* $\mathcal{T}_G$ is defined as the first time v is reached starting from u, i.e.,

$$\mathcal{T}_G = \inf\big\{t \geq 0\colon\ X_i(t) = 0\ \forall\, i \in U,\ X_i(t) = 1\ \forall\, i \in V, X(0) = u\big\}. \tag{2.3}$$

The pre-transition time plays an important role in our analysis of the transition time, because the evolution of the network is simpler on the interval $[0, \tau_G]$ than on the interval $[\tau_G, \mathcal{T}_G]$. However, we will see that $\mathcal{T}_G - \tau_G \ll \tau_G$ when the initial queue lengths are large, so that both times have the same asymptotic scaling behavior. See Figure 2.2 for a representation of the pre-transition state.

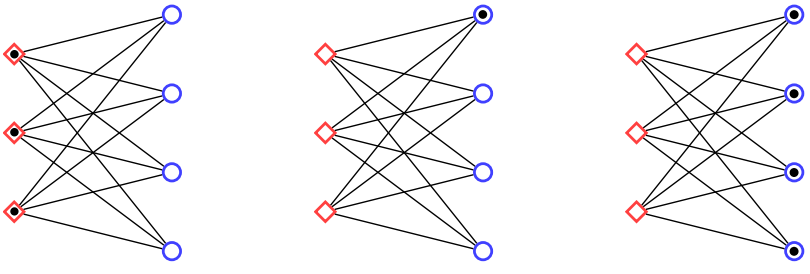


Figure 2.2: Left*: initial state u.* Center*: pre-transition state.* Right*: final state v.*

We study the transition starting from u and we set the initial queue sizes $Q_i(0)$ to be large for all $i \in U \sqcup V$. Hence, initially all the nodes in U are active virtually all the time, preventing any of the nodes in V to activate. Consequently, the queue sizes of the nodes in U will tend to decrease at rate $c - \rho_U > 0$, while the queue sizes of the nodes in V will tend to increase at rate $\rho_V > 0$. While the packet arrivals and activity periods are governed by random processes, the trajectories of the queue sizes will be roughly linear when viewed on the long time scales of interest.

As mentioned in Section 1.1.4, we focus on queue-based random-access protocols where the activation rates are functions of the queue lengths at the various nodes. We call these protocols internal models and in particular we study the internal model with activation rates as described in (1.5). Since we assume identical initial queue sizes within the sets U and V, the asymptotic distribution of the transition time in the internal model should be close to that in the external model described in (1.6) when we choose

$$h_U(t) = g_U\big(Q_U(0) - (c - \rho_U)t\big), \qquad h_V(t) = g_V\big(Q_V(0) + \rho_V t\big), \tag{2.4}$$

with $Q_U(0) = \gamma_U r$ and $Q_V(0) = \gamma_V r$. Next, we formalize our four main models of interest.

Definition 2.1.3 (Models).
Let $\delta > 0$.

- In the *internal model* the deactivation Poisson clocks tick at rate 1, while the activation Poisson clocks tick at rate

$$r_i^{\text{int}}(t) = \begin{cases} g_U(Q_i(t)), & i \in U, \\ g_V(Q_i(t)), & i \in V, \end{cases} \qquad t \geq 0. \tag{2.5}$$

- In the *external model* the deactivation Poisson clocks tick at rate 1, while the activation Poisson clocks tick at rate

$$r_i^{\text{ext}}(t) = \begin{cases} g_U(\gamma_U r - (c-\rho_U)t), & i \in U, \\ g_V(\gamma_V r + \rho_V t), & i \in V, \end{cases} \qquad t \geq 0. \tag{2.6}$$

- In the *lower external model* the deactivation Poisson clocks tick at rate 1, while the activation Poisson clocks tick at rate

$$r_i^{\text{low}}(t) = \begin{cases} g_U(\gamma_U r - (c-\rho_U)t - \delta r), & i \in U, \\ g_V(\gamma_V r + \rho_V t + \delta r), & i \in V, \end{cases} \qquad t \geq 0. \tag{2.7}$$

- In the *upper external model* the deactivation Poisson clocks tick at rate 1, while the activation Poisson clocks tick at rate

$$r_i^{\text{upp}}(t) = \begin{cases} g_U(\gamma_U r - (c-\rho_U)t + 2\delta r), & i \in U, \\ g_V(\gamma_V r + \rho_V t - \delta r), & i \in V, \end{cases} \qquad t \geq 0. \tag{2.8}$$

Note that in the three external models the activation rates depend on time via certain fixed parameters, while in the internal model they depend on time via the actual queue lengths at the nodes. In the lower external model the activation rates in U tend to be less aggressive than in the internal model (i.e., the activation clocks tick less frequently), while the activation rates in V tend to be more aggressive. In the upper external model the reverse is true: the activation rates in U are more aggressive and the activation rates in V are less aggressive. For simplicity, when considering the external model we sometimes write $r_U(t)$ and $r_V(t)$ for the activation rates at time t of nodes in U and nodes in V, respectively. We will see that the upper external model is actually defined only for $t \in [0, T_U]$ with $T_U = \frac{\gamma_U}{c-\rho_U} r$ (see Section 2.2 for details). However, with high probability as $r \to \infty$, the transition occurs before time T_U.

§2.1.2 Main theorems

The main goal of the chapter is to compare the transition time of the internal model with that of the external model. Through a large deviation analysis of the queue length process at each of the nodes, we define a notion of *good behavior* that allows us to define perturbed models with externally driven activation rates that sandwich the queue lengths of the internal model and its transition time. We show with the help of coupling that, with high probability as $r \to \infty$, the asymptotic behavior of the mean transition time for the internal model is the same as for the external model.

The metastable behavior and the transition time $\mathcal{T}_G$ of a network in which the activation rates are time-dependent in a deterministic way was characterized in [14], with the help of the metastability analysis for hard-core interaction models developed in [59]. For $s \geq 0$, let

$$\nu(s) = \frac{1}{\mathbb{E}_u[\mathcal{T}_G](s)} \tag{2.9}$$

be the inverse mean transition time of the time-homogeneous model where we freeze the the activation rates r_U and r_V at time s, i.e., we consider the model with constant

activation rates

$$r_i^{\text{ext}}(t) = \begin{cases} r_U(s), & i \in U, \\ r_V(s), & i \in V, \end{cases} \qquad t \geq 0. \tag{2.10}$$

Then, for any time scale $M = M(r)$ and any threshold $x \in [0, \infty)$,

$$\lim_{r\to\infty} \mathbb{P}_u\left(\frac{\mathcal{T}_G}{M} > x\right) = \begin{cases} 0, & \text{if } M\nu(Mx) \succ 1, \\ e^{-\int_0^x M\nu(Ms)ds}, & \text{if } M\nu(Mx) \asymp 1, \\ 1, & \text{if } M\nu(Mx) \prec 1. \end{cases} \tag{2.11}$$

(Here, as $r \to \infty$, $a \succ b$ means $b = o(a)$, $a \prec b$ means $a = o(b)$, while $a \asymp b$ means $a = \Theta(b)$.) If we let M_c be the unique solution of the equation

$$M\nu(M) = 1, \tag{2.12}$$

then the transition occurs on the time scale M_c, in the sense that $\mathbb{P}_u(\mathcal{T}_G > t) \approx 1$ for $t \prec M_c$ and $\mathbb{P}_u(\mathcal{T}_G > t) \approx 0$ for $t \succ M_c$. On the critical time scale M_c, the transition time follows an exponential distribution with time-varying rate. It was proven in [59] that, for a complete bipartite graph and $s \in [0, \infty)$,

$$\mathbb{E}_u[\mathcal{T}_G](s) = \frac{1}{|U|}\, r_U(s)^{|U|-1}\,[1 + o(1)], \qquad r \to \infty. \tag{2.13}$$

The following two theorems will be proven in Sections 2.4.1–2.4.2 with the help of (2.9)–(2.13).

Theorem 2.1.4 (Critical time scale in the external model).
The time scale on which the transition occurs is given by

$$M_c = F_c\, r^{1\wedge\beta(|U|-1)}\,[1 + o(1)], \qquad r \to \infty, \tag{2.14}$$

with

$$F_c = \begin{cases} \frac{\gamma_U^{\beta(|U|-1)}}{|U|B^{-(|U|-1)}}, & \text{if } \beta \in (0, \frac{1}{|U|-1}), \\ \frac{\gamma_U}{|U|B^{-(|U|-1)}+(c-\rho_U)}, & \text{if } \beta = \frac{1}{|U|-1}, \\ \frac{\gamma_U}{c-\rho_U}, & \text{if } \beta = (\frac{1}{|U|-1}, \infty). \end{cases} \tag{2.15}$$

Theorem 2.1.5 (Transition time in the external model).
The transition time in the external model satisfies

$$\mathbb{E}_u[\mathcal{T}_G^{\text{ext}}] = F_c\, r^{1\wedge\beta(|U|-1)}\,[1 + o(1)], \qquad r \to \infty. \tag{2.16}$$

with F_c as in (2.15)*, and*

$$\lim_{r\to\infty} \mathbb{P}_u\left(\frac{\mathcal{T}_G^{\text{ext}}}{\mathbb{E}_u[\mathcal{T}_G^{\text{ext}}]} > x\right) = \mathcal{P}(x), \qquad x \in [0, \infty), \tag{2.17}$$

with

$$\mathcal{P}(x) = \begin{cases} e^{-x}, & \text{if } \beta \in (0, \frac{1}{|U|-1}),\ x \in [0, \infty), \\ (1 - Cx)^{\frac{1-C}{C}}, & \text{if } \beta = \frac{1}{|U|-1},\ x \in [0, \frac{1}{C}), \\ 0, & \text{if } \beta = \frac{1}{|U|-1},\ x \in [\frac{1}{C}, \infty), \\ 1, & \text{if } \beta \in (\frac{1}{|U|-1}, \infty),\ x \in [0, 1), \\ 0, & \text{if } \beta \in (\frac{1}{|U|-1}, \infty),\ x \in [1, \infty), \end{cases} \tag{2.18}$$

and $C = \frac{F_c(c-\rho_U)}{\gamma_U} \in (0,1)$.

In other words, the mean transition time scales like M_c, while the law of the transition time divided by its mean is exponential, truncated polynomial or deterministic (see Figure 2.3). We distinguish between these three regimes of behavior and refer to them as *subcritical regime*, *critical regime* and *supercritical regime*, respectively. The deterministic behavior observed in the supercritical regime is also known in the literature as *cut-off*.

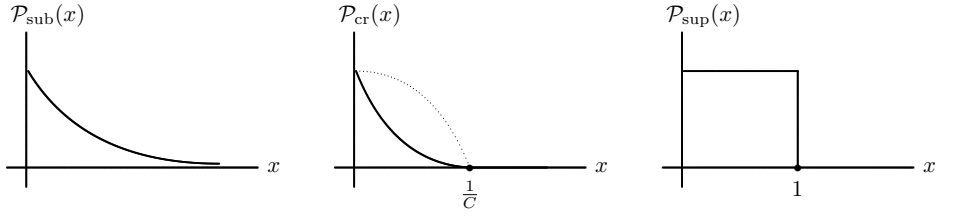


Figure 2.3: Trichotomy for $x \mapsto \mathcal{P}(x)$*:* $\beta \in (0, \frac{1}{|U|-1}]$*, subcritical regime (left);* $\beta = \frac{1}{|U|-1}$*, critical regime (middle);* $\beta \in (\frac{1}{|U|-1}, \infty)$*, supercritical regime (right). The curve in the middle is convex when* $C \in (0, 1/2)$ *and concave when* $C \in (1/2, 1)$*. The curve on the right is the limit of the curve in the middle as* $C \to 1$*.*

As shown in Remark 2.4.1, we can even include the case $\beta = 0$, and get that if $g_U(x) = \hat{\mathcal{L}}(x)$ with $\lim_{x\to\infty} \hat{\mathcal{L}}(x) = \infty$, then

$$\mathbb{E}_u[\mathcal{T}_G^{\text{ext}}] = M_c\,[1 + o(1)], \qquad M_c = \frac{1}{|U|}\hat{\mathcal{L}}(\gamma_U r)^{|U|-1}\,[1 + o(1)], \qquad r \to \infty, \quad (2.19)$$

and $\mathcal{P}(x) = e^{-x}$, $x \in [0, \infty)$. Similar properties hold for the lower and the upper external model, with perturbed $F_{c,\delta}^{\text{low}}$ and $F_{c,\delta}^{\text{upp}}$ satisfying

$$\lim_{\delta\to 0} F_{c,\delta}^{\text{low}} = \lim_{\delta\to 0} F_{c,\delta}^{\text{upp}} = F_c. \quad (2.20)$$

The main result of the chapter is the following sandwich of $\mathcal{T}_G^{\text{int}}$ between $\mathcal{T}_G^{\text{low}}$ and $\mathcal{T}_G^{\text{upp}}$, for which we already know the asymptotic behavior. Because of this sandwich we can deduce the asymptotics of the transition time in the internal model.

Theorem 2.1.6 (Transition time in the internal model).
For $\delta > 0$ *small enough, there exists a coupling such that*

$$\lim_{r\to\infty} \hat{\mathbb{P}}_u\big(\mathcal{T}_G^{\text{low}} \le \mathcal{T}_G^{\text{int}} \le \mathcal{T}_G^{\text{upp}}\big) = 1, \quad (2.21)$$

where $\hat{\mathbb{P}}_u$ *is the joint law induced by the coupling, with all three models starting from* u*. Consequently, the transition time in the internal model satisfies*

$$\mathbb{E}_u[\mathcal{T}_G^{\text{int}}] = F_c\, r^{1\wedge\beta(|U|-1)}\,[1 + o(1)], \qquad r \to \infty, \quad (2.22)$$

with F_c as in (2.15)*, and*

$$\lim_{r\to\infty} \mathbb{P}_u\left(\frac{\mathcal{T}_G^{\text{int}}}{\mathbb{E}_u[\mathcal{T}_G^{\text{int}}]} > x\right) = \mathcal{P}(x), \qquad x \in [0,\infty). \tag{2.23}$$

with $\mathcal{P}(x)$ as in (2.18)*.*

§2.1.3 Discussion and outline

Theorems. Theorem 2.1.5 gives the leading-order asymptotics of the transition time in the external model, including the lower and the upper external model. Theorem 2.1.6 is the main result of the chapter and provides the leading-order asymptotics of the transition time in the internal model, via the coupling in (2.21) and the continuity property in (2.20). Equations (2.15)–(2.16) identify the scaling of the transition time in terms of the model parameters. The trichotomy between $\beta \in (0, \frac{1}{|U|-1})$, $\beta = \frac{1}{|U|-1}$ and $\beta \in (\frac{1}{|U|-1}, \infty)$ is particularly interesting, and leads to different limit laws for the transition time on the scale of its mean.

Interpretation of the trichotomy. In order to interpret the above trichotomy, observe first of all that the activation rates of each of the nodes in U remain of order r^β almost all the way up T_U. Specifically, in the absence of the nodes in V, by time yT_U, $y \in [0,1)$, the queue lengths of the nodes in U have decreased by roughly a fraction y, and their activation rates are approximately $B(1-y)^\beta r^\beta$. Hence the fraction of joint inactivity time of the nodes in U is of order $(1/r^\beta)^{|U|} = r^{-\beta|U|}$. Since the tike it takes to leave the joint inactivity state is of order $r^{-\beta}$, all nodes in U become simultaneously inactive for the first time after a period of order $r^{-\beta}/r^{-\beta|U|} = r^{\beta(|U|-1)}$, which is $o(r)$ in the subcritical regime when $\beta < \frac{1}{|U|-1}$. When the nodes in V are actually present, with high probability as $r \to \infty$, they all activate quickly and the transition occurs almost immediately (see Section 2.4.3). Note that the queue lengths of the nodes in U have only decreased by an amount of order $r^{\beta(|U|-1)} = o(r)$, and hence are still of order r. In contrast, in the critical regime when $\beta = \frac{1}{|U|-1}$, the probability that all nodes in U become simultaneously inactive before time yT_U is approximately $\pi(y)$ with $\pi(y) = 1-(1-y)^{(1-C)/C}$, $y \in [0,1)$ (see (2.18)). Again, with high probability as $r \to \infty$, all the nodes in V activate quickly and the transition occurs almost immediately. Note that the queue lengths in the nodes in U have then dropped by a non-negligible fraction, but are still of order r. A potential scenario is that the nodes in U do not all become simultaneously inactive until their activation rates have become of a smaller order than r^β, due to the queue lengths no longer being of order r just before time T_U. However, the fact that $\pi(y) \to 1$ as $y \to 1$ implies that this scenario has negligible probability in the limit. In contrast, this scenario does occur in the supercritical regime when $\beta > \frac{1}{|U|-1}$, implying that the crossover occurs in a narrow window around T_U (see Sections 2.4.1–2.4.2 for details). We will see that this window has size $O(r^{1/\beta(|U|-1)}) = o(r)$. In particular, the window gets narrower as the activation rates of nodes in U increase.

Proofs. We look at a single-node queue length process $t \mapsto Q(t)$ and prove that with high probability it follows a path that lies in a narrow tube around its mean path (see

Figure 2.4). We study separately the input process $t \mapsto Q^+(t)$ and the output process $t \mapsto Q^-(t)$: we use Mogulskii's theorem (a pathwise large deviation principle) for the first, and Cramér's theorem (a pointwise large deviation principle) for the second. We derive upper and lower bounds for the queue length process and we use these bounds to construct two couplings that allow us to compare the different models.

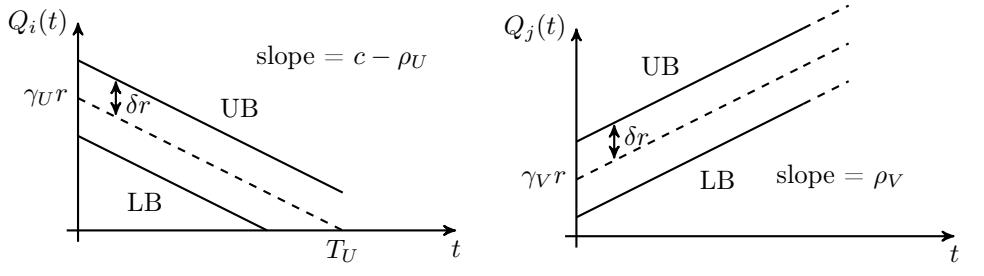


Figure 2.4: Sketches of the tubes around the mean of the queue length processes, respectively, for a node $i \in U$ and a node $j \in V$.

Dependent packet arrivals. Our large deviation estimates are so sharp that we can actually allow the Poisson processes of packet arrivals at the different nodes to be dependent. Indeed, as long at the marginal processes are Poisson, our large deviation estimates are valid at every single node, and since the network is finite a simple union bound shows that they are also valid for all nodes simultaneously, at the expense of a negligible factor that is proportional to the number of nodes. For modeling purposes independent arrivals are natural, but it is interesting to allow for dependent arrivals when we want to study activation protocols that are more involved.

Open problems. If we want to understand how small the term $o(1)$ in (2.22) actually is, then we need to derive sharper estimates in the coupling. One possibility would be to study moderate deviations for the queue length processes and to look at shrinking tubes. We do not pursue such refinements here. Our main focus for the future will be to extend the model to more complicated settings, where the activation rate at node i depends also on the queue length at the neighboring nodes of i. We want to be able to compare models with (externally driven) time-dependent activation rates and models with (internally driven) queue-dependent activation rates, and show again that their metastable behavior is similar. We also want to move away from the complete bipartite interference graph and consider more general graphs that capture more realistic wireless networks.

Other models. There are other ways to define an internal model. We mention a few examples.

(i) A simple variant of our model is obtained by fixing the activation rates, but letting the rate at time t of the Poisson deactivation clock of node i depend on the reciprocal of the queue length at time t, i.e., $1/g_i(Q_i(t))$ for some $g_i \in \mathcal{G}$. This can be equivalently seen as a unit-rate Poisson deactivation clock, where

node i either deactivates with a probability reciprocal to $g_i(Q_i(t))$, or starts a second activity period. Nodes with a large queue length are more likely to remain active for a long time before deactivating, while nodes with a short queue length have extremely short activity periods. If at time t the activation clock of an inactive node with $Q_i(t) = 0$ ticks, then the node does not activate. On the other hand, if during an activity period the queue length of an active node hits zero, then the node deactivates independently of its deactivation rate. For fixed activation and deactivation rates, this model and our internal model with $r_i^{\text{int}}(t) = g_i(Q_i(t))$ for each node i are equivalent up to a time scaling factor. In particular, they have similar stationary distributions.

(ii) An alternative approach is to use a discrete notion of queue length, namely, $Q_i(t) = N_i(t) - S_i(t)$, where $N_i(t)$ is a Poisson process with rate λ, denoting the number of packets arriving at node i during $[0, t]$, while $S_i(t)$ indicates the total number of times node i activates (or deactivates) during $[0, t]$ (we may use λ_U and λ_V to represent different arrival rates for the two sets U and V). The processes $t \mapsto S_i(t)$ and $t \mapsto N_i(t)$ are assumed to be independent. We can define a model where each time a node activates it serves exactly one packet and then deactivates again. The activation clocks still have rates $g_i(Q_i(t))$ with $g_i \in \mathcal{G}$. We can establish results similar to our internal model by adapting the arguments to the discrete setting.

Outline of the chapter. The remainder of this chapter is organized as follows. In Section 2.2 we state large deviation bounds for the input and the output process, which allow us to show that the queue length process at every node has specific lower and upper bounds that hold with very high probability. The proofs of these bounds are deferred to Appendices A–B. In Section 2.3 we use the bounds to couple the lower and the upper external model (with activation rates (2.7) and (2.8), respectively) to the internal model (with activation rates (2.5)). In Section 2.4 we derive the scaling results for the external model, and combine these with the coupling to derive Theorem 2.1.6 (as stated in Section 2.1.2).

§2.2 Bounds for the input and output processes

In this section we state the main results of our analysis of the input process and the output process at a fixed node (recall Definition 1.1.3). With the help of path large deviation techniques, we show that, with high probability as $r \to \infty$, the input process lies in a narrow tube around the deterministic path $t \mapsto (\lambda/\mu)t$ (Proposition 2.2.1). For simplicity, we suppress the index for the arrival rates λ_U and λ_V, and consider a general rate λ. The same holds for $\rho = \lambda/\mu$. We study the output process only for nodes in U, and we give lower and upper bounds in (2.27) and Proposition 2.2.4. We look at a single node and suppress its index, since the queues are independent of each other as long as the nodes remain active or inactive. The proofs of the propositions below for the input process and the output process are given in Appendices A–B, respectively.

Proposition 2.2.1 (Tube for the input process).
For $\delta > 0$ small enough and time horizon $S > 0$, let

$$\Gamma_{S,\delta S} = \left\{ \gamma \in L_\infty([0,S]) \colon \frac{\lambda}{\mu} s - \delta S < \gamma(s) < \frac{\lambda}{\mu} s + \delta S \ \ \forall\, s \in [0,S] \right\}. \tag{2.24}$$

With high probability as $r \to \infty$, the input process lies inside $\Gamma_{S,\delta S}$ as $S \to \infty$, namely,

$$\mathbb{P}\big(Q^+([0,S]) \notin \Gamma_{S,\delta S}\big) = e^{-K_\delta S\,[1+o(1)]}, \qquad S \to \infty. \tag{2.25}$$

$$K_\delta = (\lambda + \delta\mu) + \lambda - 2\sqrt{\lambda(\lambda+\delta\mu)} \in (0,\infty). \tag{2.26}$$

(Note that $\Gamma_{S,\delta S}$ contains negative values. This is of no concern because the path is always non-negative.)

We want to derive lower and upper bounds for the output process for a node in U. An upper bound is trivial by definition, namely,

$$Q^-(t) \le ct, \qquad t \ge 0. \tag{2.27}$$

It is more delicate to compute a lower bound, for which we need some preparatory definitions. We first introduce an auxiliary time that will be useful in our analysis.

Definition 2.2.2 (Auxiliary time).
Consider the internal model and recall that the initial queue lengths at nodes in U are $\gamma_U r$. Define T_U to be the expected time at which the queue length at a node in U hits zero if the transition has not occurred yet. We can write

$$T_U = T_U(r) \sim \alpha r, \qquad r \to \infty, \tag{2.28}$$

with

$$\alpha = \frac{\gamma_U}{c - \rho_U}. \tag{2.29}$$

Note that the quantity αr is the expected time at which the queue length at a node in U hits zero when the node is always active. Since the total inactivity time of a node in U before time T_U will turn out to be negligible compared to αr, we have $T_U \sim \alpha r$ as $r \to \infty$.

Next, we introduce the isolated model, an auxiliary model that will help us to derive a lower bound for the output process. We will see later that the internal model behaves in exactly the same way as the isolated model up to the pre-transition time, in particular, the pre-transition times in the internal and the isolated model coincide in distribution.

Definition 2.2.3 (Isolated model).
In the *isolated model* the activation of nodes in U is not affected by the activity states of nodes in V, i.e., they behave as if they were in isolation. On the other hand, nodes in V are still affected by nodes in U, i.e., they cannot activate until every node in U deactivates. Nodes in V have zero output process.

We study the output process for the isolated model up to time T_U. We will see later in Corollary 2.4.3 that, with high probability as $r \to \infty$, the transition in the internal model occurs before T_U, so it is enough to look at the time interval $[0, T_U]$. In the rare case when the transition does not occur before T_U, we expect it to occur in a very short time after T_U. We are now ready to give the lower bound for the output process.

Proposition 2.2.4 (The output process in the isolated model).
Consider a node in U. For $\delta, \epsilon, \epsilon_1, \epsilon_2 > 0$ small enough, the output process satisfies

$$\begin{aligned}\mathbb{P}_u\big(Q^-(t) < ct - \epsilon r \ \ \forall t \in [0, T_U]\big) \leq\, & e^{-K_\delta \alpha r\,[1+o(1)]} + e^{-K_1 r\,[1+o(1)]} \\ & + e^{-\left(K_2 r + K_3 \frac{r}{g_U(r)} + K_4 r \log g_U(r)\right)\,[1+o(1)]}, \quad r \to \infty,\end{aligned} \tag{2.30}$$

with

$$\begin{aligned} K_1 &= \left(\gamma_U - \frac{2\delta\alpha}{c-\rho_U}\right)\frac{\epsilon_1 - \log(1+\epsilon_1)}{1+\epsilon_1}, \\ K_2 &= \left(\gamma_U - \frac{2\delta\alpha}{c-\rho_U}\right)(1+\epsilon_1)\left(-1 - \log\left(\frac{\epsilon_2}{\big(\gamma_U - \frac{2\delta\alpha}{c-\rho_U}\big)(1+\epsilon_1)}\right)\right), \\ K_3 &= \epsilon_2, \\ K_4 &= \left(\gamma_U - \frac{2\delta\alpha}{c-\rho_U}\right)(1+\epsilon_1), \end{aligned} \tag{2.31}$$

satisfying $K_1, K_2, K_3, K_4 \in (0, \infty)$.

By combining the bounds for the input process and the output process, and picking $\delta = \epsilon$ and $S = r$, we obtain lower and upper bounds for the queue length process $Q(t)$ of a node in U.

Corollary 2.2.5 (The queue length process in the isolated model).
For $\delta > 0$ small enough, with high probability as $r \to \infty$, the following bounds hold for a node in U:

$$\begin{aligned} (LB)_U\colon &\quad Q(t) \geq Q_U^{\mathrm{LB}}(t) = \gamma_U r - (c-\rho_U)t - \delta r, \quad t \geq 0, \\ (UB)_U\colon &\quad Q(t) \leq Q_U^{\mathrm{UB}}(t) = \gamma_U r - (c-\rho_U)t + 2\delta r, \quad t \in [0, T_U]. \end{aligned} \tag{2.32}$$

Similarly, with high probability as $r \to \infty$, the following bounds hold for a node in V:

$$\begin{aligned} (LB)_V\colon &\quad Q(t) \geq Q_V^{\mathrm{LB}}(t) = \gamma_V r + \rho_V t - \delta r, \quad t \geq 0, \\ (UB)_V\colon &\quad Q(t) \leq Q_V^{\mathrm{UB}}(t) = \gamma_V r + \rho_V t + \delta r, \quad t \geq 0. \end{aligned} \tag{2.33}$$

Proof. The claim follows from Propositions 2.2.1 and 2.2.4 in combination with the bound in (2.27). □

§2.3 Coupling the internal and the external model

In Sections 2.3.1–2.3.2 we use the bounds defined in Section 2.2 to construct two couplings that allow us to compare the internal and the external model (Proposition 2.3.5,

respectively, Proposition 2.3.8 and Corollary 2.3.9). Throughout the sequel we assume that the deactivation rates are fixed, i.e., the deactivation Poisson clocks ring at rate 1. A node can activate only if all its neighbors are inactive. If a node is inactive, then the activation Poisson clocks ring at rates that vary over time in a deterministic way, or as functions of the queue lengths.

We are interested in coupling the models in the time interval $[0, T_U]$ and on the following event.

Definition 2.3.1 (Good behavior).
Let $\mathcal{E}_\delta$ be the event that the queue length processes in the internal model all have *good behavior* in the interval $[0, T_U]$, in the sense that

$$\begin{aligned}\mathcal{E}_\delta = &\left\{Q_U^{\mathrm{LB}}(t) \le Q_i(t) \le Q_U^{\mathrm{UB}}(t)\ \forall\, t \in [0, T_U]\ \forall\, i \in U\right\} \\ &\cup \left\{Q_V^{\mathrm{LB}}(t) \le Q_i(t) \le Q_V^{\mathrm{UB}}(t)\ \forall\, t \in [0, T_U]\ \forall\, i \in V\right\},\end{aligned} \tag{2.34}$$

i.e., the paths lie between their respectively lower and upper bounds for nodes in U and V. This event depends on the perturbation parameter δ.

Lemma 2.3.2 (Probability of good behavior).
For $\delta > 0$ small enough,

$$\lim_{r\to\infty} \mathbb{P}_u(\mathcal{E}_\delta) = 1. \tag{2.35}$$

Proof. The claim follows from Corollary 2.2.5. □

In what follows we couple on the event $\mathcal{E}_\delta$ only. The coupling can be extended in an arbitrary way off the event $\mathcal{E}_\delta$. The way this is done is irrelevant because of Lemma 2.3.2.

§2.3.1 Coupling the internal and the lower external model

The lower external model defined in (2.7) can also be described in the following way. At time $t \ge 0$ the activation rates are

$$r_i^{\mathrm{low}}(t) = \begin{cases} g_U(Q_U^{\mathrm{LB}}(t)), & i \in U, \\ g_V(Q_V^{\mathrm{UB}}(t)), & i \in V. \end{cases} \tag{2.36}$$

Note that when the lower bound $Q_U^{\mathrm{LB}}(t)$ becomes negative the activation function g_U is zero by definition. In this way we are able extend the coupling to any time $t \ge 0$, even though we consider only the interval $[0, T_U]$.

Lemma 2.3.3 (Upper bound in the lower external model).
With high probability as $r \to \infty$, the transition time $\mathcal{T}_G^{\mathrm{low}}$ in the lower external model is smaller than T_U, i.e.,

$$\lim_{r\to\infty} \mathbb{P}_u(\mathcal{T}_G^{\mathrm{low}} \le T_U) = 1. \tag{2.37}$$

Proof. As we will see in Section 2.4.2, with high probability as $r \to \infty$, the transition time in the external model is smaller than T_U. Since the lower external model is defined for an arbitrarily small perturbation $\delta > 0$, we conclude by using the continuity of g_U, g_V. □

We introduce a system that allows us to couple the internal model with the lower external model.

Definition 2.3.4 (Coupling system for the lower external model). Suppose that $h_i(t) \geq \max\{Q_U^{\mathrm{UB}}(t), Q_V^{\mathrm{UB}}(t)\}$ for all $i \in U \sqcup V$ and all $t \in [0, T_U]$. Consider a system $\mathcal{H}^{\mathrm{low}}$ where clocks are associated with each node in the following way.

- A Poisson deactivation clock ticks at rate 1. Both the nodes in the lower external model and in the internal model are governed by this clock:
  - if both nodes are active, then they deactivate together;
  - if only one node is active, then it deactivates;
  - if both nodes are inactive, then nothing happens.
- A Poisson activation clock ticks at rate $g_U(h_i(t))$ at time t for a node $i \in U$. Both the nodes in the lower external model and in the internal model are governed by this clock:
  - if both nodes are active, or both are inactive but have active neighbors, then nothing happens;
  - if the node in the internal model is active and the node in the lower external model is not, then the latter node activates (if it can) with probability
  $$\frac{r_i^{\mathrm{low}}(t)}{g_U(h_i(t))}; \tag{2.38}$$
  - if both nodes are inactive but can be activated, then this happens with probabilities
  $$\begin{aligned} &\frac{r_i^{\mathrm{low}}(t)}{g_U(h_i(t))} && \text{for the lower external model,} \\ &\frac{r_i^{\mathrm{int}}(t)}{g_U(h_i(t))} && \text{for the internal model,} \end{aligned} \tag{2.39}$$
  where
  $$\frac{r_i^{\mathrm{low}}(t)}{g_U(h_i(t))} \leq \frac{r_i^{\mathrm{int}}(t)}{g_U(h_i(t))}, \tag{2.40}$$
  in such a way that if the node in the lower external model activates, then it also activates in the internal model.

- A Poisson activation clock ticks at rate $g_V(h_i(t))$ at time t for a node $i \in V$. The same happens as for the nodes in U, but the activation probabilities are

$$\begin{aligned} &\frac{r_i^{\text{low}}(t)}{g_V(h_i(t))} \quad \text{for the lower external model,} \\ &\frac{r_i^{\text{int}}(t)}{g_V(h_i(t))} \quad \text{for the internal model,} \end{aligned} \tag{2.41}$$

  where

$$\frac{r_i^{\text{low}}(t)}{g_U(h_i(t))} \geq \frac{r_i^{\text{int}}(t)}{g_U(h_i(t))}, \tag{2.42}$$

  in such a way that if the node in the internal model activates, then it also activates in the lower external model.

With the constructions above, we are now able to compare the transition times of the two models.

Proposition 2.3.5 (Coupling the internal and the lower external model). *The following statements hold.*

(i) Under the coupling $\mathcal{H}^{\text{low}}$, the joint activity processes in the internal and in the lower external model are ordered for all $t \in [0, T_U]$, i.e.,

$$\begin{aligned} X_i^{\text{low}}(t) \leq X_i^{\text{int}}(t), \quad i \in U, \\ X_i^{\text{int}}(t) \leq X_i^{\text{low}}(t), \quad i \in V. \end{aligned} \tag{2.43}$$

(ii) With high probability as $r \to \infty$, the transition time $\mathcal{T}_G^{\text{int}}$ in the internal model is at least as large as the transition time $\mathcal{T}_G^{\text{low}}$ in the lower external model, i.e.,

$$\lim_{r\to\infty} \hat{\mathbb{P}}_u(\mathcal{T}_G^{\text{low}} \leq \mathcal{T}_G^{\text{int}}) = 1, \tag{2.44}$$

where $\hat{\mathbb{P}}_u$ is the joint law induced by the coupling with starting u.

Proof. We prove the two statements separately.

(i) For each node $i \in U$ and for all $t \in [0, T_U]$, we have that $Q_i^{\text{LB}}(t) \leq Q_i(t)$ and $g_U(Q_i^{\text{LB}}(t)) \leq g_U(Q_i(t))$ by the monotonicity of the function g_U. On the other hand, for each node $i \in V$, $Q_i(t) \leq Q_i^{\text{UB}}(t)$ and $g_V(Q_i(t)) \leq g_V(Q_i^{\text{UB}}(t))$ by the monotonicity of the function g_V. Under the system $\mathcal{H}^{\text{low}}$, at any moment the random variable describing the state of a node $i \in U$ in the lower external model is dominated by the one in the internal model, i.e., by (2.40) for all $t \in [0, T_U]$,

$$X_i^{\text{low}}(t) \leq X_i^{\text{int}}(t). \tag{2.45}$$

On the other hand, the random variable describing the state of a node $j \in V$ in the lower external model dominates the one in the internal model, i.e., by (2.42) for all $t \in [0, T_U]$,

$$X_i^{\text{int}}(t) \leq X_i^{\text{low}}(t). \tag{2.46}$$

Hence the joint activity processes in the two models are ordered.

(ii) Using the coupling construction and the ordering above, we can show that, on the event $\mathcal{E}_\delta$, the nodes in U in the lower external model deactivate earlier than in the internal model, and the nodes in V activate earlier in the lower external model. Hence the transition occurs earlier in the lower external model.

Note that we are able to compare the transition times only when $\mathcal{T}_G^{\text{low}} \leq T_U$, so we look at the coupling also on the event $\{\mathcal{T}_G^{\text{low}} \leq T_U\}$, which has high probability as $r \to \infty$ (Lemma 2.3.3). On this event we have $\mathcal{T}_G^{\text{low}} \leq \mathcal{T}_G^{\text{int}}$. Therefore

$$1 = \lim_{r\to\infty} \hat{\mathbb{P}}_u(\mathcal{E}_\delta, \mathcal{T}_G^{\text{low}} \leq T_U, \mathcal{T}_G^{\text{low}} \leq \mathcal{T}_G^{\text{int}}) = \lim_{r\to\infty} \hat{\mathbb{P}}_u(\mathcal{T}_G^{\text{low}} \leq \mathcal{T}_G^{\text{int}}). \tag{2.47}$$

□

§2.3.2 Coupling the isolated and the upper external model

The upper external model defined in (2.8) can also be described in the following way. At time $t \in [0, T_U]$ the activation rates are

$$r_i^{\text{upp}}(t) = \begin{cases} g_U(Q_U^{\text{UB}}(t)), & i \in U, \\ g_V(Q_V^{\text{LB}}(t)), & i \in V. \end{cases} \tag{2.48}$$

Lemma 2.3.6 (Upper bound in the upper external model).
With high probability as $r \to \infty$, the transition time $\mathcal{T}_G^{\text{upp}}$ in the upper external model is smaller than T_U, i.e.,

$$\lim_{r\to\infty} \mathbb{P}_u(\mathcal{T}_G^{\text{upp}} \leq T_U) = 1. \tag{2.49}$$

This statement is to be read as follows. Let δ be the perturbation parameter in the upper external model appearing in (2.8)*. Then for every $\delta > 0$ there exists a $\delta'(\delta) > 0$, satisfying $\lim_{\delta\to 0} \delta'(\delta) = 0$, such that $\lim_{r\to\infty} \mathbb{P}_u(\mathcal{T}_G^{\text{upp}} \leq [1 + \delta'(\delta)]T_U) = 1$.*

Proof. Analogous to the proof of Lemma 2.3.3. □

We introduce a system that allows us to couple the isolated model with the upper external model up to time τ_G^{iso}.

Definition 2.3.7 (Coupling system for the upper external model).
Suppose that $h_i(t) \geq \max\{Q_U^{\text{UB}}(t), Q_V^{\text{UB}}(t)\}$ for all $i \in U \sqcup V$ and all $t \in [0, \tau_G^{\text{iso}}]$. Couple the processes in the same way as in Definition 2.3.4 for $\mathcal{H}^{\text{low}}$, but with different activation probabilities. The probabilities for the isolated model and for the upper external model are such that

$$\begin{aligned} \frac{r_i^{\text{iso}}(t)}{g_U(h_i(t))} &\leq \frac{r_i^{\text{upp}}(t)}{g_U(h_i(t))}, \quad i \in U, \\ \frac{r_i^{\text{upp}}(t)}{g_V(h_i(t))} &\leq \frac{r_i^{\text{iso}}(t)}{g_V(h_i(t))}, \quad i \in V, \end{aligned} \tag{2.50}$$

where for $t \in [0, \tau_G^{\text{iso}}]$

$$r_i^{\text{iso}}(t) = \begin{cases} g_U(Q_i(t)), & i \in U, \\ g_V(Q_i(t)), & i \in V. \end{cases} \tag{2.51}$$

Note that when $\tau_G^{\text{iso}} \leq T_U$, the isolated model behaves exactly as the internal model in the interval $[0, \tau_G^{\text{iso}}]$, as shown in Appendix B.2. Moreover, the coupling is defined only when $\tau_G^{\text{iso}} \leq T_U$. We look then at the coupling also on the event $\{\mathcal{T}_G^{\text{upp}} \leq T_U\}$, which has high probability as $r \to \infty$ (Lemma 2.3.6). In the following proposition we see how this ensures that the coupling is well defined, and we compare the pre-transition times of the two models.

Proposition 2.3.8 (Coupling the isolated and the upper external model). *The following statements hold.*

(i) Under the coupling $\mathcal{H}^{\text{upp}}$, the joint activity processes in the isolated model and in the upper external model are ordered up to time τ_G^{iso}, i.e., for all $t \in [0, \tau_G^{\text{iso}}]$,

$$\begin{aligned} X_i^{\text{iso}}(t) &\leq X_i^{\text{upp}}(t), \quad i \in U, \\ X_i^{\text{upp}}(t) &\leq X_i^{\text{iso}}(t), \quad i \in V. \end{aligned} \tag{2.52}$$

(ii) With high probability as $r \to \infty$, the pre-transition time τ_G^{upp} in the upper external model is at least as large as the pre-transition time τ_G^{iso} in the isolated model, i.e.,

$$\lim_{r\to\infty} \hat{\mathbb{P}}_u(\tau_G^{\text{iso}} \leq \tau_G^{\text{upp}}) = 1, \tag{2.53}$$

where $\hat{\mathbb{P}}_u$ is the joint law induced by the coupling with starting u.

Proof. We prove the two statements separately.

(i) The proof is analogous to that of Proposition 2.3.5, but this time we use the system $\mathcal{H}^{\text{upp}}$ up to time τ_G^{iso} and all the inequalities are reversed.

(ii) Using the coupling construction and the ordering above, we can show that, on the event $\mathcal{E}_\delta \cap \{\mathcal{T}_G^{\text{upp}} \leq T_U\}$, the nodes in U in the isolated model deactivate earlier than in the upper external model, and the first activating node in V activates earlier in the isolated model. Hence the pre-transition occurs earlier in the isolated model, and we have $\tau_G^{\text{iso}} \leq \tau_G^{\text{upp}} \leq \mathcal{T}_G^{\text{upp}} \leq T_U$. Therefore the coupling is well defined and

$$1 = \lim_{r\to\infty} \hat{\mathbb{P}}_u(\mathcal{E}_{\delta,T_U}, \mathcal{T}_G^{\text{upp}} \leq T_U, \tau_G^{\text{iso}} \leq \tau_G^{\text{upp}}) = \lim_{r\to\infty} \hat{\mathbb{P}}_u(\tau_G^{\text{iso}} \leq \tau_G^{\text{upp}}). \tag{2.54}$$

□

Corollary 2.3.9 (Comparing times between models).
With high probability as $r \to \infty$, the transition time $\mathcal{T}_G^{\text{upp}}$ in the upper external model is at least as large as the pre-transition time τ_G^{int} in the internal model, i.e.,

$$\lim_{r\to\infty} \hat{\mathbb{P}}_u(\tau_G^{\text{int}} \leq \mathcal{T}_G^{\text{upp}}) = 1. \tag{2.55}$$

Proof. Since $\lim_{r\to\infty} \mathbb{P}(\tau_G^{\text{iso}} \le T_U) = 1$, we have, as shown in Proposition B.6 in Appendix B.2, that the pre-transition times in the isolated model and in the internal model coincide. Hence

$$1 = \lim_{r\to\infty} \hat{\mathbb{P}}_u(\tau_G^{\text{iso}} \le \tau_G^{\text{upp}}) = \lim_{r\to\infty} \hat{\mathbb{P}}_u(\tau_G^{\text{int}} \le \tau_G^{\text{upp}}) \le \lim_{r\to\infty} \hat{\mathbb{P}}_u(\tau_G^{\text{int}} \le \mathcal{T}_G^{\text{upp}}), \tag{2.56}$$

which completes the proof. □

§2.4 Proofs of the main results

The goal of this section is to identify the asymptotic behavior of the transition time in the internal model. In Sections 2.4.1–2.4.2 we look at the external model and prove Theorems 2.1.4–2.1.5, respectively. In Section 2.4.3 we show that the difference between the transition time and the pre-transition time is negligible for all the models considered. In Section 2.4.4 we put these results together to prove Theorem 2.1.6.

§2.4.1 Proof: critical time scale in the external model

In this section we prove Theorem 2.1.4. From now on we write $a(r) \sim b(r)$ to indicate that $\lim_{r\to\infty} a(r)/b(r) = 1$, while we write $a(r) \asymp b(r)$ to indicate that $0 < \liminf_{r\to\infty} a(r)/b(r) \le \limsup_{r\to\infty} a(r)/b(r) < \infty$.

Proof of Theorem 2.1.4. In order to compute the critical time scale M_c, we must solve the equation $M\nu(M) = 1$ in (2.12). We know from (2.9) and (2.13) that

$$\nu(s) \sim |U| r_U(s)^{1-|U|}, \qquad r \to \infty. \tag{2.57}$$

We want to identify how the transition time is related to the choice of g_U in Definition 2.1.1. Consider the time scale $M_c = F_c r^\gamma$, where $\gamma \in (0,1]$ and $F_c \in (0,\infty)$. As $r \to \infty$, we have

$$\begin{aligned} 1 = r^0 = M_c\nu(M_c) &= F_c r^\gamma\, \nu(F_c r^\gamma) \sim F_c r^\gamma |U| r_U(F_c r^\gamma)^{-(|U|-1)} \\ &= F_c r^\gamma |U| g_U\big(\gamma_U r - (c-\rho_U)F_c r^\gamma\big)^{-(|U|-1)} \\ &\sim F_c r^\gamma |U| B^{-(|U|-1)}\big(\gamma_U r - (c-\rho_U)F_c r^\gamma\big)^{-\beta(|U|-1)}. \end{aligned} \tag{2.58}$$

Recall from (2.29) that $\alpha = \frac{\gamma_U}{c-\rho_U}$. We distinguish between three cases.

(I) Case $\gamma \in (0,1)$ and $F_c \in (0,\infty)$. As $r \to \infty$, the criterion in (2.58) reads

$$1 = r^0 \sim F_c r^\gamma |U| B^{-(|U|-1)} (\gamma_U r)^{-\beta(|U|-1)}. \tag{2.59}$$

In order for the exponents of r to match, we need

$$\beta = \frac{\gamma}{|U|-1}. \tag{2.60}$$

Inserting (2.60) into (2.59), we get

$$F_c |U| B^{-(|U|-1)} \gamma_U^{-\beta(|U|-1)} = 1, \tag{2.61}$$

which gives

$$F_c = \frac{\gamma_U^{\beta(|U|-1)} B^{(|U|-1)}}{|U|}. \tag{2.62}$$

Hence

$$M_c = \frac{(B\gamma_U^{\beta})^{|U|-1}}{|U|} r^{\beta(|U|-1)}, \qquad r \to \infty. \tag{2.63}$$

(II) Case $\gamma = 1$ and $F_c \in (0, \alpha)$. As $r \to \infty$, the criterion in (2.58) reads

$$1 = r^0 \sim F_c |U| B^{-(|U|-1)} (\gamma_U - (c - \rho_U)F_c)^{-\beta(|U|-1)} r^{1-\beta(|U|-1)}. \tag{2.64}$$

In order for the exponents of r to match, we need

$$\beta = \frac{1}{|U|-1}. \tag{2.65}$$

Inserting (2.65) into (2.64), we get

$$\frac{F_c |U| B^{-(|U|-1)}}{\gamma_U - (c - \rho_U)F_c} = 1, \tag{2.66}$$

which gives

$$F_c = \frac{\gamma_U}{|U| B^{-(|U|-1)} + (c - \rho_U)}. \tag{2.67}$$

Hence

$$M_c = \frac{\gamma_U}{|U| B^{-(|U|-1)} + (c - \rho_U)} r, \qquad r \to \infty. \tag{2.68}$$

Recall from (2.28) that $T_U \sim \alpha r$ is the expected time at which the queue length at a node in U hits zero. We will see in Section 2.4.2 that the transition in the external model typically occurs before the queues are empty.

(III) Case $\gamma = 1$ and $F_c = \alpha - Dr^{-\delta}$, $\delta \in (0, 1)$. As $r \to \infty$, the criterion in (2.58) reads

$$1 = r^0 \sim \alpha r |U| B^{-(|U|-1)} \big((c - \rho_U) D r^{1-\delta}\big)^{-\beta(|U|-1)}. \tag{2.69}$$

In order for the exponents of r to match, we need

$$\beta = \frac{1}{(1-\delta)(|U|-1)}. \tag{2.70}$$

Inserting (2.70) into (2.69), we get

$$\alpha |U| B^{-(|U|-1)} ((c - \rho_U) D)^{-\beta(|U|-1)} = 1, \tag{2.71}$$

which gives

$$D = \frac{(\alpha |U| B^{-(|U|-1)})^{1/\beta(|U|-1)}}{c - \rho_U}. \tag{2.72}$$

Hence

$$M_c = \alpha r - \frac{(\alpha |U| B^{-(|U|-1)})^{1/\beta(|U|-1)}}{c - \rho_U} r^{1/\beta(|U|-1)} = \alpha r, \qquad r \to \infty, \tag{2.73}$$

and so the crossover takes place in a window of size $O(r^{1/\beta(|U|-1)}) = o(r)$ around αr. Note that this window gets narrower as β increases, i.e., as the activation rates of nodes in U increase.

□

In the following remark we discuss more general activation functions g_U.

Remark 2.4.1 (Modulation with slowly varying functions).
Consider activation functions of the form $g_U(x) = x^\beta \hat{\mathcal{L}}(x)$ with $\beta \in (0,\infty)$ and $\hat{\mathcal{L}}(x)$ a slowly varying function (i.e., $\lim_{x\to\infty} \hat{\mathcal{L}}(ax)/\hat{\mathcal{L}}(x) = 1$ for all $a > 0$). Let $M_c = r^\gamma \mathcal{L}(r)$ with $\gamma \in (0,1)$ and $\mathcal{L}(r)$ a slowly varying function. As $r \to \infty$, we have

$$\begin{aligned} 1 = r^0 &\sim r^\gamma \mathcal{L}(r)|U|\big(\gamma_U r - (c-\rho_U) r^\gamma \mathcal{L}(r)\big)^{-\beta(|U|-1)} \hat{\mathcal{L}}(\gamma_U r - (c-\rho_U) r^\gamma \mathcal{L}(r))^{-(|U|-1)} \\ &\sim r^\gamma \mathcal{L}(r)|U|(\gamma_U r)^{-\beta(|U|-1)} \hat{\mathcal{L}}(\gamma_U r)^{-(|U|-1)}. \end{aligned} \tag{2.74}$$

In order for the exponents of r to match, we again need

$$\beta = \frac{\gamma}{|U|-1}. \tag{2.75}$$

We get

$$\mathcal{L}(r) = \frac{\gamma_U^{\beta(|U|-1)}}{|U|} \hat{\mathcal{L}}(\gamma_U r)^{|U|-1}, \qquad r \to \infty. \tag{2.76}$$

Hence

$$M_c = \frac{\gamma_U^{\beta(|U|-1)}}{|U|} r^{\beta(|U|-1)} \hat{\mathcal{L}}(\gamma_U r)^{|U|-1}, \qquad r \to \infty. \tag{2.77}$$

We can even include the case $\beta = 0$, in which we obtain that if $g_U(x) = \hat{\mathcal{L}}(x)$ with $\lim_{x\to\infty} \hat{\mathcal{L}}(x) = \infty$, then

$$M_c = \frac{1}{|U|} \hat{\mathcal{L}}(\gamma_U r)^{|U|-1}, \qquad r \to \infty. \tag{2.78}$$

§2.4.2 Proof: transition time in the external model

In this section we prove Theorem 2.1.5. We already know that the transition occurs on the critical time scale M_c computed in Section 2.4.2.

Proof of Theorem 2.1.5. Knowing the critical time scale M_c, we can compute the mean transition time from (2.11). As $r \to \infty$, we have

$$\begin{aligned} \mathbb{E}_u[\mathcal{T}_G^{\text{ext}}] &= \int_0^\infty \mathbb{P}_u(\mathcal{T}_G^{\text{ext}} > x)\, dx = M_c \int_0^\infty \mathbb{P}_u\left(\frac{\mathcal{T}_G^{\text{ext}}}{M_c} > x\right) dx \\ &\sim M_c \int_0^\infty e^{-\int_0^x M_c \nu(M_c s)\, ds}\, dx = M_c \int_0^\infty e^{-\int_0^x \frac{M_c \nu(M_c s)}{M_c \nu(M_c)} ds}\, dx \\ &= M_c \int_0^\infty e^{-\int_0^x \left(\frac{\gamma_U r - (c-\rho_U) M_c s}{\gamma_U r - (c-\rho_U) M_c}\right)^{-\beta(|U|-1)} ds}\, dx, \end{aligned} \tag{2.79}$$

where the choice of β is important. We distinguish between the three cases.

(I) Case $\beta \in (0, \frac{1}{|U|-1})$, $M_c = F_c r^\gamma$, $\gamma \in (0,1)$. We have

$$\lim_{r\to\infty} \left(\frac{\gamma_U r - (c-\rho_U) M_c s}{\gamma_U r - (c-\rho_U) M_c} \right)^{-\beta(|U|-1)} = 1. \tag{2.80}$$

Hence, as $r \to \infty$,

$$\mathbb{E}_u[\mathcal{T}_G^{\text{ext}}] \sim M_c \int_0^\infty e^{-\int_0^x ds}\, dx = M_c \int_0^\infty e^{-x}\, dx = M_c. \tag{2.81}$$

The law of $\mathcal{T}_G^{\text{ext}}$ is exponential, i.e.,

$$\lim_{r\to\infty} \mathbb{P}_u\left(\frac{\mathcal{T}_G^{\text{ext}}}{\mathbb{E}_u[\mathcal{T}_G^{\text{ext}}]} > x \right) = e^{-x}, \qquad x \in [0,\infty). \tag{2.82}$$

(II) Case $\beta = \frac{1}{|U|-1}$, $M_c = F_c r$, $F_c \in (0, \alpha)$. We have

$$\begin{aligned} \lim_{r\to\infty} \left(\frac{\gamma_U r - (c-\rho_U) F_c r s}{\gamma_U r - (c-\rho_U) F_c r} \right)^{-\beta(|U|-1)} &= \frac{\gamma_U - (c-\rho_U) F_c}{\gamma_U - (c-\rho_U) F_c s} = \frac{1 - \frac{c-\rho_U}{\gamma_U} F_c}{1 - \frac{c-\rho_U}{\gamma_U} F_c s} \\ &= \frac{1 - \frac{F_c}{\alpha}}{1 - \frac{F_c}{\alpha} s} = \frac{1-C}{1-Cs}, \end{aligned} \tag{2.83}$$

with $C = F_c/\alpha$. Hence, as $r \to \infty$,

$$\begin{aligned} \mathbb{E}_u[\mathcal{T}_G^{\text{ext}}] &\sim M_c \int_0^{\frac{1}{C}} e^{-\int_0^x \frac{1-C}{1-Cs} ds}\, dx = M_c \int_0^{\frac{1}{C}} e^{-\log(1-Cx)^{-\frac{1-C}{C}}}\, dx \\ &= M_c \int_0^{\frac{1}{C}} (1-Cx)^{\frac{1-C}{C}}\, dx = M_c \left[(1-Cx)^{1+\frac{1-C}{C}} \frac{1}{(1+\frac{1-C}{C})(-C)} \right]_0^{\frac{1}{C}} \\ &= M_c \left[-(1-Cx)^{\frac{1}{C}} \right]_0^{\frac{1}{C}} = M_c. \end{aligned} \tag{2.84}$$

Here, the integral must be truncated at $x = 1/C$ because for larger x the integrand becomes negative. Indeed, note that when $x = 1/C = \alpha/F_c$, which corresponds to time $T_U = \alpha r$, we have

$$\begin{aligned} \lim_{r\to\infty} \mathbb{P}_u\big(\mathcal{T}_G^{\text{ext}} > T_U\big) = \lim_{r\to\infty} \mathbb{P}_u\left(\mathcal{T}_G^{\text{ext}} > \frac{\alpha}{F_c} F_c r \right) &= \lim_{r\to\infty} \mathbb{P}_u\left(\frac{\mathcal{T}_G^{\text{ext}}}{M_c} > \frac{\alpha}{F_c} \right) \\ &= \left(1 - C\frac{\alpha}{F_c} \right)^{\frac{1-C}{C}} = 0, \end{aligned} \tag{2.85}$$

because $C = F_c/\alpha$. This means that, with high probability as $r \to \infty$, the transition occurs before time T_U. The law of $\mathcal{T}_G^{\text{ext}}$ is truncated polynomial:

$$\lim_{r\to\infty} \mathbb{P}_u\left(\frac{\mathcal{T}_G^{\text{ext}}}{\mathbb{E}_u[\mathcal{T}_G^{\text{ext}}]} > x \right) = \begin{cases} (1-Cx)^{\frac{1-C}{C}}, & x \in [0, \frac{1}{C}), \\ 0, & x \in [\frac{1}{C}, \infty). \end{cases} \tag{2.86}$$

(III) Case $\beta \in (\frac{1}{|U|-1}, \infty)$, $M_c = \alpha r$. This case corresponds to the limit $C \to 1$ of the previous case. In this limit, (2.86) becomes

$$\lim_{r\to\infty} \mathbb{P}_u\left(\frac{\mathcal{T}_G^{\text{ext}}}{\mathbb{E}_u[\mathcal{T}_G^{\text{ext}}]} > x\right) = \begin{cases} 1, & x \in [0,1), \\ 0, & x \in [1,\infty). \end{cases} \tag{2.87}$$

□

Note that the three cases above corresponds to the three regimes of behavior: respectively, the subcritical regime, the critical regime and the supercritical regime.

§2.4.3 Negligible gap in the internal model

In this section we focus on the internal model and estimate the length of the interval $[\tau_G^{\text{int}}, \mathcal{T}_G^{\text{int}}]$, which, with high probability as $r \to \infty$, turns out to be very small with respect to τ_G^{int}. This implies that the transition time has the same asymptotic behavior as the pre-transition time.

We know that the queue at a node $i \in V$ is of order r at time τ_G^{int}, i.e., $Q_i(\tau_G^{\text{int}}) \asymp r$, since it starts at $\gamma_V r$, with $\gamma_V > 0$, and only the input process is present until this time. Hence all the activation Poisson clocks at nodes in V tick at a very aggressive rate. The idea is that within the activation period (which has an exponential distribution with mean 1) of the first node activating in V, all the other nodes in V activate because they are not "blocked" by any node in U. Consequently, the network quickly reaches v.

Theorem 2.4.2 (Negligible gap).
In the internal model

$$\lim_{r\to\infty} \mathbb{P}_u\left(\mathcal{T}_G^{\text{int}} - \tau_G^{\text{int}} = o\left(\frac{1}{g_V(r)}\right)\right) = 1. \tag{2.88}$$

Proof. Starting from τ_G^{int}, a node $x \in V$ remains inactive for an exponential period with mean $1/r_x^{\text{int}}(\tau_G) = 1/g_V(Q(\tau_G)) \asymp 1/g_V(r)$. Denote by W_x the length of an inactivity period for a node $x \in V$. Let x_1 be the first node activating in V. We then have i.i.d. inactivity periods $W_x \simeq \text{Exp}(g_V(Q(\tau_G)))$ for all $x \in V \setminus \{x_1\}$. We label the remaining nodes $x_2, \dots, x_{|V|}$ in an arbitrary way. We also have i.i.d. activity periods $Z_x \simeq \text{Exp}(1)$ for all $x \in V$.

Consider a time $t_1 = o(1/g_V(r))$. With high probability as $r \to \infty$, all the nodes in V activate before time t_1, i.e.,

$$\begin{aligned} \lim_{r\to\infty} \mathbb{P}_u\big(W_{x_i} < t_1,\, \forall\, i = 2, \dots, |V|\big) &= \lim_{r\to\infty} \mathbb{P}_u\big(W_{x_2} < t_1\big)^{|V|-1} \\ &= \lim_{r\to\infty} \big(1 - e^{-g_V(Q(\tau_G))t_1}\big)^{|V|-1} = 1. \end{aligned} \tag{2.89}$$

Moreover, with high probability as $r \to \infty$, once they activated, all nodes in V stay active for a period of length at least $t_2 \asymp 1/g_V(r) > t_1$, i.e.,

$$\begin{aligned} \lim_{r\to\infty} \mathbb{P}_u\big(Z_{x_i} > t_2 \;\; \forall\, i = 1, \dots, |V|\big) &= \lim_{r\to\infty} \mathbb{P}_u\big(Z_{x_1} > t_2\big)^{|V|} \\ &= \lim_{r\to\infty} (e^{-t_2})^{|V|} = 1. \end{aligned} \tag{2.90}$$

In conclusion, with high probability as $r \to \infty$, every node in V activates before time t_1 and remains active for at least a time $t_2 > t_1$. This ensures that the transition occurs before time t_2. In particular, it occurs when the last node in V activates (which occurs even before time t_1), so that $\mathcal{T}_G^{\text{int}} - \tau_G^{\text{int}} = o(1/g_V(r))$. □

Note that this argument extends to any "external" model with activation rates that tend to infinity with r, in particular, to all the models considered in the chapter. The transition always happens quickly after the pre-transition, due to the high level of aggressiveness of nodes in V.

Corollary 2.4.3 (Upper bound on the transition time).
With high probability as $r \to \infty$, the transition time in the internal model is smaller than T_U, i.e.,

$$\lim_{r\to\infty} \mathbb{P}_u(\mathcal{T}_G^{\text{int}} \le T_U) = 1. \tag{2.91}$$

Proof. The claim follows from Lemma 2.3.6, Corollary 2.3.9 and Theorem 2.4.2. □

§2.4.4 Proof: transition time in the internal model

In this section we prove Theorem 2.1.6. First we derive the sandwich of the transition times in the lower external, the internal and the upper external model. After that we identify the asymptotics of the transition time for the internal model by using the results for the external models.

Proof of Theorem 2.1.6. Using Proposition 2.3.5, Corollary 2.3.9 and Theorem 2.4.2, we have that there exists a coupling such that

$$\begin{aligned} 1 &= \lim_{r\to\infty} \hat{\mathbb{P}}_u\left(\mathcal{T}_G^{\text{low}} \le \mathcal{T}_G^{\text{int}}, \mathcal{T}_G^{\text{int}} = \tau_G^{\text{int}} + o\left(\frac{1}{g_V(r)}\right), \tau_G^{\text{int}} \le \mathcal{T}_G^{\text{upp}}\right) \\ &= \lim_{r\to\infty} \hat{\mathbb{P}}_u\left(\mathcal{T}_G^{\text{low}} \le \mathcal{T}_G^{\text{int}} \le \mathcal{T}_G^{\text{upp}} + o\left(\frac{1}{g_V(r)}\right)\right) \\ &= \lim_{r\to\infty} \hat{\mathbb{P}}_u\big(\mathcal{T}_G^{\text{low}} \le \mathcal{T}_G^{\text{int}} \le \mathcal{T}_G^{\text{upp}}\big), \end{aligned} \tag{2.92}$$

where $\hat{\mathbb{P}}_u$ is the joint law of the three models on the same probability space all three starting from u.

By Theorem 2.1.5, we know the law of the transition time in the external model. By construction, we have $\mathbb{E}_u[\mathcal{T}_G^{\text{low}}] \le \mathbb{E}_u[\mathcal{T}_G^{\text{ext}}] \le \mathbb{E}_u[\mathcal{T}_G^{\text{upp}}]$. When considering the lower and the upper external model, the transition time asymptotics are controlled by the prefactors $F_{c,\delta}^{\text{low}}$ and $F_{c,\delta}^{\text{upp}}$, respectively, which are perturbations of the prefactor F_c due to the perturbations of the activation rates. In particular, we know from (2.20) that $\lim_{\delta\to 0} F_{c,\delta}^{\text{low}} = \lim_{\delta\to 0} F_{c,\delta}^{\text{upp}} = F_c$. Hence, for all $\epsilon > 0$,

$$\mathbb{E}_u[\mathcal{T}_G^{\text{int}}] = (F_c \pm \epsilon) r^{\beta(|U|-1)} [1 + o(1)], \qquad r \to \infty, \tag{2.93}$$

and since ϵ can be taken arbitrarily small, it may be absorbed into the $o(1)$-term, as

$$\mathbb{E}_u[\mathcal{T}_G^{\text{int}}] = F_c r^{\beta(|U|-1)} [1 + o(1)], \qquad r \to \infty. \tag{2.94}$$

The same kind of argument applies to the law of the transition time, since for any $x \in [0, \infty)$,

$$\lim_{r\to\infty} \mathbb{P}_u(\mathcal{T}_G^{\text{low}} > x) \leq \lim_{r\to\infty} \mathbb{P}_u(\mathcal{T}_G^{\text{int}} > x) \leq \lim_{r\to\infty} \mathbb{P}_u(\mathcal{T}_G^{\text{upp}} > x). \tag{2.95}$$

This completes the proof. □

§A Appendix: the input process

The main target of this appendix is to prove Proposition 2.2.1 in Section 2.2. We use path large deviation techniques. For simplicity, we suppress the index for the arrival rates λ_U and λ_V, and consider a general rate λ. We show that, with high probability as $r \to \infty$, the input process lies in a narrow tube around the deterministic path $t \mapsto (\lambda/\mu)t$.

Consider a single queue, and for simplicity suppress its index. For $T > 0$, define the scaled process

$$Q_n^+(t) = \frac{1}{n} Q^+(nt) = \frac{1}{n} \sum_{j=1}^{N(nt)} Y_j, \qquad t \in [0, T], \tag{2.96}$$

with $Q_n^+(0) = 0$. We have

$$\mathbb{E}[Q_n^+(t)] = \frac{1}{n} \frac{\lambda n t}{\mu} = \frac{\lambda}{\mu} t, \tag{2.97}$$

and, by the strong law of large numbers, $Q_n^+(t) \to (\lambda/\mu)t$ almost surely for every t as $n \to \infty$.

When studying the process $t \mapsto Q_n^+(t)$, we need to take into account that this is a combination of the Poisson arrival process $t \mapsto N(nt)$ and the exponential service times Y_j, $j \in \mathbb{N}$. Two different types of fluctuations can occur: packets arrive at a slower or faster rate than λ, respectively, service times for each packet are shorter or longer than their mean $1/\mu$. Both need to be considered for a proper large deviation analysis.

§A.1 Large deviation principle for the two components

Definition A.1 (Space of paths).
Consider the space $L_\infty([0, T])$ of *essentially bounded* functions in $[0, T]$, with the norm $\|f\|_\infty = \operatorname{ess\,sup}_{x\in[0,T]} |f(x)|$ called the essential norm. A function f is essentially bounded, i.e., $f \in L_\infty([0, T])$, when there is a measurable function g on $[0, T]$ such that $f = g$ except on a set of measure zero and g is bounded. Let $\mathcal{AC}_T \subset L_\infty([0, T])$ denote the space of *absolutely continuous* functions $f\colon [0, T] \to \mathbb{R}$ such that $f(0) = 0$.

Given the Poisson arrival process $t \mapsto N(nt)$ with rate λ, define the scaled process $t \mapsto Z_n(t)$ by

$$Z_n(t) = \frac{1}{n} N(nt) = \frac{1}{n} \sum_{i=1}^{nt} X_i = \frac{1}{n} \sum_{i=1}^{\lfloor nt \rfloor} X_i, \qquad t \in [0,T], \tag{2.98}$$

where $X_i \simeq \text{Poisson}(\lambda)$ are i.i.d. random variables and $\lfloor x \rfloor$ denotes the greatest integer smaller than or equal to x. Note that $N(nt) \simeq \text{Poisson}(\lambda nt)$. Let ν_n be the law of $(Z_n(t))_{t\in[0,T]}$ on $L_\infty([0,T])$. Note that $Z_n(t)$ is asymptotically equivalent to $N(t)$ with mean $\mathbb{E}[Z_n(t)] = \lambda t$, and $(Z_n(t))_{t\in[0,T]}$ tends to $(\lambda t)_{t\in[0,T]}$ as $n \to \infty$.

We recall the definition of large deviation principle.

Definition A.2 (Large deviation principle (LDP)).
A family of probability measures $(P_n)_{n\in\mathbb{N}}$ on a Polish space $\mathcal{X}$ is said to satisfy the *large deviation principle (LDP)* with rate n and with good rate function $I\colon \mathcal{X} \to [0,\infty]$ if

$$\begin{aligned} \limsup_{n\to\infty} \frac{1}{n} \log P_t(C) &\le -I(C) \qquad \forall\, C \subset \mathcal{X} \text{ closed}, \\ \liminf_{n\to\infty} \frac{1}{n} \log P_t(O) &\ge -I(O) \qquad \forall\, O \subset \mathcal{X} \text{ open}, \end{aligned} \tag{2.99}$$

where $I(S) = \inf_{x\in S} I(x)$, $S \subset \mathcal{X}$. A good rate function satisfies: (1) $I \not\equiv \infty$, (2) I is lower semi-continuous, (3) I has compact level sets.

We begin by stating the LDP for the arrival process $(Z_n(t))_{t\in[0,T]}$.

Lemma A.3 (LDP for the arrival process).
The family of probability measures $(\nu_n)_{n\in\mathbb{N}}$ satisfies the LDP on $L_\infty([0,T])$ with rate n and with good rate function I_N given by

$$I_N(\eta) = \begin{cases} \int_0^T \Lambda_N^*(\dot\eta(t))\, dt, & \eta \in \mathcal{AC}_T, \\ \infty, & \textit{otherwise}, \end{cases} \tag{2.100}$$

where $\Lambda_N^*(x) = x\,\log(x/\lambda) - x + \lambda$, $x \in (0,\infty)$.

Proof. Apply Mogulskii's theorem (see [40, Theorem 5.1.2]). Use the fact that Λ_N^* is the Fenchel-Legendre transform of the cumulant generating function Λ defined by $\Lambda(\theta) = \log \mathbb{E}(e^{\theta X_1})$, $\theta \in \mathbb{R}$. □

For $\Gamma \subset L_\infty([0,T])$, define $I_N(\Gamma) = \inf_{\eta\in\Gamma} I_N(\eta)$. Consequently, the LDP implies that, if $\Gamma \subset L_\infty([0,T])$ is an I_N-continuous set, i.e., $I_N(\Gamma) = I_N(\text{int}(\Gamma)) = I_N(\text{cl}(\Gamma))$, then

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(Z_n([0,T]) \in \Gamma\big) = -I_N(\Gamma). \tag{2.101}$$

Informally, the LDP reads as the approximate statement

$$\mathbb{P}\big(Z_n([0,T]) \approx \eta([0,T]\big) = e^{-nI_N(\eta)[1+o(1)]}, \qquad n \to \infty, \tag{2.102}$$

where $\approx$ stands for close in the essential norm. Informally, on this event we may approximate

$$Q_n^+(t) = \frac{1}{n}\sum_{j=1}^{N(nt)} Y_j = \frac{1}{n}\sum_{j=1}^{nZ_n(t)} Y_j \approx \frac{1}{n}\sum_{j=1}^{n\eta(t)} Y_j = \frac{1}{n}\sum_{j=1}^{\lfloor n\eta(t)\rfloor} Y_j, \qquad t\in[0,T], \quad (2.103)$$

where $\approx$ now stands for close in the Euclidean norm. Given $\eta \in L_\infty([0,T])$, let μ_n^η denote the law of the last sum in (2.103). Below we state the LDP for the input process subject to the arrival process.

Lemma A.4 (LDP for the input process subject to the arrival process).
Given $\eta \in L_\infty([0,T])$, the family of probability measures $(\mu_n^\eta)_{n\in\mathbb{N}}$ satisfies the LDP on $L_\infty([0,T])$ with rate n and with good rate function I_Q^η given by

$$I_Q^\eta(\phi) = \begin{cases} \int_0^T \Lambda_Q^*\left(\frac{d\phi(t)}{d\eta(t)}\right) d\eta(t), & \phi \in \mathcal{AC}_\mathcal{T}, \\ \infty, & \text{otherwise}, \end{cases} \quad (2.104)$$

where $\Lambda_Q^*(x) = x\mu - 1 - \log(x\mu)$, $x \in (0,\infty)$.

Proof. Again apply Mogulskii's theorem, this time with $\eta(t)$ as the time index. Use that Λ^* is the Fenchel-Legendre transform of the cumulant generating function Λ defined by $\Lambda(\theta) = \log \mathbb{E}(e^{\theta Y_1})$, $\theta \in \mathbb{R}$. □

§A.2 Measures in product spaces

The rate function I_Q^η describes the large deviations for the sequence of processes $(Q_n^+(t))_{t\in[0,T]}$ given the path η. To derive the LDP averaged over η, we need a small digression into measures in product spaces.

Definition A.5 (Product measures).
Define the family of probability measures $(\rho_n)_{n\in\mathbb{N}}$ such that $\rho_n = \nu_n\mu_n^\eta$. These measures are defined on the product space $L_\infty([0,T]) \times L_\infty([0,T])$ given by the Cartesian product of the space $L_\infty([0,T])$ with itself, equipped with the product topology.

The open sets in the product topology are unions of sets of the form $U_1 \times U_2$ with U_1, U_2 open in $L_\infty([0,T])$. Moreover, the product of base elements of $L_\infty([0,T])$ gives a basis for the product space $L_\infty([0,T]) \times L_\infty([0,T])$. Define the projections $\mathrm{Pr}_i\colon L_\infty([0,T]) \times L_\infty([0,T]) \to L_\infty([0,T])$, $i = 1,2$, onto the first and the second coordinates, respectively. The product topology on $L_\infty([0,T]) \times L_\infty([0,T])$ is the topology generated by sets of the form $\mathrm{Pr}_i^{-1}(U_i)$, $i = 1,2$, where and U_1, U_2 are open subsets of $L_\infty([0,T])$.

Lemma A.6 (Product LDP).
The family of probability measures $(\rho_n)_{n\in\mathbb{N}}$ satisfies the LDP on $L_\infty([0,T])\times L_\infty([0,T])$ with rate n and with good rate function I given by

$$I(\phi,\eta) = \begin{cases} \int_0^T \Lambda_Q^*\left(\frac{d\phi(t)}{d\eta(t)}\right) d\eta(t) + \int_0^T \Lambda_N^*(\dot\eta(t))\, dt, & \phi,\eta \in \mathcal{AC}_T, \\ \infty, & \text{otherwise.} \end{cases} \quad (2.105)$$

Proof. The claim follows from standard large deviation theory (see [40]). □

§A.3 Large deviation principle for the input process

The contraction principle allows us to derive the LDP averaged over η. Indeed, let $\mathcal{X} = L_\infty([0,T]) \times L_\infty([0,T])$ and $\mathcal{Y} = L_\infty([0,T])$, let $(\rho_n)_{n\in\mathbb{N}}$ be a sequence of product measures on $\mathcal{X}$, and consider the projection Pr_1 onto $\mathcal{Y}$, which is a continuous map. Then the sequence of induced measures $(\mu_n)_{n\in\mathbb{N}}$ given by $\mu_n = \rho_n \,\mathrm{Pr}_1^{-1}$ satisfies the LDP on $L_\infty([0,T])$ with good rate function

$$\tilde{I}_Q(\phi) = \inf_{(\phi,\eta)\in Pr_1^{-1}(\{\phi\})} I(\phi,\eta) = \inf_{\eta\in L_\infty([0,T])} I(\phi,\eta). \tag{2.106}$$

We can now state the LDP for the input process $(Q_n^+(t))_{t\in[0,T]}$.

Proposition A.7 (LDP for the input process).
The family of probability measures $(\mu_n)_{n\in\mathbb{N}}$ satisfies the LDP on $L_\infty[0,T]$ with rate n and with good rate function $\hat{I}$ given by

$$\hat{I}_Q(\Gamma) = \inf_{\phi\in\Gamma} \tilde{I}_Q(\phi). \tag{2.107}$$

In particular, if Γ is $\hat{I}_Q$-continuous, i.e., $\hat{I}_Q(\Gamma) = \hat{I}_Q(\mathrm{int}(\Gamma)) = \hat{I}_Q(\mathrm{cl}(\Gamma))$, then

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(Q_n^+([0,T]) \in \Gamma\big) = -\hat{I}_Q(\Gamma). \tag{2.108}$$

Proof. The claim follows from the contraction principle (see [40]). □

It is interesting to look at a specific subset of $L_\infty([0,T])$ that gives good bounds for the input process. We are now in a position to prove Proposition 2.2.1.

Proof. If we compute the Fenchel-Legendre transforms Λ_Q^* and Λ_N^*, and we pick $\eta(t) = \lambda t$ and $\phi(t) = (1/\mu)\eta(t) = (1/\mu)\lambda t$, we can easily check that the rate function attains its minimal value zero. Hence, with high probability as $r \to \infty$, the input process is close to this deterministic path.

We can now estimate the probability of the scaled input process to go outside $\Gamma_{T,\delta}$, which represents a tube of width 2δ around the mean path in the interval $[0,T]$. More precisely,

$$\Gamma_{T,\delta} = \left\{ \gamma \in L_\infty([0,T]) \colon \frac{\lambda}{\mu} t - \delta < \gamma(t) < \frac{\lambda}{\mu} t + \delta \ \ \forall\, t \in [0,T] \right\}. \tag{2.109}$$

We may set $T = 1$ for simplicity and look at the scaled input process in the time interval $[0,1]$. We have

$$\hat{I}_Q((\Gamma_{1,\delta})^c) = \hat{I}_Q(\mathrm{int}((\Gamma_{1,\delta}))^c) = \hat{I}_Q(\mathrm{cl}((\Gamma_{1,\delta}))^c). \tag{2.110}$$

Hence $(\Gamma_{1,\delta})^c$ is $\hat{I}_Q$-continuous, and so according to (2.108),

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\big(Q_n^+([0,1]) \notin \Gamma_{1,\delta}\big) = -\hat{I}_Q((\Gamma_{1,\delta})^c). \tag{2.111}$$

Since

$$\begin{aligned}
&\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}\big(Q_n^+([0,1])\notin\Gamma_{1,\delta}\big)\\
&=\lim_{n\to\infty}\frac{1}{n}\log\mathbb{P}\bigg(\bigg\{\frac{\lambda}{\mu}t-\delta<Q_n^+(t)<\frac{\lambda}{\mu}t+\delta\ \ \forall\, t\in[0,1]\bigg\}^c\bigg)\\
&=\lim_{S\to\infty}\frac{1}{S}\log\mathbb{P}\bigg(\bigg\{\frac{\lambda}{\mu}s-\delta S<Q^+(s)<\frac{\lambda}{\mu}s+\delta S\ \ \forall\, s\in[0,S]\bigg\}^c\bigg),
\end{aligned}\tag{2.112}$$

where we put $s=nt$ and $S=n$, we conclude that the probability to go out of $\Gamma_{S,\delta S}$ is

$$\mathbb{P}\bigg(\bigg\{\frac{\lambda}{\mu}s-\delta S<Q^+(s)<\frac{\lambda}{\mu}s+\delta S\ \ \forall\, s\in[0,S]\bigg\}^c\bigg)=e^{-S\,\hat{I}_Q((\Gamma_{1,\delta})^c)\,[1+o(1)]},\qquad S\to\infty.\tag{2.113}$$

Because I_Q is convex, to compute $\hat{I}_Q((\Gamma_{1,\delta})^c)$ it suffices to minimise over the linear paths. The minimizer turns out to be one of the two linear paths that go from the origin $(0,0)$ to $(1,\lambda/\mu\pm\delta)$, i.e., $\gamma^*(t)=kt$ with $k=(\lambda\pm\delta\mu)/\mu$. By construction, $\hat{I}_Q((\Gamma_{1,\delta})^c)=\tilde{I}_Q(\gamma^*)=\inf_{\eta\in L_\infty([0,1])}I(\gamma^*,\eta)$, where

$$I(\gamma^*,\eta)=\int_0^1\Lambda_Q^*\left(\frac{d\gamma^*(t)}{d\eta(t)}\right)d\eta(t)+\int_0^1\Lambda_N^*(\dot\eta(t))\,dt.\tag{2.114}$$

We want to minimize the sum over all paths η such that $\eta(0)=0$. Both integrals are convex as a function of γ^* and η, hence they are minimized by linear paths. Our choice of $\gamma^*(t)=kt$ is linear, so we set $\eta(t)=ct$ with some constant $c>0$. We can then write

$$\begin{aligned}
I(\gamma^*,\eta)&=\int_0^{\eta(1)}\Lambda_Q^*\left(\frac{d\gamma^*(t)}{cdt}\right)cdt+\int_0^1\Lambda_N^*(\dot\eta(t))\,dt\\
&=\int_0^c\Lambda_Q^*\left(\frac{k}{c}\right)cdt+\int_0^1\Lambda_N^*(c)\,dt\\
&=c\left[\frac{k}{c}\mu-1-\log\left(\frac{k\mu}{c}\right)\right]+c\log\left(\frac{c}{\lambda}\right)-c+\lambda.
\end{aligned}\tag{2.115}$$

The value of c that minimizes the right-hand side is $c=\sqrt{\lambda k\mu}$. Substituting this into the formula above, we get

$$K_\delta=\tilde{I}_Q(\gamma^*)=k\mu+\lambda-2\sqrt{\lambda k\mu}=(\lambda+\delta\mu)+\lambda-2\sqrt{\lambda(\lambda+\delta\mu)}.\tag{2.116}$$

Note that $K_\delta>0$ for all $\delta>0$ and $\lim_{\delta\to0}K_\delta=0$. This completes the proof. □

§B Appendix: the output process

The main goal of this appendix is to prove Proposition 2.2.4 in Section 2.2. In Section B.1 we show a lower bound for the output process for the nodes in U, in a setting where the nodes in U are not influenced by the nodes in V. We study the network evolution up to time T_U. In Section B.2 we show that, until the pre-transition time, the network in the internal model behaves actually as we described.

§B.1 The output process in the isolated model

Recall that in the isolated model a node in U keeps activating and deactivating independently of the nodes in V, until its queue length hits zero. We again consider a single queue for a node in U and for simplicity suppress its index. In order to show that, with high probability as $r \to \infty$, the output process $t \mapsto Q^-(t) = cT(t)$ when properly rescaled is close to a deterministic path, we will provide a lower bound for the output process. The upper bound $Q^-(t) \leq ct$ is trivial and holds for any $t \geq 0$, by the definition of output process.

Lemma B.1 (Auxiliary output process).
For all $\delta > 0$ and T large, the following statements hold.

(i) With high probability as $r \to \infty$, the process

$$Q^{\mathrm{LB},T}(t) = \gamma_U r + \rho_U t - \delta T - ct, \qquad t \in [0,T], \tag{2.117}$$

is a lower bound for the actual queue length process $(Q(t))_{t\in[0,T]}$.

(ii) The probability of the lower bound in (i) failing is

$$\frac{1}{2} e^{-K_\delta T\,[1+o(1)]}, \qquad T \to \infty, \tag{2.118}$$

with $K_\delta = (\lambda + \delta\mu) + \lambda - 2\sqrt{\lambda(\lambda+\delta\mu)}$.

Proof. We prove the two statements separately.

(i) By Proposition 2.2.1, with high probability as $r \to \infty$, we have $Q^+(t) \geq \rho_U t - \delta T$ for any $\delta > 0$. Trivially, $Q^-(t) \leq ct$. It is therefore immediate that, with high probability as $r \to \infty$, $Q^{\mathrm{LB},T}(t) \leq Q(t)$.

(ii) The exponentially small probability of $Q^+(t)$ going below the lower bound is half of the probability given by Proposition 2.2.1, i.e.,

$$\frac{1}{2} e^{-K_\delta T\,[1+o(1)]}, \qquad T \to \infty, \tag{2.119}$$

with $K_\delta = (\lambda + \delta\mu) + \lambda - 2\sqrt{\lambda(\lambda+\delta\mu)}$.

□

We study the network evolution up to time T_U defined in Definition 2.2.2, the expected time a single node queue takes to hit zero. We will prove in Appendix B.2 that, with high probability as $r \to \infty$, the pre-transition time in the internal model coincides in distribution with the pre-transition time in the isolated model, which occurs before T_U. Hence it is enough to study the isolated model up to T_U.

Definition B.2 (Auxiliary times).
We next define two times that will be useful in our analysis.

(T_U^*) Consider the auxiliary output process $Q^{\mathrm{LB},T_U}(t)$ up to time T_U. We define T_U^* as the time needed for the process to hit zero, i.e.,

$$T_U^* = T_U^*(r) = \frac{\gamma_u r - \delta T_U}{c - \rho_U} = \frac{\gamma_u - \delta\alpha}{c - \rho_U} r = \alpha' r, \tag{2.120}$$

with $\alpha' = \frac{\gamma_u - \delta\alpha}{c-\rho_U}$. The difference $T_U - T_U^* = \frac{\delta\alpha}{c-\rho_U} r$ is of order r. The queue length at time T_U^* is not zero, but still of order r.

(T_U^{**}) We define a smaller time T_U^{**} in such a way that, not only $Q(T_U^{**}) \asymp r$, but also $Q^{\mathrm{LB},T_U}(T_U^{**}) \asymp r$, i.e.,

$$T_U^{**} = T_U^{**}(r) = T_U - 2(T_U - T_U^*) = \left(\frac{\gamma_U - 2\delta\alpha}{c - \rho_U}\right) r = \alpha'' r, \tag{2.121}$$

with $\alpha'' = \frac{\gamma_U - 2\delta\alpha}{c-\rho_U}$.

Definition B.3 (Inactivity process).
Define the *inactivity process* by setting

$$W(t) = t - T(t), \tag{2.122}$$

which equals the total amount of inactivity time until time t.

Recall that the service process $t \mapsto Q^-(t)$ with $Q^-(0) = 0$ is an alternating sequence of activity periods and inactivity periods. The activity periods Z_i, $i \in \mathbb{N}$, are i.i.d. exponential random variables with mean 1. The inactivity periods W_m, $m \in \mathbb{N}$, are exponential random variables with a mean that depends on the actual queue length at the time when each of these periods starts, namely, if $W_m = \left[t_m^{(i)}, t_m^{(f)}\right]$, then $W_m \simeq \mathrm{Exp}(g_U(Q(t_m^{(i)})) + O(1/r))$. The queue length during this inactivity intervals is actually increasing, but we are considering very small intervals, whose lengths are of order $1/r$, so that the queue length does not change much and the error is then $O(1/r)$.

To state our lower bound on the output process, we need the following two lemmas.

Lemma B.4 (Upper bound on number of activity periods).
Let $M(t)$ be the number of activity periods that end before time t. Then, for all $\epsilon_1 > 0$, the following statements hold.

(i) With high probability as $r \to \infty$,

$$M(T_U^{**}) \le (1 + \epsilon_1) T_U^{**}. \tag{2.123}$$

(ii) The probability of the upper bound in (i) failing is

$$\mathbb{P}_u(M(T_U^{**}) > (1 + \epsilon_1) T_U^{**}) \le e^{-K_1 r\,[1+o(1)]} + \frac{1}{2} e^{-K_\delta \alpha r\,[1+o(1)]}, \qquad r \to \infty, \tag{2.124}$$

with $K_1 = \alpha'' \frac{\epsilon_1 - \log(1+\epsilon_1)}{1+\epsilon_1}$, K_δ as in Lemma B.1

Proof. We prove the two statements separately.

(i) Note that $M(T_U^{**})$ counts the number of activity periods before time T_U^{**}, each of which has an average duration 1. Since activity periods alternate with inactivity periods, we expect $M(T_U^{**})$ to be less than T_U^{**}. Assume now, for small $\epsilon_1 > 0$, that $M(T_U^{**}) > (1+\epsilon_1)T_U^{**}$, which means that the number of activity periods before T_U^{**} is greater than the length of the interval $[0, T_U^{**}]$. This implies that the average length of each activity period before time T_U^{**} is strictly less than 1, namely, that $\frac{1}{T_U^{**}} \sum_{i=1}^{T_U^{**}} Z_i \leq 1/(1+\epsilon_1)$. According to Cramér's theorem, we can compute the probability of this last event as

$$\mathbb{P}_u \left(\sum_{i=1}^{T_U^{**}} Z_i \leq \left(\frac{1}{1+\epsilon_1} \right) T_U^{**} \right) = e^{-T_U^{**} I\left(\frac{1}{1+\epsilon_1}\right)[1+o(1)]}, \qquad r \to \infty, \tag{2.125}$$

with rate function $I(x) = x\log(x) - x + 1$. Therefore, it occurs with exponentially small probability. Hence $M(T_U^{**}) > (1+\epsilon_1)T_U^{**}$ must also occur with a probability which is also exponentially small. With high probability as $r \to \infty$, we then have that

$$M(T_U^{**}) \leq (1+\epsilon_1)T_U^{**}. \tag{2.126}$$

Recall that $T_U^{**} = \alpha'' r$. The counting of alternating activity and inactivity periods gets affected when the queue length hits zero, since then the node deactivates and the lengths of the activity periods are not regular anymore. At time T_U^{**}, with high probability as $r \to \infty$, the queue length is still of order r. Hence, the probability that it hits zero at any time in the interval $[0, T_U^{**}]$ is very small, since this event would imply the node to have a queue length that is below the lower bound, $Q(T_U^{**}) \leq Q^{\mathrm{LB},T_U}(T_U^{**})$, which happens with an exponentially small probability by Lemma B.1.

(ii) We can write

$$\begin{aligned} \mathbb{P}_u(M(T_U^{**}) > (1+\epsilon_1)T_U^{**}) &\leq e^{-T_U^{**} I\left(\frac{1}{1+\epsilon_1}\right)[1+o(1)]} + \frac{1}{2} e^{-K_\delta T_U [1+o(1)]} \\ &= e^{-K_1 r [1+o(1)]} + \frac{1}{2} e^{-K_\delta \alpha r [1+o(1)]}, \qquad r \to \infty, \end{aligned} \tag{2.127}$$

with $K_1 = \alpha'' I\left(\frac{1}{1+\epsilon_1}\right) = \alpha'' \frac{\epsilon_1 - \log(1+\epsilon_1)}{1+\epsilon_1}$, K_δ as in Lemma B.1.

□

Lemma B.5 (Upper bound on inactivity process).
For all $\delta, \epsilon_1, \epsilon_2 > 0$ small, the following statements hold.

(i) With high probability as $r \to \infty$,

$$W(T_U^{**}) \leq \epsilon_2 r. \tag{2.128}$$

(ii) The probability of the upper bound in (i) failing is

$$\mathbb{P}_u\big(W(T_U^{**}) > \epsilon_2 r\big) \le e^{-K_\delta \alpha r\,[1+o(1)]} + e^{-K_1 r\,[1+o(1)]} + e^{-\left(K_2 r + K_3 \frac{r}{g_U(r)} + K_4 r \log g_U(r)\right)[1+o(1)]}, \qquad r \to \infty, \tag{2.129}$$

with $K_2 = \alpha''(1+\epsilon_1)\big(-1-\log\big(\frac{\epsilon_2}{\alpha''(1+\epsilon_1)}\big)\big), K_3 = \epsilon_2, K_4 = \alpha''(1+\epsilon_1)$.

Proof. We prove the two statements separately.

(i) Since $M(t)$ counts the number of activity periods, and we start with an active node (initially all nodes in U are active), we have

$$W(T_U^{**}) \le \sum_{m=1}^{M(T_U^{**})} W_m \le \sum_{m=1}^{M(T_U^{**})} \hat{W}_m, \tag{2.130}$$

where $\hat{W}_m$ are i.i.d. exponential random variables with rate $g_U(Q^{\mathrm{LB},T_U}(T_U^{***}))$, and T_U^{***} is the starting point of the last inactivity period before time T_U^{**}. By the construction of T_U^{**}, we know that $Q^{\mathrm{LB},T_U}(T_U^{***})$ is of order r. The last inactivity period is expected to be longer than the previous ones, since the activation rates depend on the actual queue length, which is decreasing over time. To make the inactivity periods $\hat{W}_m$ longer, we consider the lower bound $Q^{\mathrm{LB},T_U}(t)$ for the actual queue length given in Lemma B.1.

By Lemma B.4, with high probability as $r \to \infty$, $M(T_U^{**}) \le (1+\epsilon_1)T_U^{**}$, and so

$$W(T_U^{**}) \le \sum_{m=1}^{M(T_U^{**})} \hat{W}_m \le \sum_{m=1}^{(1+\epsilon_1)T_U^{**}} \hat{W}_m. \tag{2.131}$$

Define $n = [(1+\epsilon_1)T_U^{**}]$. By Cramér's theorem, for small $\epsilon_3 > 0$,

$$\begin{aligned}\mathbb{P}_u\left(\sum_{m=1}^{(1+\epsilon_1)T_U^{**}} \hat{W}_m \ge \epsilon_3 T_U^{**}\right) &\le \mathbb{P}_u\left(\sum_{m=1}^{n} \hat{W}_m \ge \frac{\epsilon_3}{1+\epsilon_1} n\right) \\ &= e^{-nI\left(\frac{\epsilon_3}{1+\epsilon_1}\right)[1+o(1)]} \\ &= e^{-T_U^{**}(1+\epsilon_1)I\left(\frac{\epsilon_3}{1+\epsilon_1}\right)[1+o(1)]}, \qquad n \to \infty,\end{aligned} \tag{2.132}$$

where I is the rate function given by

$$I(x) = \frac{x}{g_U\big(Q^{\mathrm{LB},T_U}(T_U^{***})\big)} - 1 - \log x + \log g_U\big(Q^{\mathrm{LB},T_U}(T_U^{***})\big). \tag{2.133}$$

We take $\epsilon_3 > (1+\epsilon_1)/g_U\big(Q^{\mathrm{LB},T_U}(T_U^{***})\big) \asymp 1/g_U(r)$ arbitrarily small, so that we can apply Cramér's theorem. Combining (2.131)–(2.132), we obtain that, with high probability as $r \to \infty$,

$$W(T_U^{**}) \le \epsilon_3 T_U^{**} = \epsilon_3 \alpha'' r = \epsilon_2 r, \tag{2.134}$$

where $\epsilon_2 = \epsilon_3 \alpha''$ can be taken arbitrarily small.

(ii) We can write

$$\begin{aligned}
&\mathbb{P}_u\left(\sum_{m=1}^{(1+\epsilon_1)T_U^{**}} \hat{W}_m > \epsilon_3 T_U^{**}\right) \\
&\leq e^{-T_U^{**}(1+\epsilon_1)I\left(\frac{\epsilon_3}{1+\epsilon_1}\right)[1+o(1)]} \\
&= e^{-\alpha'' r(1+\epsilon_1)\left(\frac{\epsilon_3}{(1+\epsilon_1)g_U(r)} - 1 - \log\left(\frac{\epsilon_3}{1+\epsilon_1}\right) + \log g_U(r)\right)[1+o(1)]} \\
&= e^{-\left[\alpha''(1+\epsilon_1)\left(-1-\log\left(\frac{\epsilon_3}{1+\epsilon_1}\right)\right)r + \epsilon_3\alpha''\frac{r}{g_U(r)} + \alpha''(1+\epsilon_1)r\log(g_U(r))\right][1+o(1)]} \\
&= e^{-\left(K_2 r + K_3\frac{r}{g_U(r)} + K_4 r\log g_U(r)\right)[1+o(1)]}, \qquad r\to\infty,
\end{aligned} \tag{2.135}$$

where the constants $K_2 = \alpha''(1+\epsilon_1)\big(-1-\log\big(\frac{\epsilon_2}{\alpha''(1+\epsilon_1)}\big)\big), K_3 = \epsilon_3\alpha'' = \epsilon_2$ and $K_4 = \alpha''(1+\epsilon_1)$. We also have to consider the probabilities computed in (2.118) and (2.124), and the claim in (2.129) is settled.

□

We are now in a position to prove Proposition 2.2.4.

Proof. The equation $Q^-(t) \geq ct - \epsilon r$ can be read as $T(t) \geq t - \epsilon r/c$. This is equivalent to saying that $W(t) \leq \epsilon r/c$ for all $t \in [0, T_U]$. By taking $\epsilon_2 = \epsilon/(3c)$ in Lemma B.5, we know that, for all $t \in [0, T_U^{**}]$, $W(t) \leq W(T_U^{**}) \leq \epsilon r/(3c)$. Moreover, in the interval $[T_U^{**}, T_U]$, the cumulative amount of inactivity time is trivially bounded from above by the length of the interval, which is $\frac{2\delta r}{c-\rho_U} \leq 2\epsilon r/(3c)$, and ϵ can be taken arbitrarily small, since δ can be taken arbitrarily small. Putting the two bounds together, we find that, with high probability as $r \to \infty$,

$$W(t) \leq \epsilon_2 r + \frac{2\delta r}{c-\rho_U} \leq \frac{1}{3}\frac{\epsilon r}{c} + \frac{2}{3}\frac{\epsilon r}{c} = \frac{\epsilon r}{c}, \qquad t \in [0, T_U], \tag{2.136}$$

and the probability of this not happening is given by (2.129). □

The above lower bound $Q^-(t) \geq ct - \epsilon r$ and the trivial upper bound $Q^-(t) \leq ct$ imply that, with high probability as $r \to \infty$, the output process $Q^-(t)$ stays close to the path $c \mapsto ct$ by sending ϵ to zero. In other words, the node stays almost always active all the time before T_U.

§B.2 The output process in the internal model

In this section we want to couple the isolated model and the internal model and show that they have identical behavior in the time interval $[0, \tau_G^{\text{int}}]$. Hence it follows that the output process in the internal model for nodes in U actually behaves as in the isolated model described in Section B.1, until the pre-transition time.

Proposition B.6 (Coupling the internal and the isolated model).
Let $X_i^{\text{int}}(t)$ and $X_i^{\text{iso}}(t)$ denote the activity state of a node i at time t in the internal and the isolated model, respectively. Then

$$\lim_{r\to\infty} \mathbb{P}_u\big(X_i^{\text{int}}(t) = X_i^{\text{iso}}(t) \ \forall\, i \in U \sqcup V \ \forall\, t \in [0, \tau_G^{\text{int}}]\big) = 1. \tag{2.137}$$

Consequently, with high probability as $r \to \infty$, the pre-transition times in the internal and the isolated model coincide, i.e.,

$$\lim_{r\to\infty} \mathbb{P}_u(\tau_G^{\text{int}} = \tau_G^{\text{iso}}) = 1. \tag{2.138}$$

Proof. In Section B.1 we determined upper and lower bounds for the output process for nodes in U in the isolated model up to time T_U. Assume now that $\tau_G^{\text{int}} \leq T_U$. When considering the internal model and the set of nodes in V, note that these bounds are not true for the whole interval $[0, T_U]$, since at time τ_G^{int} some nodes in V already start to activate and influence the behavior of nodes in U.

If we look at the interval $[0, \tau_G^{\text{int}}]$, then we note that the queue length process for a node $i \in U$ is not affected by nodes in V, and so it behaves in exactly the same way as if the node were isolated. The activation and deactivation Poisson clocks at node i are synchronized, and are ticking at the same time in the isolated model and in the internal model, so that $X_i^{\text{int}}(t) = X_i^{\text{iso}}(t)$. Moreover, the activity states of nodes in V are always equal to 0 in both models. Hence we conclude that the activity states of every node coincide up to the pre-transition time τ_G^{int}. Consequently, the pre-transition times in the internal and the isolated model coincide on the event $\{\tau_G^{\text{int}} \leq T_U\}$, which can then be written as the event $\{\tau_G^{\text{iso}} \leq T_U\}$. For the latter we know that it has a high probability as $r \to \infty$ (see proof of Proposition 2.3.8 in Section 2.3.2). □

CHAPTER 3

Arbitrary bipartite interference graphs

This chapter is based on:
S.C. Borst, F. den Hollander, F.R. Nardi, M. Sfragara. *Wireless random-access networks with bipartite interference graphs.* [arXiv:2001.02841], 2020.

Abstract

We consider random-access networks where each node represents a server with a queue. Each node can be either active or inactive. A node deactivates at unit rate, and activates a rate that depends on its queue length, provided none of its neighbors is active. We consider arbitrary bipartite graphs in the limit as the queues become large, and we identify the transition time between the two states where one half of the network is active and the other half is inactive. We decompose the transition into a succession of transitions on complete bipartite subgraphs, and formulate a greedy algorithm that takes the graph as input and gives as output the set of transition paths the system is most likely to follow. Along each path we determine the mean transition time and its law on the scale of its mean. Depending on the activation rate functions, we identify three regimes of behavior.

§3.1 Introduction

This chapter is a continuation of Chapter 2. We turn our attention to general bipartite interference graphs: the node set can be partitioned into two nonempty sets U and V, but not necessarily all nodes in U interfere with all nodes in V.

In Section 3.1.1 we describe the setting and the mathematical model of interest in this chapter. In Section 3.1.2 we introduce the key idea behind our main results and give an outline of the remainder of the chapter.

§3.1.1 Setting

We refer to Sections 1.1.5 and 2.1.1 for a general introduction to the mathematical model. In this section we refine it with some extra notions we will need in the chapter.

Consider the bipartite graph $G = ((U, V), E)$, where $U \sqcup V$ is the set of nodes and E is the set of (undirected) edges that connect a node in U to a node in V, and vice versa (see Figure 3.1 for examples). Through the chapter we assume $|V| = N$.

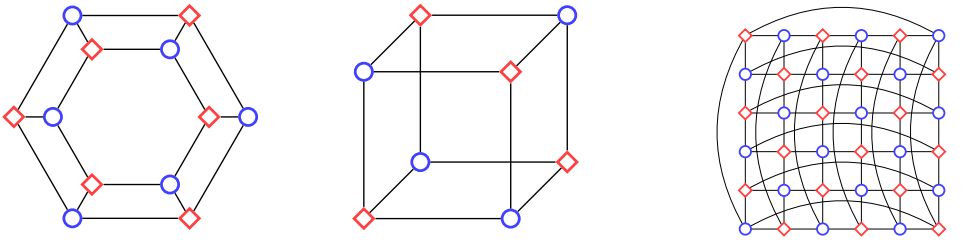


Figure 3.1: Examples of bipartite graphs: cyclic ladder (left), hypercube (center), even torus (right).

We study the internal model with activation rates depending on the queue lengths as in (1.5). We assume the activation rates to satisfy Definition 1.1.4 and we focus on the following.

Definition 3.1.1 (Assumptions on the activation rates).
We assume *polynomial activation functions* of the form

$$\begin{aligned} g_U(x) &\sim Bx^{\beta}, \quad x \to \infty, \\ g_V(x) &\sim B'x^{\beta'}, \quad x \to \infty, \end{aligned} \tag{3.1}$$

with $B, B', \beta, \beta' \in (0, \infty)$. We assume that nodes in V are *much more aggressive* than nodes in U, namely,

$$\beta' > \beta + 1. \tag{3.2}$$

As we will see later, this ensures that the transition from u to v can be decomposed into a succession of transitions on complete bipartite subgraphs.

We begin by recalling the results for complete bipartite graphs from Chapter 2 (Theorem 2.1.6). Note that they are strongly related to the initial queue lengths at the nodes in U, which are assumed in (1.11) to be $Q_U(0) = \gamma_U r$.

Theorem (Theorem 2.1.6).
Let G be a complete bipartite graph.

(I) $\beta \in (0, \frac{1}{|U|-1})$*: subcritical regime. The transition time satisfies*

$$\mathbb{E}_u[\mathcal{T}_G] = F_{\text{sub}}\, Q_U(0)^{\beta(|U|-1)}\,[1 + o(1)], \qquad r \to \infty, \tag{3.3}$$

with $F_{\text{sub}} = \frac{1}{|U|B^{-(|U|-1)}}$*, and*

$$\lim_{r\to\infty} \mathbb{P}_u\left(\frac{\mathcal{T}_G}{\mathbb{E}_u[\mathcal{T}_G]} > x\right) = \int_x^\infty \mathcal{P}_{\text{sub}}(y)\, dy = e^{-x}, \qquad x \in [0,\infty) \tag{3.4}$$

with

$$\mathcal{P}_{\text{sub}}(z) = e^{-z}, \qquad z \in [0,\infty). \tag{3.5}$$

(II) $\beta = \frac{1}{|U|-1}$*: critical regime. The transition time satisfies*

$$\mathbb{E}_u[\mathcal{T}_G] = F_{\text{cr}}\, Q_U(0)\,[1 + o(1)], \qquad r \to \infty, \tag{3.6}$$

with $F_{\text{cr}} = \frac{1}{|U|B^{-(|U|-1)}+(c-\rho_U)}$*, and*

$$\begin{aligned}\lim_{r\to\infty} \mathbb{P}_u\left(\frac{\mathcal{T}_G}{\mathbb{E}_u[\mathcal{T}_G]} > x\right) &= \int_x^\infty \mathcal{P}_{\text{cr}}(y)\, dy \\ &= \begin{cases} (1-Cx)^{\frac{1-C}{C}}, & \text{if } x \in [0, \frac{1}{C}), \\ 0, & \text{if } x \in [\frac{1}{C}, \infty), \end{cases}\end{aligned} \tag{3.7}$$

with

$$\mathcal{P}_{\text{cr}}(z) = \begin{cases} (1-C)(1-Cz)^{\frac{1}{C}-2}, & \text{if } z \in [0, \frac{1}{C}), \\ 0, & \text{if } z \in [\frac{1}{C}, \infty), \end{cases} \tag{3.8}$$

and $C = F_{\text{cr}}\,(c - \rho_U) \in (0,1)$.

(III) $\beta \in (\frac{1}{|U|-1}, \infty)$*: supercritical regime. The transition time satisfies*

$$\mathbb{E}_u[\mathcal{T}_G] = F_{\text{sup}}\, Q_U(0)\,[1 + o(1)], \qquad r \to \infty, \tag{3.9}$$

with $F_{\text{sup}} = \frac{1}{c-\rho_U}$*, and*

$$\lim_{r\to\infty} \mathbb{P}_u\left(\frac{\tau_{1_V}}{\mathbb{E}_u[\mathcal{T}_G]} > x\right) = \int_x^\infty \mathcal{P}_{\text{sup}}(y)\, dy = \begin{cases} 1, & \text{if } x \in [0,1), \\ 0, & \text{if } x \in [1,\infty), \end{cases} \tag{3.10}$$

with

$$\mathcal{P}_{\text{sup}}(z) = \delta_1(z), \qquad z \in [0,\infty), \tag{3.11}$$

where $\delta_1(z)$ is the Dirac function at 1.

Theorem 2.1.6 shows that there is a trichotomy: depending on the value of β the transition exhibits a subcritical regime, a critical regime or a supercritical regime. Our goal is to extend Theorem 2.1.6 to arbitrary bipartite graphs. Note how the mean transition time depends on the actual value of the initial queue lengths at nodes in U: for complete bipartite graphs, those are fixed and equal to $\gamma_U r$; for arbitrary bipartite graphs, we will see how the mean transition time depends on the way the queue lengths are changing when nodes in V activate.

Next, we define some key notions that we will need in the chapter.

Definition 3.1.2 (Fork).
For a node $v \in V$, we define the set of neighbors of v as $N(v) = \{u \in U\colon uv \in E\}$ and the degree of v as $d(v) = |N(v)|$. Given a node $v \in V$, we refer to *fork* of v as the complete bipartite subgraph of G containing only node v, its neighbors $N(v) \subseteq U$ and the edges between them. We talk about a d-fork when $d(v) = d$ with $d \in \mathbb{N}$.

Definition 3.1.3 (Updated queue lengths).
Let $Q_U = \{Q_{U,i}\}_{i=1}^{|U|}$ be the sequence of queues associated with the nodes in U, and $Q_V = \{Q_{V,j}\}_{j=1}^{|V|}$ the sequence of queues associated with the nodes in V. Put $Q = (Q_U, Q_V)$, and let $Q^k = (Q_U^k, Q_V^k)$ be the pair of sequences representing the *updated queue lengths* after k nodes in V activated (see Definition 3.2.9 later for more details).

We denote by $\mathcal{T}_G^Q$ the transition time of the graph G when the initial queue lengths are $Q = (Q_U, Q_V)$. It represents the time it takes to reach v starting from u. Below, we define the nucleation time in order to distinguish between the full transition of G and the successive transitions (nucleations) of the subgraphs of G related to each node activating in V.

Definition 3.1.4 (Nucleation time).
We call *nucleation time* of the fork of v the time it takes for the nodes $N(v)$ to deactivate and for v to activate. We denote this time by $\mathcal{T}_v^Q = \mathcal{T}_{N(v),v}^Q$, where v represents the activating node and Q represents the initial queue lengths. It can be seen as the transition time of the complete bipartite subgraph of G represented by the fork of v. Note that, for $v, w \in V$, $\mathcal{T}_v^Q$ and $\mathcal{T}_w^Q$ are dependent random variables when $N(v) \cap N(w) \neq \emptyset$.

§3.1.2 Key idea and outline

The key idea behind this chapter is to define an algorithm that allows us to identify the set of paths $\mathcal{A}$ the network is mostly likely to follow while nodes in U deactivate and nodes in V activate. We label the nodes in V based on their first activation and we denote by a^* the path that the network follows. More precisely, $a^* = (v_1^*, \dots, v_N^*)$ with $v_1^*, \dots, v_N^*$ all distinct, where v_1^* is the first node that activates and v_N^* the last one. Let $\mathcal{E}(a^*)$ denote the event that any of the paths in $\mathcal{A}$ occurs. We will prove that

$$\lim_{r\to\infty} \mathbb{P}_u(\mathcal{E}(a^*)) = 1. \tag{3.12}$$

In particular, we will show that if we condition on the event

$$A_a = \{\text{the network follows path } a \in \mathcal{A}\}, \tag{3.13}$$

then we are able to identify how the mean transition time $\mathbb{E}_u[\mathcal{T}_G^Q \mid A_a]$ depends on the sequence of nucleation times of the forks of the nodes in V, ordered as in the path a (Theorem 3.3.2). We derive the asymptotics of the mean transition time as $r \to \infty$ (Theorem 3.3.3) and identify the law of the transition time divided by its mean (Theorem 3.3.5). To do so, we determine how the queue lengths change along the given path (Theorem 3.4.8). Similarly as for the complete bipartite graph in Theorem 2.1.6, we distinguish between three regimes for the value of β (subcritical, critical and supercritical), in which the queues behave differently and, consequently, so does the transition time.

Outline of the chapter. The remainder of this chapter is organized as follows. In Section 3.2 we introduce the algorithm, show that it has two important properties, *greediness* and *consistency*, and give an example of how it works. In Section 3.3 we state our main theorems. In particular, we show how both the mean transition time and its law on the scale of its mean can be determined according to the path that the algorithm chooses. In Section 3.4 we show how the nucleation times depend on the graph structure and we analyze how the queue lengths at the nodes change along each path that the algorithm chooses. In Section 3.5 we provide the proof of the two algorithm properties mentioned above and we discuss the algorithm complexity. In Section 3.6 we prove our main theorems. In Appendix C, we show some technical computations for the mean nucleation time in the special setting of disjoint forks competing for activation.

§3.2 The algorithm

In this section we introduce the algorithm that describes, step by step, how the network behaves while nodes in U deactivate and nodes in V activate. The presentation is organised into a series of definitions and lemmas. In Section 3.2.1 we define how the algorithm works iteratively. In Section 3.2.2 we show that the algorithm is greedy and consistent (Propositions 3.2.6–3.2.7). In Section 3.2.3 we explain how the algorithm is used to capture the nucleation of the forks. An example of a bipartite graph and how the algorithm acts on it are given in Section 3.2.4.

§3.2.1 Definition of the algorithm

Let $N = |V|$ be the number of nodes in V. The algorithm takes as input the bipartite graph $G = ((U, V), E)$ and gives as output a sequence of triples that is needed to characterise the transition time, namely,

$$G \to (Y_k, \bar{d}_k, n_k)_{k=1}^N, \tag{3.14}$$

where Y_k is a random variable with values in $\{1, \dots, N\}$ describing the index of the node selected at step k, $\bar{d}_k \in \mathbb{N}$ is the degree of the selected node and $n_k \in \mathbb{N}$ is

a parameter that counts how many possibilities there are at step k to choose the next node in V (uniformly at random) from the remaining nodes with least degree. Sometimes we will write v_k^* instead of v_{Y_k} to emphasise that the network is following a specific order of activation for the nodes in V.

Definition 3.2.1 (Algorithm).
Set $G = G_1 = ((U_1, V_1), E_1)$. Given the graph $G_k = ((U_k, V_k), E_k)$, find the graph $G_{k+1} = ((U_{k+1}, V_{k+1}), E_{k+1})$ by iterating the following procedure until V_{k+1} is empty.

- Start from the graph G_k.
- Look at the nodes in V_k and at the *minimum degree* $\bar{d}_k$ in G_k.
- Pick a node uniformly at random from the ones with minimum degree in G_k.
- Denote the chosen node by v_k^* and the *number of choices* by n_k.
- Eliminate the node v_k^* and all its neighbors in U_k, together with all their adjacent edges. Denote the resulting bipartite graph by G_{k+1}.

The idea of eliminating step by step the nodes in U that deactivated comes from the fact that when a node in V activates, it "blocks" all its neighbors in U, which, with high probability as $r \to \infty$, will remain inactive for the rest of the time. This is due to the aggressiveness of the nodes in V compared to the nodes in U (recall Definition 3.1.1). The following lemma will be proved in Section 3.6.2.

Lemma 3.2.2 (Activation sticks).
Consider a node $u \in U$ and let $N(u) \subseteq V$ be the set of neighbors of u. Denote by t_u the first time a node $v \in N(u)$ activates. Then, with high probability as $r \to \infty$, u remains inactive after t_u, i.e., $X_u(t) = 0$ for all $t \geq t_u$.

Definition 3.2.3 (Mean nucleation time for the algorithm).
The algorithm generates a sequence $v_1^*, \ldots, v_N^*$ of successively activating nodes in V. Associated with step k of the algorithm is the nucleation time of the fork of node v_k^* (see Definition 3.1.2), which according to Theorem 2.1.6 satisfies

$$\mathbb{E}_u[\mathcal{T}_{v_k^*}^{Q^{k-1}}] = F^k \left(\mathbb{E}_u[Q_U^{k-1}]\right)^{1 \wedge \beta(\bar{d}_k - 1)} [1 + o(1)], \qquad r \to \infty. \tag{3.15}$$

Here F^k is a pre-factor that depends on the degree $\bar{d}_k$, which plays the role of $|U|$ in Theorem 2.1.6, and on its relation with β. The term $\mathbb{E}_u[Q_U^{k-1}]$ represents the mean updated queue lengths at the nodes in U_k in the subgraph G_{k-1} (see Definition 3.1.3), and plays the role of the initial queue lengths in Theorem 2.1.6. Note that Q_U^0 is fixed, while $Q_U^1, Q_U^2, \ldots, Q_U^{N-1}$ are random.

Intuitively, the sum of the mean nucleation times associated with the path generated by the algorithm gives the mean transition time along that path. We will see in Section 3.4.2 that the pre-factors F^k actually need to be adjusted by certain weights that depend on the graph structure.

§3.2.2 Properties of the algorithm

Definition 3.2.4 (Maximum least degree).
Given the sequence $(\bar{d}_k)_{k=1}^N$ generated by the algorithm, let

$$d^* = \max_{1 \le k \le N} \bar{d}_k \tag{3.16}$$

be the *maximum least degree* of the path associated with $(\bar{d}_k)_{k=1}^N$.

Each time we run the algorithm, it may generate a different sequence, since it decides uniformly at random which node in V with minimum degree to pick next. We know that the set of paths $\mathcal{A}$ generated by the algorithm is the set of most likely paths the network follows. The order of the nodes in a path is given by their successive activation in V.

The following lemma and two propositions will be proved in Section 3.5.2.

Lemma 3.2.5 (Comparing maximum least degrees of different paths).
Consider two different paths a, b such that $a \in \mathcal{A}$ is generated by the algorithm. For $k = 1, \dots, N$, denote by $\bar{d}_{k,a}$ and $\bar{d}_{k,b}$ the minimum degrees at step k in paths a and b. Let $d_a^* = \max_{1 \le k \le N} \bar{d}_{k,a}$ and $d_b^* = \max_{1 \le k \le N} \bar{d}_{k,b}$. Then $d_a^* \le d_b^*$.

In other words, given any path b, its maximum least degree cannot be smaller than the maximum least degree of a path a generated by the algorithm. We will see how the maximum least degree d^* determines the order of the mean transition time. Depending on how β is related to d^*, we distinguish between the following three different regimes.

(I) *Sucritical regime*, if $\beta \in (0, \frac{1}{d^*-1})$.

(II) *Critical regime*, if $\beta = \frac{1}{d^*-1}$.

(III) *Supercritical regime*, if $\beta \in (\frac{1}{d^*-1}, \infty)$.

The algorithm is greedy, in the sense that it always chooses the node that adds the least to the total transition time along the path, simply because this node is likely to be the first to activate. The greedy way in which the algorithm picks the nodes ensures that the transition time along the chosen path is the shortest possible.

Proposition 3.2.6 (Greediness).
The mean transition time along a path generated by the algorithm is the shortest possible.

The algorithm is consistent, in the sense that d^* is unique. Different paths generated by the algorithm lead to the same order of the mean transition time.

Proposition 3.2.7 (Consistency).
All the paths generated by the algorithm lead to the same order of the mean transition time.

§3.2.3 Structure of the algorithm

A node in V activates because it is the one whose complete bipartite fork has the fastest nucleation, and occurs because of the randomness in the activation and deactivation Poisson clocks and the randomness of the queue length processes that appear as the arguments of the activation rates.

Definition 3.2.8 (Next nucleation time).
Given that $k-1$ nodes in V already activated, we define by *next nucleation time* $\bar{\tau}_k$ the time it subsequently takes for the k-th node in V to activate, i.e.,

$$\bar{\tau}_k = \min_{v \in V_k} \mathcal{T}_v^{Q^{k-1}}. \tag{3.17}$$

By keeping track of which nodes have been picked, we can compute the updated queue lengths for the successive mean nucleation times.

Definition 3.2.9 (Updated queue lengths).
For $k = 1, \dots, N$, define the *updated queue lengths* Q^{k-1} by

$$Q^{k-1} = (Q_U^{k-1}, Q_V^{k-1}) = \left(Q_U\Big(\sum_{l=1}^{k-1} \bar{\tau}_l \Big), Q_V\Big(\sum_{l=1}^{k-1} \bar{\tau}_l \Big) \right). \tag{3.18}$$

When a node in V activates, its fork can be of three different types depending on how its degree is related to β.

Definition 3.2.10 (Subcritical, critical and supercritical nodes).
Given that $k-1$ nodes in V already activated, consider the k-th activating node and its fork of degree $\bar{d}_k$. If $\beta \in (0, \frac{1}{\bar{d}_k - 1})$, then the node (or its fork) is subcritical. If $\beta = \frac{1}{\bar{d}_k - 1}$, then it is critical. If $\beta \in (\frac{1}{\bar{d}_k - 1}, \infty)$, then it is supercritical.

In the subcritical and critical regimes, with high probability as $r \to \infty$, the next nucleation time $\bar{\tau}_k$ is given by the minimum over the nodes with least degree in V_k. Indeed, with high probability as $r \to \infty$, nodes with least degree activate first. The following lemma will be proved in Section 3.6.2.

Lemma 3.2.11 (Activation selects low degree).
For $k = 1, \dots, N$, consider two nodes $v, w \in V_k$ such that $d_k(w) > d_k(v) = \bar{d}_k$. Suppose that $\beta \in (0, \frac{1}{\bar{d}_k - 1}]$. Then the probability of w activating before v satisfies

$$\lim_{r \to \infty} \mathbb{P}_u\big(\mathcal{T}_w^{Q^{k-1}} < \mathcal{T}_v^{Q^{k-1}} \big) = 0. \tag{3.19}$$

In the supercritical regime the situation is more delicate. If at step k the least degree fork has degree $\bar{d}_k$ such that $\beta \in (\frac{1}{\bar{d}_k - 1}, \infty)$, then the mean nucleation time of the next activating fork is the same for all the remaining forks in the graph. The network does not distinguish between the nodes according to their degree anymore, since all possibilities contribute equally to the total mean transition time. Indeed, the mean nucleation time is given by the expected time it takes for the queue lengths at nodes in U to hit zero. Hence, after the nucleation of the first supercritical fork,

all the queues in U are $o(r)$ and the transition occurs very fast (see Section 3.4.3 for more details).

In Section 3.3 we will see how the transition time can be computed given the set of possible paths generated by the algorithm. Moreover, for each fixed path we will identify the mean transition time and its law on the scale of its mean. Given a path, we know in which order the nodes activate. In Section 3.6 we will see how we can identify the nucleation time of a node given in Definition 3.2.8 with the nucleation time of the complete bipartite fork of the activating node, as written in (3.15). The sum of all the nucleation times gives us the transition time of the graph. Not all the terms in the sum contribute significantly as $r \to \infty$. We will need to identify which are the leading order terms. The answer depends on the sequence of degrees $(\bar{d}_k)_{k=1}^N$ generated by the algorithm and on how the queue lengths change along the path.

§3.2.4 Example

Consider the bipartite graph $G = ((U, V), E)$ with $|U| = 6$ and $|V| = 4$ in Figure 3.2. This graph serves as a simple example of how the algorithm works.

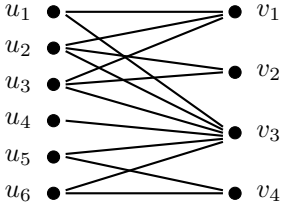


Figure 3.2: The initial bipartite graph $G = G_1 = ((U_1, V_1), E_1)$.

Step $k = 1$. We start with $G = G_1 = ((U_1, V_1), E_1)$. There are two nodes v_2, v_4 with minimum degree $\bar{d}_1 = 2$, so $n_1 = 2$. Pick uniformly at random one of them (with probability $\frac{1}{n_1} = \frac{1}{2}$), say $Y_1 = 2$. Eliminate node v_2, all its neighbors u_2, u_3, and all their edges $u_2v_1, u_2v_2, u_2v_3, u_3v_1, u_3v_2, u_3v_3$. Denote the new bipartite graph by $G_2 = ((U_2, V_2), E_2)$. The nucleation time associated with this node satisfies

$$\mathbb{E}_u[\mathcal{T}_{v_{Y_1}}^{Q^0}] = \mathbb{E}_u[\mathcal{T}_{v_2}^{Q^0}] = F^1 \left(Q_U^0\right)^{1\wedge\beta} [1 + o(1)], \qquad r \to \infty. \tag{3.20}$$

Step $k = 2$. Node v_1 has the minimum degree $\bar{d}_2 = 1$, so $Y_2 = 1$. Eliminate node v_1, all its neighbors, and all their edges. Denote the new graph by $G_3 = ((U_3, V_3), E_3)$. The nucleation time associated with this node satisfies

$$\mathbb{E}_u[\mathcal{T}_{v_{Y_2}}^{Q^0}] = \mathbb{E}_u[\mathcal{T}_{v_1}^{Q^1}] = F^2 \left(\mathbb{E}_u[Q_U^1]\right)^0 [1 + o(1)] = o(1), \qquad r \to \infty. \tag{3.21}$$

Step $k = 3$. Node v_4 has the minimum degree $\bar{d}_3 = 2$, so $Y_3 = 4$. Eliminate node v_4, all its neighbors, and all their edges. Denote the new graph by $G_4 = ((U_4, V_4), E_4)$.

The nucleation time associated with this node satisfies

$$\mathbb{E}_u[\mathcal{T}_{v_{Y_3}}^{Q^0}] = \mathbb{E}_u[\mathcal{T}_{v_4}^{Q^2}] = F^3\left(\mathbb{E}_u[Q_U^2]\right)^{1\wedge\beta}[1+o(1)], \qquad r\to\infty. \tag{3.22}$$

Step $k=4$. Node v_3 is the only node left, with degree $\bar{d}_4 = 1$, so $Y_4 = 3$. Eliminate node v_3, all its neighbors, and all their edges, after which the empty graph is left. The nucleation time associated with this node satisfies

$$\mathbb{E}_u[\mathcal{T}_{v_{Y_4}}^{Q^0}] = \mathbb{E}_u[\mathcal{T}_{v_3}^{Q^3}] = F^4\left(\mathbb{E}_u[Q_U^3]\right)^0[1+o(1)] = o(1), \qquad r\to\infty. \tag{3.23}$$

The above scenario forms a path that is described by nodes in V activating in the order v_2, v_1, v_4, v_3 (see Figure 3.3).

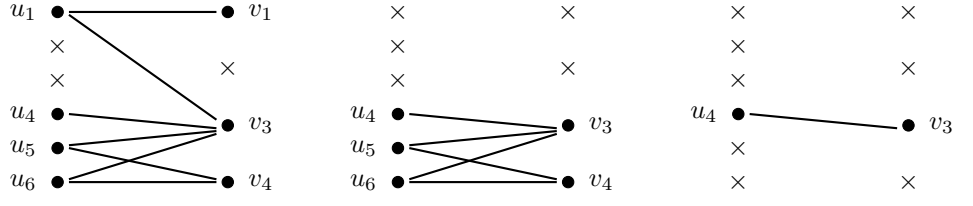


Figure 3.3: The sequence of bipartite graphs $G_2 = ((U_2, V_2), E_2)$, $G_3 = ((U_3, V_3), E_3)$ and $G_4 = ((U_4, V_4), E_4)$ generated by the algorithm.

Note that the algorithm may pick node v_4 at the first step by setting $Y_1 = 4$, since the choice of the node with minimum degree is uniformly at random. If so, then the algorithm would follow a different path. At the first step we would get $Y_1 = 4$ and $\mathbb{E}_u[\mathcal{T}_{v_4}^{Q^0}] = F^1\left(Q_U^0\right)^{1\wedge\beta}[1+o(1)]$. At the second step, $Y_2 = 2$ and $\mathbb{E}_u[\mathcal{T}_{v_2}^{Q^1}] = F^2\left(\mathbb{E}_u[Q_U^1]\right)^{1\wedge\beta}[1+o(1)]$. At the third step, $Y_3 = 1$ and $\mathbb{E}_u[\mathcal{T}_{v_1}^{Q^2}] = o(r)$. At the fourth step, $Y_4 = 3$ and $\mathbb{E}_u[\mathcal{T}_{v_3}^{Q^3}] = o(r)$. This choice leads to a different path, where the nodes in V activate in the order v_4, v_2, v_1, v_3.

Each possible scenario is identified with a path in the algorithm, described by the nodes in V according to the order of their first activation. The total mean transition time along a path can be thought as a sum of the mean nucleation times associated with each activating node in the path (see Theorem 3.3.2). We will prove in Section 3.5.2 that all the paths generated by the algorithm lead to the same order of the mean transition time.

§3.3 Main results

In this section we present our main theorems regarding the transition time. In Section 3.3.1 we show that $\mathcal{E}$, the event that the network follows the algorithm, occurs with high probability as $r\to\infty$ (Theorem 3.3.2(i)). We analyze the contributions along a given path, noting that not all the nucleation times are significant for the total mean transition time (Theorem 3.3.2(ii)). In Section 3.3.2 we compute the asymptotics of the mean transition time, including the pre-factor, focusing on the significant

terms only (Theorem 3.3.3). In Section 3.3.3 we identify the law of the transition time divided by its mean, which turns out to be a convolution of the laws found for the complete bipartite graph in Theorem 2.1.6 (Theorem 3.3.5). There is again a trichotomy, depending on the value of β. Proofs will be given in Section 3.6.

§3.3.1 Most likely paths

Let Ω be the set of all possible orderings (permutations) of nodes in V. Denote by $\mathcal{A} \subseteq \Omega$ the subset of orderings generated by the algorithm, and denote by $\mathcal{A}_{sc}$ the subset of orderings generated by the algorithm truncated at the first supercritical node (if there is any). Recall that, according to Definition 3.2.10, a supercritical node is a node that activates through a supercritical fork. If $a = (v_1, \dots, v_N)$ is an element of $\mathcal{A}$, then $a_{sc} = (v_1, \dots, v_{sc})$ is an element of $\mathcal{A}_{sc}$, where v_{sc} denotes the last node of each truncated ordering. We allow this node to be any of the remaining supercritical nodes not already present in the sequence.

Definition 3.3.1 (The network follows the algorithm).
Denote by $a^* = (v_1^*, \dots, v_N^*)$ the ordering of the nodes in V along the path a^* followed by the network. For fixed a^*, let

$$\mathcal{E}(a^*) = \{\exists\, a \in \mathcal{A}\colon\ a = a^*\} \cup \left\{\exists\, a_{sc} = (v_1, \dots, v_{sc}) \in \mathcal{A}_{sc}\colon\ v_1 = v_1^*, \dots, v_{sc} = v_{sc}^*\right\} \tag{3.24}$$

be the event that the network follows any of the paths generated by the algorithm up to the first supercritical node (if there is any).

Our first main theorem shows how the algorithm helps us to find the mean transition time. The first statement holds for all three regimes. The second and third statements hold in the subcritical and critical regimes only (for which the network follows the algorithm until the last activating node). The idea is that the mean transition time can be seen as a weighted sum of the mean nucleation times associated with each activation and of negligible terms representing the time it takes after each activation to bring the network back in the state with all the remaining nodes in U active. In the supercritical regime we do not need any statement, because the mean transition time is known to be the expected time it takes for the queue lengths to hit zero.

Theorem 3.3.2 (Most likely paths).
Consider the bipartite graph G with initial queue lenghts Q^0.

(i) *With high probability as $r \to \infty$, the network follows the algorithm, i.e.,*

$$\lim_{r\to\infty} \mathbb{P}_u(\mathcal{E}(a^*)) = 1. \tag{3.25}$$

Consider $\beta \in (0, \frac{1}{d^*-1}]$: subcritical or critical regime.

(ii) *With high probability as $r \to \infty$, the transition time satisfies*

$$\mathbb{E}_u[\mathcal{T}_G^{Q^0} \mathbb{1}_{\mathcal{E}(a^*)}] = \sum_{k=1}^{N} \sum_{\substack{i_1,\dots,i_k: \\ (v_{i_1},\dots,v_{i_k}) \in V_1 \times \cdots \times V_k}} \left(\prod_{l=1}^{k} \frac{1}{n_l} \right) f_k \, \mathbb{E}_u[\mathcal{T}_{v_{i_k}}^{Q^{k-1}} \mathbb{1}_{\mathcal{E}(a^*)}] \, [1 + o(1)], \quad r \to \infty, \tag{3.26}$$

where $n_k \in \mathbb{N}$ is the number of possible nodes that the algorithm can pick at step k, while the factor $f_k \in (0,1)$ (to be identified in Theorem 3.3.3) comes from the fact that the node activating at step k is the one that activates first among the n_k nodes with the same least degree. Both n_k and f_k depend on the sequence of nodes that activated before step k.

(iii) *Conditional on the path $a = (v_1, \dots, v_N) \in \mathcal{A}$ and the event*

$$A_a = \{a^* = a\} = \{v_1 = v_1^*, \dots, v_N = v_N^*\}, \tag{3.27}$$

with high probability as $r \to \infty$, the transition time satisfies

$$\mathbb{E}_u[\mathcal{T}_G^{Q^0} | A_a] = \sum_{k=1}^{N} f_k \, \mathbb{E}_u[\mathcal{T}_{v_k}^{Q^{k-1}}] \, [1 + o(1)], \qquad r \to \infty. \tag{3.28}$$

Theorem 3.3.2 will be proved in Section 3.6.3. Note that the mean transition time can be split as

$$\mathbb{E}_u[\mathcal{T}_G^{Q^0}] = \mathbb{E}_u[\mathcal{T}_G^{Q^0} \mathbb{1}_{\mathcal{E}(a^*)}] + \mathbb{E}_u[\mathcal{T}_G^{Q^0} \mathbb{1}_{\mathcal{E}(a^*)^C}]. \tag{3.29}$$

The second term in the right-hand side represents the mean transition time when the network does not follow the algorithm, and equals

$$\mathbb{E}_u[\mathcal{T}_G^{Q^0} \mathbb{1}_{\mathcal{E}(a^*)^C}] = \mathbb{E}_u[\mathcal{T}_G^{Q^0} | \mathcal{E}(a^*)^C] \, \mathbb{P}_u(\mathcal{E}(a^*)^C). \tag{3.30}$$

Even though we know from Theorem 3.3.2(i) that $\mathbb{P}_u[\mathcal{E}(a^*)^C]$ tends to zero as $r \to \infty$, a priori this term may still affect the total mean transition time, since the conditional expectation may be substantial. In what follows we focus on the first term in the right-hand side, since this captures the typical behavior of the network.

We will see in Theorem 3.3.3 below that, in the supercritical regime, the mean transition time is the expected time it takes for the queues in U to hit zero, independently of which path the network took before the activation of the first supercritical node. Theorem 3.3.2(ii) gives us a way, in the subcritical and critical regimes, to split the total mean transition time into a sum of mean nucleation times of successive forks, by taking into account all possible paths that the algorithm may follow, each with its own probability. Theorem 3.3.2(iii) shows that we can also think of the total mean transition time as a sum over all possible paths, each with its own probability and mean transition time, namely,

$$\mathbb{E}_u[\mathcal{T}_G^{Q^0} \mathbb{1}_{\mathcal{E}(a^*)}] = \sum_{a \in \mathcal{A}} \mathbb{E}_u[\mathcal{T}_G^{Q^0} \mathbb{1}_{A_a}] = \sum_{a \in \mathcal{A}} \mathbb{E}_u[\mathcal{T}_G^{Q^0} | A_a] \, \mathbb{P}_u(A_a). \tag{3.31}$$

The above expression allows us to compute the mean transition time along a single path. For every $a \in \mathcal{A}$,

$$\mathbb{P}_u(A_a) = \prod_{k=1}^{N} \frac{1}{n_k}. \tag{3.32}$$

We already saw in Proposition 3.2.7 that the order of the mean transition time does not depend on which path the algorithm generates.

§3.3.2 Mean of the transition time

Consider a path $a \in \mathcal{A}$ generated by the algorithm and the event A_a that the network follows this path. Recall that $d^* = \max_{1\le k\le N} \bar{d}_k$ is the maximum degree among the sequence of minimum degrees $(\bar{d}_k)_{k=1}^N$. Let v_k^* be the k-th activating node in path a. According to Definition 3.2.3, the mean nucleation time $\mathbb{E}_u[\mathcal{T}_{v_k^*}^{Q^{k-1}}]$ is given by

$$\mathbb{E}_u[\mathcal{T}_{v_k^*}^{Q^{k-1}}] = \begin{cases} F_{\text{sub}}^k \left(\mathbb{E}_u[Q_U^{k-1}]\right)^{\beta(\bar{d}_k-1)} [1+o(1)], & \text{if } \beta \in (0, \frac{1}{\bar{d}_k-1}), \\ F_{\text{cr}}^k \, \mathbb{E}_u[Q_U^{k-1}] \, [1+o(1)], & \text{if } \beta = \frac{1}{\bar{d}_k-1}, \\ F_{\text{sup}}^k \, \mathbb{E}_u[Q_U^{k-1}] \, [1+o(1)], & \text{if } \beta = (\frac{1}{\bar{d}_k-1}, \infty), \end{cases} \qquad r \to \infty, \tag{3.33}$$

with

$$F_{\text{sub}}^k = \frac{1}{\bar{d}_k B^{-(\bar{d}_k-1)}}, \qquad F_{\text{cr}}^k = \frac{1}{\bar{d}_k B^{-(\bar{d}_k-1)} + (c-\rho_U)}, \qquad F_{\text{sup}}^k = \frac{1}{c-\rho_U}, \tag{3.34}$$

are constants depending on $\bar{d}_k, B, c, \rho_U$. Note that F_{sub}^k really depends on k, while $F_{\text{cr}}^k = \frac{1}{\bar{d}_k B^{-(\bar{d}_k-1)}+(c-\rho_U)}$ is the same for every critical node, and $F_{\text{sup}}^k = F_{\text{sup}}$ is independent of k. Moreover, note that the first mean nucleation time depends on the initial queue lengths Q_U^0 at the nodes in U, but in general the mean nucleation time associated with a fork depends on the mean queue lengths at the nodes in U at the moment the fork starts the nucleation.

Our second main theorem identifies the mean transition time along a given path.

Theorem 3.3.3 (Mean transition time).
Consider the bipartite graph G with initial queue lengths Q^0.

(I) *$\beta \in (0, \frac{1}{d^*-1})$: subcritical regime. The transition time satisfies*

$$\mathbb{E}_u[\mathcal{T}_G^{Q^0} | A_a] = \sum_{\substack{1\le k\le N \\ k:\, \bar{d}_k = d^*}} f_k \, \frac{\gamma_U^{\beta(d^*-1)}}{d^* B^{-(d^*-1)}} \, r^{\beta(d^*-1)} \, [1+o(1)], \qquad r \to \infty, \tag{3.35}$$

with

$$f_k = \frac{1}{n_k}. \tag{3.36}$$

(II) $\beta = \frac{1}{d^*-1}$*: critical regime. Denote by* $h_k \in \mathbb{N}_0$ *the number of nodes in* V *at step* k *that already activated through a fork of degree* d^**. Then the transition time satisfies*

$$\mathbb{E}_u[\mathcal{T}_G^{Q^0}|A_a] = \sum_{\substack{1\le k\le N\\ k:\,\bar{d}_k=d^*}} f_k \frac{\gamma_U^{(h_k)}}{d^* B^{-(d^*-1)} + (c-\rho_U)}\, r\,[1+o(1)], \qquad r\to\infty, \tag{3.37}$$

with

$$f_k = \frac{\bar{d}_k B^{-(\bar{d}_k-1)} + (c-\rho_U)}{n_k \bar{d}_k B^{-(\bar{d}_k-1)} + (c-\rho_U)} \tag{3.38}$$

and

$$\gamma_U^{(h_k)} = \gamma_U - (c-\rho_U) \sum_{\substack{1\le i\le k\\ i:\,\bar{d}_i=d^*}} f_i', \tag{3.39}$$

where for a critical node v_i *the coefficient* f_i' *is defined in a recursive way as*

$$f_i' = \frac{1}{n_i \bar{d}_i B^{-(\bar{d}_i-1)} + (c-\rho_U)} \left(\gamma_U - (c-\rho_U) \sum_{\substack{1\le j\le i-1\\ j:\,\bar{d}_j=d^*}} f_j'\right) > 0. \tag{3.40}$$

(III) $\beta \in (\frac{1}{d^*-1}, \infty)$*: supercritical regime. The transition time satisfies*

$$\mathbb{E}_u[\mathcal{T}_G^{Q^0}] = \frac{\gamma_U}{c-\rho_U} r\,[1+o(1)], \qquad r\to\infty. \tag{3.41}$$

Theorem 3.3.3 will be proved in Section 3.6.4. Both in the subcritical and supercritical regimes, Theorem 3.3.3 provides explicit formulas for the mean transition time in terms of the parameters c, γ_U, ρ_U and B, β in our model and the sequence of numbers $(\bar{d}_k, n_k)_{k=1}^N$ that are produced by the algorithm, with $d^* = \max_{1\le k\le N} \bar{d}_k$. In the critical regime, however, the formula is more delicate, since the pre-factor depends on how long the critical nucleations take. Indeed, $\gamma_U^{(h_k)}$ in (3.39) represent the mean updated queue lengths at step k after h_k nodes in V activate through critical forks (see Section 3.4.3 for more details). Recall from Chapter 2 that the queue lengths all have a good behavior, in the sense that, with high probability as $r\to\infty$, they are always close to their mean (see Remark 3.4.7). Note that the mean transition time in the subcritical and critical regimes depends on the path, while in the supercritical regime it does not.

§3.3.3 Law of the transition time

Theorem 3.3.2 shows how the mean transition time along a path is a sum of terms related to the successive mean nucleation times of complete bipartite subgraphs of G. Theorem 3.3.3 tells us that, depending on the value of β, this sum reduces to a smaller sum of only a few significant terms. It also tells us how to compute the pre-factors of these terms.

Definition 3.3.4 (Multiplicity of d^*).
Consider a path $a \in \mathcal{A}$ generated by the algorithm and its associated degree sequence $(\bar{d}_k)_{k=1}^N$. Write m^a_{sub} and m^a_{cr} to denote the multiplicity of d^* in the path a in the subcritical and critical regimes, i.e.,

$$m^a_{\text{sub}} = |\{k\colon\ \bar{d}_k = d^* < \beta^{-1} + 1\}|, \tag{3.42}$$

$$m^a_{\text{cr}} = |\{k\colon\ \bar{d}_k = d^* = \beta^{-1} + 1\}|. \tag{3.43}$$

Our third main theorem identifies the law of $\mathcal{T}_G^{Q^0}/\mathbb{E}_u[\mathcal{T}_G^{Q^0}]$. Recall the laws $\mathcal{P}_{\text{sub}}, \mathcal{P}_{\text{cr}}, \mathcal{P}_{\text{sup}}$ arising from Theorem 2.1.6. Write $\circledast$ to denote convolution.

Theorem 3.3.5 (Law of the transition time).
Consider the bipartite graph G with initial queue lengths Q^0.

(I) *$\beta \in (0, \frac{1}{d^*-1})$: subcritical regime. With f_k as in (3.36) and m^a_{sub} as in (3.42), the transition time satisfies*

$$\lim_{r\to\infty} \mathbb{P}_u\left(\frac{\mathcal{T}_G^{Q^0}}{\mathbb{E}_u[\mathcal{T}_G^{Q^0}|A_a]} > x \mid A_a\right) = \int_x^\infty \left(\circledast_{k=1}^{m^a_{\text{sub}}} \mathcal{P}_{\text{sub}}^{f_k, S_{m^a_{\text{sub}}}}\right)(y)\, dy, \quad x \in [0,\infty), \tag{3.44}$$

with

$$\mathcal{P}_{\text{sub}}^{f_k, S_{m^a_{\text{sub}}}}(z) = \frac{S_{m^a_{\text{sub}}}}{f_k} \exp\left(-\frac{S_{m^a_{\text{sub}}}}{f_k}\, z\right), \qquad z \in [0,\infty), \tag{3.45}$$

and with $S_{m^a_{\text{sub}}} = \sum_{i\colon\ \bar{d}_i = d^*} f_i$.

(III) *$\beta \in (\frac{1}{d^*-1}, \infty)$: supercritical regime. The transition time satisfies*

$$\lim_{r\to\infty} \mathbb{P}_u\left(\frac{\mathcal{T}_G^{Q^0}}{\mathbb{E}_u[\mathcal{T}_G^{Q^0}]} > x\right) = \int_x^\infty \mathcal{P}_{\text{sup}}(y)\, dy = \begin{cases} 1, & \text{if } x \in [0,1), \\ 0, & \text{if } x \in [1,\infty), \end{cases} \tag{3.46}$$

with

$$\mathcal{P}_{\text{sup}}(z) = \delta_1(z), \qquad z \in [0,\infty), \tag{3.47}$$

where $\delta_1(z)$ is the Dirac function at 1.

Theorem 3.3.5 will be proved in Section 3.6.5. There we will also see why there is no statement for the critical regime (II).

§3.3.4 Discussion

Intuition. Analyzing the transition time for arbitrary bipartite graphs is much harder than for complete bipartite graphs. The key idea is to view the transition time as a sum of subsequent nucleation times for complete bipartite subgraphs. The order in which nodes activate in V is random, because it depends on the fluctuations of the activation rates via the queue lengths. However, with high probability as $r \to \infty$, the nodes with the least number of active neighbors in U activate first. After each activation, the underlying bipartite graph changes according to which node activates

and which nodes deactivate. Hence the subsequent activations in V depend on how the graph changes, as well as on the evolution of the network, since the queue lengths (and hence the activation rates) change over time as well. Recall that the queue lengths all have a good behavior (see Remark 3.4.7).

Theorems. To keep track of this evolution, we defined a greedy algorithm in Section 3.2. If we run the algorithm once, then it generates a specific path of activating nodes in V. This is enough to determine the leading order of the transition time as $r \to \infty$, since it only depends on the maximum least degree d^*, which is the same for all the paths that can be generated. Moreover, given d^*, we can immediately determine whether we are in the subcritical, critical or supercritical regime. If we are interested in the pre-factor of the mean transition time and in its law, then we need to generate all possible paths. Theorem 3.3.2 shows that we can split the mean transition time into a weighted sum over all possible paths of the mean nucleation times associated with each activation in the path. Theorem 3.3.3 gives the mean transition time conditional on the path and shows that the outcome is non-trivial both in the subcritical and critical regimes. Theorem 3.3.5 gives the law conditional on the path, but fails to capture the critical regime. The reason is that there are intricate dependencies between the subsequent nucleation times along the path.

§3.4 Nucleation times and queue lengths

In Section 3.4.1 we introduce the concept of asymptotic independence of forks and we show that in the subcritical and critical regimes competing forks can be treated as if they were disjoint, in the limit as $r \to \infty$ (Proposition 3.4.1). In Section 3.4.2 we study the mean and the law of the next nucleation time by using techniques from metastability and results from Section 3.4.1 (Propositions 3.4.3 and 3.4.6). In Section 3.4.3 we show how the mean queue lengths change according to which node activates in V (Theorem 3.4.8).

§3.4.1 Asymptotic independence of forks

In this section we show that, as $r \to \infty$, forks can be treated as being independent of each other even when they share some nodes. We introduce the concept of *asymptotic independence* of forks, which holds only as $r \to \infty$ and which allows us to treat overlapping forks as if they were disjoint. We show that the nucleation time of a fork is not influenced by the behavior of other forks sharing nodes with *it*.

In Chapter 2 it is shown that, as soon as all the nodes in U of a complete bipartite graph become simultaneously inactive, the first node in V (and subsequently all the others nodes) activates in a very short time interval, negligible compared to the time it takes for all the nodes in U to deactivate. Hence, the time it takes for the nodes in U to become all simultaneously inactive is the same as the time it takes for the first node in V to activate, up to an error term that is negligible as $r \to \infty$. In our setting, to study the nucleation times of forks it is enough to study the time it takes for all their respective nodes in U to deactivate, without considering the set V.

Proposition 3.4.1 (Asymptotic independence).
Consider the graph G_k and the $\bar{d}_k$-fork W, where $\bar{d}_k$ is the minimum degree of the nodes in V_k. Denote by T_W the time it takes for fork W to nucleate for the first time. Consider the event

$$\mathcal{E} = \Big\{ \exists \{s_1, \dots, s_\alpha\} \subset \{u_1, \dots, u_{\bar{d}_k}\} = W \cap U_k \colon \exists\, 0 \le t < \bar{\tau}_k \text{ s.t.} \\ X_{s_i}\Big(\sum_{j=1}^{k-1} \bar{\tau}_j + t \Big) = 0 \ \forall\, i = 1, \dots, \alpha \Big\} \tag{3.48}$$

of having a subset of α nodes in U_k belonging to fork W that are simultaneously inactive at a time t after the last nucleation. The following statements hold.

(i) The mean nucleation time of W satisfies

$$\mathbb{E}_u[T_W \,|\, \mathcal{E}] = \mathbb{E}_u[T_W]\,[1 + o(1)], \qquad r \to \infty. \tag{3.49}$$

(ii) The law of the nucleation time of W satisfies

$$\lim_{r\to\infty} \mathbb{P}_u(T_W > x \mid \mathcal{E}) = \lim_{r\to\infty} \mathbb{P}_u(T_W > x), \quad x \in [0, \infty). \tag{3.50}$$

Proof. We prove the two statements separately.

(i) We denote by $\mathcal{S}$ the event that after time t all the nodes of W that are still active become simultaneously inactive before any of the inactive nodes in $\{s_1, \dots, s_\alpha\}$ activates again. We know that the time it takes for $\bar{d}_k - \alpha$ nodes to become simultaneously inactive is an exponential random variable $T_{\mathcal{S}}$ with mean of order $r^{\beta(\bar{d}_k - \alpha - 1)}$, while the time it takes for one of the α inactive nodes to activate is an exponential random variable with mean of order $1/r^\beta$. Hence the probability of $\mathcal{S}$ is of order $r^{-\beta(\bar{d}_k - \alpha)} = o(1)$. If $\mathcal{S}$ occurs, then W nucleates in time $T_W = t + T_{\mathcal{S}}$. Note that t must be of order $r^{\beta(\alpha-1)}$, hence

$$\mathbb{E}_u[T_W \,|\, \mathcal{E} \cap \mathcal{S}] = O\big(r^{\beta(\alpha-1)}\big) + O\big(r^{\beta(\bar{d}_k - \alpha - 1)}\big) = o\big(r^{\beta(\bar{d}_k - 1)}\big), \qquad r \to \infty. \tag{3.51}$$

On the other hand, if the complementary event $\mathcal{S}^C$ occurs, then, with high probability as $r \to \infty$, in a negligible time $o(1)$ the network reaches the state with all the nodes $u_1, \dots, u_{\bar{d}_k}$ active, and from there it takes time $\mathbb{E}_u[T_W]$ for W to nucleate. Hence

$$\mathbb{E}_u[T_W \,|\, \mathcal{E} \cap \mathcal{S}^C] = o(1) + \mathbb{E}_u[T_W], \qquad r \to \infty. \tag{3.52}$$

Putting the two complementary events together, we obtain that

$$\begin{aligned} \mathbb{E}_u[T_W \mid \mathcal{E}] &= \mathbb{E}_u\big[T_W \mid \mathcal{E} \cap \mathcal{S}\big]\, \mathbb{P}_u(\mathcal{S}) + \mathbb{E}_u\big[T_W \mid \mathcal{E} \cap \mathcal{S}^C\big]\, \mathbb{P}_u(\mathcal{S}^C) \\ &= o\big(r^{\beta(\bar{d}_k - 1)}\big)\, o(1) + \big(o(1) + \mathbb{E}_u[T_W]\big)\,(1 - o(1)) \\ &= \mathbb{E}_u[T_W]\,[1 + o(1)], \qquad r \to \infty. \end{aligned} \tag{3.53}$$

(ii) Using the complementary events $\mathcal{S}$ and $\mathcal{S}^C$, we can write for all $x \geq 0$

$$\begin{aligned}\lim_{r\to\infty} \mathbb{P}_u(T_W > x \mid \mathcal{E}) &= \lim_{r\to\infty} \mathbb{P}_u(\{T_W > x\} \cap \mathcal{S} \mid \mathcal{E})\, \mathbb{P}_u(\mathcal{S}) \\ &\quad + \lim_{r\to\infty} \mathbb{P}_u(\{T_W > x\} \cap \mathcal{S}^C \mid \mathcal{E})\, \mathbb{P}_u(\mathcal{S}^C) \\ &= \lim_{r\to\infty} \mathbb{P}_u(T_W > x),\end{aligned} \tag{3.54}$$

since $\lim_{r\to\infty} \mathbb{P}_u(\mathcal{S}) = 0$ and, when conditioning on $\mathcal{S}^C$, with high probability as $r \to \infty$, the network reaches the initial state in a negligible time after t, hence it behaves as if at time t all nodes in U were active.

□

The above proposition shows that, as $r \to \infty$, the mean nucleation time of a fork and its law are not influenced by the fact that some of its nodes might be simultaneously inactive at some time. The intuition is that, as $r \to \infty$, the nucleation of a fork is so hard to achieve and takes so long that sharing some nodes with other forks does not help to make the nucleation happen appreciably faster. The network tends to quickly reach the initial state with all the remaining nodes in U active, and hence the nucleation time of a fork can be seen as the time it takes for its nodes in U to deactivate starting from all of them being active. In particular, in case of overlapping forks, the nucleation time of a fork is not influenced by the behavior of other forks sharing nodes with it.

§3.4.2 Next nucleation time

Given the graph G_k, consider the next nucleation time $\bar{\tau}_k$ from Definition 3.2.8. The next node that activates is the one that completes the fastest nucleation among the n_k nodes with least degree. We want to find an expression for $\mathbb{E}_u[\bar{\tau}_k]$.

In Appendix A we show the computations for the mean next nucleation time in the case when the competing forks are disjoint, hence described by i.i.d. random variables. Recall that in the subcritical regime we are considering a minimum of nucleation times that are exponential random variables, while in the critical regime we are considering a minimum of nucleation times that follow a truncated polynomial law (see Theorem 2.1.6). By using Proposition 3.4.1, we are also able to give explicit asymptotics for the mean next nucleation time without assuming the forks being independent.

Each nucleation of a fork can be seen as a successful escape from a metastable state, which is represented by the initial state where the nodes in U_k in the fork are active and the node in V_k in the fork is inactive. When considering multiple forks, we can view the network as an ergodic Markov process on a state space Ω representing the collection of all the feasible joint activity states of G_k. The first nucleation can be described by a regenerative process where the Markov process leaves a metastable state x_0 (with all the nodes in U_k active) and reaches a stable set S, which represents the set of states where at least one of the forks of minimum degree has all its nodes in U_k simultaneously inactive. The set S is rare for the Markov process, in the sense

that the probability of reaching S starting from x_0 is small. We denote by $T^k_{x_0\to S} = \bar\tau_k$ the time it takes to go from x_0 to S.

Lemma 3.4.2 (Mean return time to metastable state).
For $k = 1, \dots, N$, suppose that $k-1$ nodes in V already activated. Then, with high probability as $r \to \infty$, the time $R^x_{U_k}$ it takes for the network G_k to reach the state with all the nodes in U_k active (the metastable state x_0) starting from any other state x is negligible, i.e.,

$$\mathbb{E}_u[R^x_{U_k}] = o(1), \qquad r \to \infty. \tag{3.55}$$

In particular, let $R^{k-1}_{U_k}$ be the time it takes for the network G_k to reach the state with all the nodes in U_k active starting from the moment the $(k-1)$-th node in V activated. Then, with high probability as $r \to \infty$,

$$\mathbb{E}_u[R^{k-1}_{U_k}] = o(1), \qquad r \to \infty. \tag{3.56}$$

Proof. At any time t, the activation and deactivation of a node $u \in U_k$ are described by i.i.d. exponential random variables with rates $g_U(Q_u(t))$ and 1, respectively. Hence, an active node takes on average one unit of time to deactivate, while an inactive node u takes on average $1/g_U(Q_u(t))$ time to activate. Since in the subcritical and critical regimes the queue lengths at any node at any moment are of order r (see Section 3.4.3 for more details), we can say that $1/g_U(Q_u(t)) = o(1)$ for each $u \in U_k$. Suppose that, at some time t, node $u_1 \in U_k$ is inactive and node $u_2 \in U_k$ is active, i.e., $X_{u_1}(t) = 0$ and $X_{u_2}(t) = 1$. Since

$$\lim_{r\to\infty} \mathbb{P}_u(u_1 \text{ activates} < u_2 \text{ deactivates}) = 1, \tag{3.57}$$

and there is a finite number of nodes in U_k, with high probability as $r \to \infty$, starting from any state x all the nodes in U_k will be active on average in time $o(1)$. Hence, as $r \to \infty$, $\mathbb{E}_u[R^x_{U_k}] = o(1)$, and in particular $\mathbb{E}_u[R^{k-1}_{U_k}] = o(1)$. □

We are now ready to state a result for the mean next nucleation time in the subcritical and critical regimes.

Proposition 3.4.3 (Mean next nucleation time).
Consider the graph G_k. Recall that $\bar d_k$ is the minimum degree of a node in V_k, n_k is the number of forks of degree $\bar d_k$ in G_k, and h_k is as in (3.74).

(I) *$\beta \in (0, \frac{1}{\bar d_k - 1})$: subcritical regime. The mean next nucleation time sastisfies*

$$\mathbb{E}_u[\bar\tau_k] = f_k\, \mathbb{E}_u[\mathcal{T}^{Q^{k-1}}_{v^*_k}] = f_k\, F^k_{\text{sub}}\, \mathbb{E}_u[Q^{k-1}_U]^{\beta(\bar d_k - 1)}\, [1+o(1)], \qquad r \to \infty, \tag{3.58}$$

with

$$f_k = \frac{1}{n_k}. \tag{3.59}$$

(II) *$\beta = \frac{1}{\bar d_k - 1}$: critical regime. The mean next nucleation time sastisfies*

$$\mathbb{E}_u[\bar\tau_k] = f_k\, \mathbb{E}_u[\mathcal{T}^{Q^{k-1}}_{v^*_k}] = f_k\, F^k_{cr}\, \mathbb{E}_u[\mathcal{T}^{Q^{k-1}}_{v^*_k}]\, [1 + o(1)], \qquad r \to \infty, \tag{3.60}$$

with

$$f_k = \frac{\bar d_k B^{-(\bar d_k - 1)} + (c - \rho_U)}{n_k \bar d_k B^{-(\bar d_k - 1)} + (c - \rho_U)}. \tag{3.61}$$

Proof. By Proposition 3.4.1, as $r \to \infty$ we may consider arbitrarily overlapping forks as if they were disjoint. Therefore the computations for the mean next nucleation time carried out in Appendix C for the case of disjoint forks can be used for the case of overlapping forks as well. For completeness, in the subcritical regime (I) we offer a proof that uses a different argument, which cannot be used in the critical regime (II) because the queues are changing on scale r over time.

Consider the stationary distribution π of the Markov process mentioned above. The probability of the set S is given by

$$\pi(S) = \sum_{j=1}^{n_k} \pi(S_j)\,[1+o(1)] = n_k \left(\frac{1}{B\,(Q_U^{k-1})^{\beta}} \right)^{\bar{d}_k} [1+o(1)], \qquad r \to \infty, \tag{3.62}$$

where S_j is the event that the j-th fork has all its nodes simultaneously inactive. The terms representing multiple forks with all their nodes simultaneously inactive contribute in a negligible way to $\pi(S)$. Moreover, we know that, for $j = 1, \ldots, n_k$,

$$\begin{aligned} \pi(S_j) &= \frac{\mathbb{E}_u[\text{time spent in } S_j]}{\mathbb{E}_u[\text{time spent in } S_j] + \mathbb{E}_u[T^k_{x_0 \to S_j}]} \\ &= \frac{\frac{1}{\bar{d}_k} \frac{1}{B\,(Q_U^{k-1})^{\beta}}}{\frac{1}{\bar{d}_k} \frac{1}{B\,(Q_U^{k-1})^{\beta}} + F^k_{\text{sub}}\,(Q_U^{k-1})^{\beta(\bar{d}_k - 1)}} [1+o(1)] \\ &= \frac{\frac{1}{\bar{d}_k} \frac{1}{B\,(Q_U^{k-1})^{\beta}}}{F^k_{\text{sub}}\,(Q_U^{k-1})^{\beta(\bar{d}_k - 1)}} [1+o(1)] = \left(\frac{1}{B\,(Q_U^{k-1})^{\beta}} \right)^{\bar{d}_k} [1+o(1)], \qquad r \to \infty. \end{aligned} \tag{3.63}$$

This proves (3.62).

Using the same type of argument, we can compute $\mathbb{E}_u[T_{x_0 \to S}]$. Indeed,

$$\begin{aligned} \pi(S) &= \frac{\mathbb{E}_u[\text{time spent in } S]}{\mathbb{E}_u[\text{time spent in } S] + \mathbb{E}_u[T^k_{x_0 \to S}]} \\ &= \frac{\frac{1}{\bar{d}_k} \frac{1}{B\,(Q_U^{k-1})^{\beta}}}{\frac{1}{\bar{d}_k} \frac{1}{B\,(Q_U^{k-1})^{\beta}} + \mathbb{E}_u[T^k_{x_0 \to S}]} [1+o(1)] = \frac{\frac{1}{\bar{d}_k} \frac{1}{B\,(Q_U^{k-1})^{\beta}}}{\mathbb{E}_u[T^k_{x_0 \to S}]} [1+o(1)], \qquad r \to \infty. \end{aligned} \tag{3.64}$$

After inverting, we get

$$\begin{aligned} \mathbb{E}_u[\bar{\tau}_k] = \mathbb{E}_u[T^k_{x_0 \to S}] &= \frac{\frac{1}{\bar{d}_k} \frac{1}{B\,(Q_U^{k-1})^{\beta}}}{\pi(S)} [1+o(1)] = \frac{\frac{1}{\bar{d}_k} \frac{1}{B\,(Q_U^{k-1})^{\beta}}}{n_k \big(\frac{1}{B\,(Q_U^{k-1})^{\beta}} \big)^{\bar{d}_k}} [1+o(1)] \\ &= f_k\, F^k_{\text{sub}}\,(Q_U^{k-1})^{\beta(\bar{d}_k - 1)}\,[1+o(1)], \qquad r \to \infty, \end{aligned} \tag{3.65}$$

with

$$f_k = \frac{1}{n_k}. \tag{3.66}$$

This completes the proof. □

Corollary 3.4.4 (Pre-factor adjustment).
Given the graph G_k, conditional on the next activating node of degree $\bar{d}_k$,

$$\mathbb{E}_u[\bar{\tau}_k | Y_k = i_k] = \mathbb{E}_u\Big[\min_{v\in V_k} \mathcal{T}_v^{Q^{k-1}} \;\Big|\; Y_k = i_k\Big] = f_k\, \mathbb{E}_u[\mathcal{T}_{v_{i_k}}^{Q^{k-1}}], \qquad r \to \infty, \quad (3.67)$$

where f_k is as in (3.59) *or* (3.61) *when a subcritical node or a critical node activates, respectively.*

Proof. The claim follows from Proposition 3.4.3. □

In the subcritical regime (I), the queue lengths do not change on scale r and therefore the renewal theory developed in [49] applies, which is tailored to exponential behavior in metastable regimes. In the critical regime (II), however, the queue lengths do change on scale r and [49] does not apply. For details, see Section 3.4.3. Recall that Ω is the state space of the Markov process and that, in our notation, $\bar{\tau}_k = T_{x_0 \to S}$.

Definition 3.4.5 (Recurrence property).
Let $H > 0$ and $h \in (0,1)$. We say that the pair (x_0, S) satisfies the property $\mathrm{Rec}(H,h)$ if

$$\sup_{x\in\Omega} \mathbb{P}\big(T_{x\to\{x_0,S\}} > H\big) \le h. \quad (3.68)$$

The following result is the equivalent of [49, Theorem 2.3].

Proposition 3.4.6 (Law of the next nucleation time).
Consider the pair (x_0, S) such that the property $\mathrm{Rec}(H,h)$ holds for $0 < H < \mathbb{E}_u[\bar{\tau}_k]$, with $\epsilon = H/\mathbb{E}_u[\bar{\tau}_k]$ and h sufficiently small. Then there exist functions $C(\epsilon,h)$ and $\lambda(\epsilon,h)$, satisfying $C(\epsilon,h), \lambda(\epsilon,h) \to 0$ as $\epsilon, h \to 0$, such that, for any $t > 0$,

$$\left| \mathbb{P}\left(\frac{\bar{\tau}_k}{\mathbb{E}_u[\bar{\tau}_k]} > t \right) - e^{-t} \right| \le C e^{-(1-\lambda)t}. \quad (3.69)$$

Proof. We choose H to be a constant, and without loss of generality set $H = 1$. We claim that the pair (x_0, S) satisfies the property $\mathrm{Rec}(H,h)$ with h sufficiently small. Indeed, starting from any state $x \in \Omega$, the network reaches the set $\{x_0, S\}$ in a small time $o(1)$.

If the starting state x is one of the states S_j, $j = 1, \dots, n_k$, corresponding to the set S, then we are done. Otherwise, by Lemma 3.4.2, the metastable state x_0 attracts in time $o(1)$ every state x for which some forks have some nodes in U inactive. It is therefore immediate that, with high probability as $r \to \infty$, $T_{x\to\{x_0,S\}}$ is smaller than H, which is what we need in order to claim that (3.68) holds when h is sufficiently small. Note that we can let $h \to 0$ as $r \to \infty$.

We recover from Proposition 3.4.3 that the ratio between H and the mean next nucleation time is sufficiently small. Indeed, $\epsilon = H/\mathbb{E}_u[\bar{\tau}_k] \to 0$ as $r \to \infty$. Hence a straightforward application of [49, Theorem 2.3] allows us to conclude that the next nucleation time divided by its mean follows an exponential law with unit rate. □

§3.4.3 Updated queue lengths

In this section we analyze in more detail how the mean queue lengths change over time and how they affect the mean nucleation times associated with each step of the algorithm.

Remark 3.4.7 (Good behavior).
Recall from Definition 2.3.1 and Lemma 2.3.2 in Chapter 2 that, with high probability as $r \to \infty$, the queue lengths all have a good behavior in the interval $[0, T_U(r)]$ with $T_U(r) = \frac{\gamma_U}{c-\rho_U} r\,[1 + o(1)]$ as $r \to \infty$, representing the expected time it takes for the queue lengths to hit zero. More precisely, for $\delta > 0$ small enough and for all $t \in [0, T_U(r)]$,

$$\lim_{r\to\infty} \mathbb{P}_u\Big(\mathbb{E}_u[Q_U(t)] - \delta r \le Q_U(t) \le \mathbb{E}_u[Q_U(t)] + \delta r\Big) = 1, \tag{3.70}$$

which means that the queue lengths are always close to their mean for all times smaller than $T_U(r)$.

Recall that we start with initial queue lengths $Q^0 = (Q_U^0, Q_V^0) = (\gamma_U r, \gamma_V r)$. We are interested in studying how the queue lengths change along a fixed path, depending on which types of forks we encounter at each activation. Fix a path and consider the sequence of nodes activating in V.

Similarly to (3.33), the next nucleation time $\bar{\tau}_k = \min_{v\in V_k} \mathcal{T}_v^{Q^{k-1}}$ (recall Definition 3.2.8) satisfies

$$\mathbb{E}_u[\bar{\tau}_k] = f'_k\, r^{1\wedge\beta(\bar{d}_k-1)}\,[1 + o(1)], \qquad r \to \infty, \tag{3.71}$$

where f'_k depends on f_k, on the constants $F^k_{\mathrm{sub}}, F^k_{\mathrm{cr}}, F^k_{\mathrm{sup}}$ (for the three regimes, respectively), and on the mean updated queue lengths. The following theorem shows how the mean queue lengths change according to which type of node activates in V.

Theorem 3.4.8 (Mean updated queue lengths).
Let $(\bar{d}_k)_{k=1}^N$ be the sequence of degrees in a fixed path and $d^* = \max_{1\le k\le N} \bar{d}_k$.

(I) *$\beta \in (0, \frac{1}{d^*-1})$: subcritical regime. After step k, the mean queue length at a node in U is*

$$\mathbb{E}_u[Q_U^k] = \gamma_U r\,[1 + o(1)], \qquad r \to \infty. \tag{3.72}$$

(II) *$\beta = \frac{1}{d^*-1}$: critical regime. After step k, the mean queue length at a node in U, after h_k critical nodes in V activated, is*

$$\mathbb{E}_u[Q_U^k] = \gamma_U^{(h_k)} r\,[1 + o(1)], \qquad r \to \infty, \tag{3.73}$$

with

$$\gamma_U^{(h_k)} = \gamma_U - (c - \rho_U) \sum_{\substack{1\le i\le k \\ i:\, \bar{d}_i = d^*}} f'_i > 0, \tag{3.74}$$

where for a critical node v_i the coefficient f'_i is defined in a recursive way as

$$f'_i = \frac{1}{n_i \bar{d}_i B^{-(\bar{d}_i-1)} + (c - \rho_U)} \left(\gamma_U - (c - \rho_U) \sum_{\substack{1\le j\le i-1 \\ j:\, \bar{d}_j = d^*}} f'_j\right) > 0. \tag{3.75}$$

(III) $\beta \in (\frac{1}{d^*-1}, \infty)$*: supercritical regime. After step* k*, the mean queue length at a node in* U*, if any supercritical node in* V *activated, is*

$$\mathbb{E}_u[Q_U^k] = o(r), \qquad r \to \infty. \tag{3.76}$$

Proof. We treat the three regimes separately.

(I) $\beta \in (0, \frac{1}{d^*-1})$. All the nodes in V are subcritical, in particular the first node $v_1 \in V$. Then $\mathbb{E}_u[\bar{\tau}_1] = o(r)$ as $r \to \infty$. The mean queue lengths at nodes in U after node v_1 activates are

$$\begin{aligned} \mathbb{E}_u[Q_U(\bar{\tau}_1)] &= \mathbb{E}_u[\gamma_U r - (c-\rho_U)\bar{\tau}_1] = \gamma_U r - (c-\rho_U)\mathbb{E}_u[\bar{\tau}_1] \\ &= \gamma_U r\,[1+o(1)], \qquad r \to \infty, \end{aligned} \tag{3.77}$$

which means that after the first activation the mean queue lengths are the same as before, up to an error term $o(1)$. Iterating this reasoning, we conclude that the mean queue lengths remain approximately the same as long as subcritical nodes in V activate.

(II) $\beta = \frac{1}{d^*-1}$. If the first node $v_1 \in V$ is subcritical, then the time it takes to nucleate its fork does not influence the mean queue lengths by much, as seen in (I). Without loss of generality, we may therefore assume that v_1 is critical. Then $\mathbb{E}_u[\bar{\tau}_1] = f_1' r$ is of order r. The mean queue lengths at nodes in U after node v_1 activates are

$$\begin{aligned} \mathbb{E}_u[Q_U(\bar{\tau}_1)] &= \mathbb{E}_u[\gamma_U r - (c-\rho_U)\bar{\tau}_1] = \gamma_U r - (c-\rho_U)\mathbb{E}_u[\bar{\tau}_1] \\ &= (\gamma_U - (c-\rho_U)f_1')r\,[1+o(1)] = \gamma_U^{(1)} r\,[1+o(1)], \qquad r \to \infty, \end{aligned} \tag{3.78}$$

where $\gamma_U^{(1)} = \gamma_U - (c-\rho_U)f_1' > 0$.

If the second node $v_2 \in V$ is subcritical, then again the time it takes to nucleate its fork does not influence the mean queue lengths by much. Assume therefore that v_2 is critical. Then the fork requires a nucleation time of order r, namely, $\mathbb{E}_u[\bar{\tau}_2] = f_2' r$. The mean queue lengths at nodes in U after node $v_2 \in V$ activates are

$$\begin{aligned} \mathbb{E}_u[Q_U(\bar{\tau}_1 + \bar{\tau}_2)] &= \mathbb{E}_u[\gamma_U r - (c-\rho_U)(\bar{\tau}_1 + \bar{\tau}_2)] \\ &= \gamma_U r - (c-\rho_U)(\mathbb{E}_u[\bar{\tau}_1] + \mathbb{E}_u[\bar{\tau}_2]) \\ &= (\gamma_U - (c-\rho_U)(f_1' + f_2'))r\,[1+o(1)] \\ &= \gamma_U^{(2)} r\,[1+o(1)], \qquad r \to \infty, \end{aligned} \tag{3.79}$$

where $\gamma_U^{(2)} = \gamma_U - (c-\rho_U)(f_1' + f_2') > 0$.

More generally, assume that h_k critical nodes activated in the first k steps. Then

$$\mathbb{E}_u[Q_U^k] = \gamma_U^{(h_k)} r\,[1+o(1)], \qquad r \to \infty, \tag{3.80}$$

with

$$\gamma_U^{(h_k)} = \gamma_U - (c-\rho_U) \sum_{\substack{1\le i\le k\\ i:\,\bar{d}_i = d^*}} f_i' > 0, \tag{3.81}$$

where the last sum is over all the h_k critical nodes. Each of them contributes with a positive coefficient f_i' which is given by the recursive relation

$$\begin{aligned} f_i' &= f_i\, F_{\rm cr}^i\, \gamma_U^{(h_{i-1})} \\ &= \frac{\bar{d}_i B^{-(\bar{d}_i-1)} + (c-\rho_U)}{n_i \bar{d}_i B^{-(\bar{d}_i-1)} + (c-\rho_U)} \, \frac{1}{\bar{d}_i B^{-(\bar{d}_i-1)} + (c-\rho_U)}\, \gamma_U^{(h_{i-1})} \\ &= \frac{1}{n_i \bar{d}_i B^{-(\bar{d}_i-1)} + (c-\rho_U)} \left(\gamma_U - (c-\rho_U) \sum_{\substack{1\le j\le i-1\\ j:\,\bar{d}_j = d^*}} f_j' \right). \end{aligned} \tag{3.82}$$

Note that the coefficients f_k' introduced in (3.71) are well defined for every $k = 1, \dots, N$, but in the above computations we are only interested in the ones associated with the critical nodes. For example,

$$f_1' = \begin{cases} \frac{1}{n_1} \, \frac{1}{\bar{d}_1 B^{-(\bar{d}_1-1)}}\, \gamma_U, & \text{if } \bar{d}_1 < d^*, \\ \frac{1}{n_1 \bar{d}_1 B^{-(\bar{d}_1-1)} + (c-\rho_U)}\, \gamma_U, & \text{if } \bar{d}_1 = d^*. \end{cases} \tag{3.83}$$

(III) $\beta \in [\frac{1}{d^*-1}, \infty)$. If the first node $v_1 \in V$ is subcritical, then its nucleation time does not influence the mean queue lengths by much, as seen in (I). If v_1 is critical, then the mean queue lengths decrease but remain of order r, as seen in (II). We therefore assume that v_1 is supercritical. Then $\mathbb{E}_u[\bar{\tau}_1] = \frac{\gamma_U}{c-\rho_U} r\,[1+o(1)]$, as $r \to \infty$. Indeed, from Theorem 2.1.6 we know that the mean nucleation time of a supercritical fork is given by the expected time it takes for the queue length to hit zero. This holds for every supercritical node in V and therefore it is true also for $\mathbb{E}_u[\bar{\tau}_1]$. Hence, the mean queue lengths at nodes in U after node $v_1 \in V$ activates are

$$\mathbb{E}_u[Q_U(\bar{\tau}_1)] = \mathbb{E}_u[\gamma_U r - (c-\rho_U)\bar{\tau}_1] = \gamma_U r - (c-\rho_U)\mathbb{E}_u[\bar{\tau}_1] = o(r), \qquad r \to \infty. \tag{3.84}$$

More generally, the mean queue lengths become $o(r)$ as soon as the first supercritical node activates, independently of which nodes activated before. Thus, after any step k the mean queue length at a node in U, if any supercritical node activated, is

$$\mathbb{E}_u[Q_U^k] = o(r), \qquad r \to \infty. \tag{3.85}$$

□

In summary, we have shown that if a subcritical node activates, then we do not change the mean queue lengths at nodes in U by much: they only decrease by a factor $o(1)$. On the other hand, if a critical node activates, then the mean queue lengths drop significantly, but still remain of order r. Finally, if a supercritical node activates,

then the mean queue lengths become $o(r)$, and remain so during all the successive nucleations. With the help of (3.71) we know how to relate the mean next nucleation times of the forks to the mean updated queue lengths after each activation. Hence we know that, once a node that contributes order r to the total mean transition time activates, we can ignore the contribution of all the previous and all the subsequent subcritical nodes. Once a supercritical node activates, we can ignore the contribution of all the subsequent nodes, since their queue lengths are $o(r)$.

§3.5 Analysis of the algorithm

In Section 3.5.1 we describe how the algorithm acts on an arbitrary bipartite graph. (In Section 3.2.4 we already illustrated this via an example.) In Section 3.5.2 we prove the greediness and the consistency of the algorithm.

§3.5.1 Recursion

Consider the graph $G = G_1 = ((U_1, V_1), E_1)$. The first node activating in V_1 is the one with the least degree, since this requires the least number of nodes in U_1 to become simultaneously inactive. Since the expected time for m nodes in U_1 to become simultaneously inactive is of order $r^{1\wedge\beta(m-1)}$, with high probability as $r \to \infty$, the first node that activates in V_1 is v_{Y_1} such that $d(v_{Y_1}) = \bar{d}_1 = \min_{v\in V_1} d(v)$, where $d(v)$ denotes the degree of node v in the graph G_1. We make the algorithm pick as first node a node v_{Y_1} with least degree in V_1. If there are multiple nodes with the same least degree, then the algorithm chooses one of them uniformly at random. If the least degree $\bar{d}_1$ is such that $\beta(\bar{d}_1 - 1) > 1$, then the algorithm chooses a node uniformly at random among all nodes in V_1. Let $G_1'(U_1', V_1')$ be the complete bipartite subgraph of G_1 with $U_1' = \{u \in U_1 : uv_{Y_1} \text{ is an edge of } G_1\}$ and $V_1' = \{v_{Y_1}\}$. According to Theorem 2.1.6, the associated nucleation time $\mathcal{T}_{v_{Y_1}}^{Q^0}$ satisfies

$$\mathbb{E}_u[\mathcal{T}_{v_{Y_1}}^{Q^0}] = F^1\,(Q_U^0)^{1\wedge\beta(\bar{d}_1-1)}\,[1 + o(1)], \qquad r \to \infty. \tag{3.86}$$

Reasoning as above, we see that the algorithm picks as second node a node v_{Y_2} with the least number of active neighbors left in G. Consider the bipartite graph $G_2(U_2, V_2)$ with $U_2 = U_1 \setminus U_1'$ and $V_2 = V_1 \setminus V_1' = V_1 \setminus \{v_{Y_1}\}$. If we denote by $d_2(v)$ the degree of a node $v \in V_2$ in G_2, then v_{Y_2} is such that $d_2(v_{Y_2}) = \bar{d}_2 = \min_{v\in V_2} d_2(v)$. If there are multiple nodes with the same least degree, then the algorithm again chooses one uniformly at random. If the least degree $\bar{d}_2$ is such that $\beta(\bar{d}_2 - 1) > 1$, then we choose a node uniformly at random among all nodes in V_2. Let $G_2'(U_2', V_2')$ be the complete bipartite subgraph of G with $U_2' = \{u \in U_2 \colon uv_{Y_2} \text{ is an edge of } G_2\}$ and $V_2' = \{v_{Y_2}\}$. The associated nucleation time $\mathcal{T}_{v_{Y_2}}^{Q^1}$ satisfies

$$\mathbb{E}_u[\mathcal{T}_{v_{Y_2}}^{Q^1}] = F^2\,(Q_U^1)^{1\wedge\beta(\bar{d}_2-1)}\,[1 + o(1)], \qquad r \to \infty, \tag{3.87}$$

Iterating this procedure until all the nodes in V_1 are active, we find one of the paths that the algorithm follows in terms of successive activation of the nodes in V_1.

Note that, depending on the choice the algorithm makes at each step, there may be different paths for the activation.

§3.5.2 Greediness and consistency

We first prove Lemma 3.2.5. After that we prove Propositions 3.2.6–3.2.7.

Proof of Lemma 3.2.5. The proof is by contradiction. Suppose that $d_a^* > d_b^*$. Denote by $d_{k,a}(v)$ and $d_{k,b}(v)$ the degrees of node $v \in V_k$ at step $k = 1, \ldots, N$ in paths a and b, respectively.

We start by considering node $w_1 \in V$ such that, at some step k_1^a in path a, $d_{k_1^a,a}(w_1) = \bar{d}_{k_1^a,a} = d_a^*$. Then $d(w_1) \geq d_a^*$ in G. On the other hand, in path b, when w_1 activates at some step k_1^b, it has degree $d_{k_1^b,b}(w_1) \leq d_b^*$. This implies that some of the edges of w_1 (at least $d_a^* - d_b^*$ edges) have already been processed via previous forks in path b. At least one of these forks must have nucleated before the fork of w_1, in path b but not in path a, say, the fork of w_2. Hence there exists a node $w_2 \in V$ such that, at some step $k_2^b < k_1^b$ in path b, $d_{k_2^b,b}(w_2) \leq d_b^*$. This node has not yet activated by step k_1^a in path a, so it must be that $d_{k_1^a,a}(w_2) \geq d_a^*$, otherwise the algorithm would choose node w_2 before node w_1. Say that node w_2 will activate at step $k_2^a > k_1^a$ in path a. Then, $d(w_2) \geq d_a^*$ in G. As before, this implies that some of its edges have already been processed with previous forks in path b. Again, at least one of these forks must have nucleated before the fork of w_2, in path b but not in path a, say, the fork of w_3. Hence there exists a node $w_3 \in V$ such that, at some step $k_3^b < k_2^b$ in path b, $d_{k_3^b,b}(w_3) \leq d_b^*$. This node has not yet activated by step k_2^a in path a, nor by step k_1^a, so $d_{k_1^a,a}(w_3) \geq d_{k_1^a,a}(w_1) \geq d_a^*$, otherwise the algorithm would choose node w_3 before node w_1. Hence $d(w_3) \geq d_a^*$ in G.

We can iterate this argument. Since there are only N nodes in V, we get a contradiction after we have considered all the nodes. □

We are now able to prove the greediness and the consistency of the algorithm.

Proof of Proposition 3.2.6. By Lemma 3.2.5, we know that the maximum least degree of a path generated by the algorithm is the smallest possible. We know that the order of the mean transition time along a path is related to d^* and depends on the value of β. Hence, Lemma 3.2.5 implies that the mean transition time along a path generated by the algorithm is the shortest possible, in the sense that it has the smallest order of r possible. □

Proof of Proposition 3.2.7. Lemma 3.2.5 proves equality for any two paths generated by the algorithm. This leads to the same order of the mean transition time. □

Despite the fact that d^* does not depend on which path the algorithm generates, its multiplicity does. Figure 3.4 shows a graph on which the algorithm can generate two different paths with the same maximum least degree but with different multiplicity.

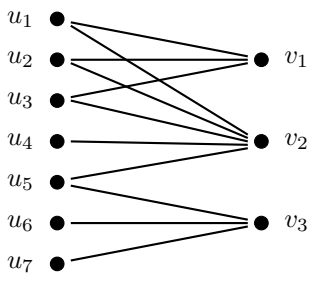


Figure 3.4: The algorithm may generate the path v_1, v_2, v_3 or the path v_3, v_1, v_2 with different multiplicity of d^*.

§3.5.3 Algorithm complexity

The algorithm we constructed can be implemented in different ways according to what we want to compute.

- In order to know the leading order of the mean transition time, it is enough to recover the maximum least degree d^* from the graph. By Proposition 3.2.7 we know that d^* is the same for all paths the algorithm can generate. Hence it is enough to run it once and comparing the value of d^* with the value of β we can immediately determine whether we are in the subcritical, critical or supercritical regime.

  In this case the computational complexity of the algorithm is polynomial in the number of nodes in V, and so the leading order of the mean transition time is quickly determined. More precisely, the algorithm has a complexity of $\mathcal{O}(|U||V|^2)$.

- If we are interested in the precise asymptotics of the mean transition time and in its law as $r \to \infty$, then we need to compute the pre-factor of the leading order term. To do so, we need to run the algorithm multiple times, until all possible paths are generated, in order to recover all the possible sequences $(\bar{d}_k)_{k=1}^N$ and (n_k). A proper approach is to let a (deterministic) depth-first search algorithm run through all possible paths and enumerate them. Theorem 3.3.2 shows that if we know the total mean transition time along each path, then we can recover the mean transition time of the graph.

  In this case the computational complexity of the algorithm is factorial in the number of nodes in V, since it depends in a delicate manner on the architecture of the graph. More precisely, the algorithm has a complexity of $\mathcal{O}(|U||V|^2|V|!)$.

See [99] for a deeper analysis of the algorithm complexity.

§3.6 Proofs of the main results

The aim of this section is to prove the theorems in Section 3.3. In Section 3.6.1 we introduce some further definitions. In Section 3.6.2 we prove Lemmas 3.2.2 and 3.2.11. In the last three sections we prove the three main theorems, respectively: in Section 3.6.3 we prove Theorem 3.3.2, in Section 3.6.4 we prove Theorem 3.3.3, and in Section 3.6.5 we prove Theorem 3.3.5.

§3.6.1 Preparatory results

Consider an arbitrary bipartite graph $G = ((U, V), E)$ with $|V| = N$ and let $v_1, \dots, v_N$ be the nodes in V. The activation path that the network follows is denoted by $v_1^*, \dots, v_N^*$, while the indices of the nodes that the algorithm picks are denoted by $Y_1, \dots, Y_N$ (as in Definition 3.2.1). We want to study the transition time when the network follows a path generated by the algorithm. When conditioning the network on a specific activation order, we can write

$$\{Y_k = i\} = \{v_k^* = v_i\}, \tag{3.88}$$

in the sense that saying that the k-th index Y_k chosen by the algorithm equals i is equivalent to saying that the k-th activating node v_k^* equals v_i.

Definition 3.6.1 (Iteration graph).
For $k = 1, \dots, N$, suppose that $k-1$ nodes in V already activated. Denote by $G_k = ((V_k, U_k), E_k)$ the subgraph of $G = ((U, V), E)$ consisting of:

- $V_k = V \setminus \{\{v_{Y_i}\}_{0<i<k}\} \subseteq V$, the set of nodes in V that have not yet activated;
- $U_k = U \setminus \bigcup_{0<i<k} N(v_{Y_i}) \subseteq U$, the set of nodes in U that are not neighbors of some of the nodes in V that already activated;
- $E_k = \{uv : u \in U_k, v \in V_k\} \subseteq E$, the set of edges between U_k and V_k.

Let $\bar{d}_k$ be the minimum degree of the nodes in V_k and n_k be the number of least degree forks in G_k.

Definition 3.6.2 (Minimum degree subset).
Define the set of nodes with minimum degree in V as

$$M(V) = \{v' \in V \colon d(v') = \min_{v \in V} d(v)\}. \tag{3.89}$$

Lemma 3.6.3 (Probability of choosing the next node).
Given the graph G_k, in the subcritical and critical regimes, the probability that the next node activating in V_k is node v_i is

$$\mathbb{P}(Y_k = i) = \begin{cases} \frac{1}{n_k}, & \text{if } \beta \in (0, \frac{1}{\bar{d}_k - 1}],\ v_i \in M(V_k), \\ 0, & \text{if } \beta \in (0, \frac{1}{\bar{d}_k - 1}],\ v_i \in V_k \setminus M(V_k), \end{cases} \tag{3.90}$$

which depends on the sequence of nodes already active in V.

Proof. By construction, the algorithm picks nodes in $M(V_k)$ before before it picks nodes in $V_k \setminus M(V_k)$. It is therefore enough to count the number of forks of minimum degree at step k, which is n_k. □

§3.6.2 Proof: activation sticks and selects low degrees

We next prove two lemmas from Section 3.2 that will be needed to prove Theorem 3.3.2 in Section 3.6.3.

Proof of Lemma 3.2.2. Recall Definition 3.1.1. We claim that if a node $u \in U$ deactivates and one of its neighbors in V activates at time t_u, then, with high probability as $r \to \infty$, it will not activate anymore after time t_u. The moments when u could possibly activate again are the moments when all its neighbors in V are simultaneously inactive. We consider the worst case scenario when u has only one active neighbor $v \in V$. Denote by t_v the first moment when v deactivates after t_u. This happens many times, since the activity period of a node is described by an exponential variable Z with rate 1. Instead, the inactivity periods are very short, since the nodes in V are very aggressive and the activation rates grow with the queue lengths, which tend to infinity as $r \to \infty$. We consider a time period of length equal to the total transition time, and we assume the transition time to be the longest possible (of order r). Then on average we have a number of possibilities for u to activate that is equal to

$$\frac{\mathbb{E}[\mathcal{T}_G^{Q^0}]}{\mathbb{E}[Z]} = \mathbb{E}[\mathcal{T}_G^{Q^0}] = Cr\,[1 + o(1)], \qquad r \to \infty, \tag{3.91}$$

with C a positive constant. At each of these times, nodes u and v are both inactive and are competing with each other to activate again. Denote by Z_u and Z_v the lengths of the inactivity periods of u and v, respectively. Then $Z_u \simeq \text{Exp}(g_U(Q_u(t_v)))$ and $Z_v \simeq \text{Exp}(g_V(Q_v(t_v)))$ and so, with high probability as $r \to \infty$, node v activates before node u, i.e.,

$$\begin{aligned} \lim_{r\to\infty} \mathbb{P}(Z_v < Z_u) &= \lim_{r\to\infty} \frac{g_V(Q_v(t_v))}{g_U(Q_u(t_v)) + g_V(Q_v(t_v))} = \lim_{r\to\infty} \frac{K' r^{\beta'}}{K r^{\beta} + K' r^{\beta'}} \\ &= \lim_{r\to\infty} \frac{1}{1 + (K/K') r^{-(\beta'-\beta)}} = 1, \end{aligned} \tag{3.92}$$

where we use that $\beta' > \beta$, and K, K' are positive constants.

Note that the queue lengths in U are always of order r, except when we are in the supercritical regime. In this regime we are not interested in the competition between u and v anymore, since we know how long the transition takes. The queue lengths in V start being of order r, increase while u is active and decrease when v is active, but remain of order r. Indeed, if there are other nodes in U that take long enough to activate so that the queue length of v becomes $o(r)$, then we must be in the supercritical regime. In the worst case scenario, nodes u and v compete with each other for the duration of the transition, i.e., order r times. The probability of v winning every competition is

$$\begin{aligned} \lim_{r\to\infty} \mathbb{P}(Z_v < Z_u)^r &= \lim_{r\to\infty} \left(\frac{1}{1 + (K/K') r^{-(\beta'-\beta)}} \right)^{Cr} = \lim_{r\to\infty} \left(e^{-(K/K') r^{-(\beta'-\beta)}} \right)^{Cr} \\ &= \lim_{r\to\infty} e^{-C(K/K') r^{-(\beta'-\beta-1)}} = 1, \end{aligned} \tag{3.93}$$

where we use that $\beta' > \beta + 1$. Hence, with $\mathbb{P}_u$-probability tending to 1 as $r \to \infty$, node u will never win any competition against node v, and hence will remain blocked for the duration of the transition. □

Proof of Lemma 3.2.11. We distinguish between node v being subcritical or critical.

(I) $\beta \in (0, \frac{1}{\bar{d}_k - 1})$. From Theorem 2.1.6 we know the law of the nucleation time for the fork of v, namely,

$$\lim_{r\to\infty} \mathbb{P}_u\left(\frac{\mathcal{T}_v^{Q^{k-1}}}{\mathbb{E}_u[\mathcal{T}_v^{Q^{k-1}}]} > x\right) = \mathcal{P}_1(x) = e^{-x}, \qquad x \in [0,\infty). \tag{3.94}$$

From the same equations we also know the law of the nucleation time for the fork of w, which depends on how β and $d(w)$ are related to each other. It is enough to verify that

$$\lim_{r\to\infty} \mathbb{P}_u\left(\frac{\mathcal{T}_w^{Q^{k-1}}}{\mathbb{E}_u[\mathcal{T}_w^{Q^{k-1}}]} > x\right) = \mathcal{P}(x), \qquad x \in [0,\infty), \tag{3.95}$$

with $\mathcal{P}(x) \uparrow 1$ when $x \downarrow 0$ and $\mathcal{P}(x) \downarrow 0$ when $x \uparrow \infty$. To that end, assume that $\mathcal{T}_v^{Q^{k-1}}$ and $\mathcal{T}_w^{Q^{k-1}}$ deviate from their mean such that $\mathcal{T}_v^{Q^{k-1}} > \mathcal{T}_w^{Q^{k-1}}$. This happens with probability tending to 0 as $r \to \infty$, since the deviations must be of order r. Thus,

$$\lim_{r\to\infty} \mathbb{P}_u\big(\mathcal{T}_v^{Q^{k-1}} > \mathcal{T}_w^{Q^{k-1}}\big) = \lim_{r\to\infty} \mathbb{P}_u\big(\mathcal{X}_{v,v}^{k-1} > \mathcal{X}_{v,w}^{k-1}\big), \tag{3.96}$$

where we abbreviate $\mathcal{X}_{v,w}^{k-1} = \mathcal{T}_w^{Q^{k-1}}/\mathbb{E}_u[\mathcal{T}_v^{Q^{k-1}}]$. For fixed $M < \infty$, we can split the right-hand side of (3.96) without the limit $r \to \infty$ as

$$\begin{aligned}
\mathbb{P}_u\big(\mathcal{X}_{v,v}^{k-1} > \mathcal{X}_{v,w}^{k-1}\big) &= \mathbb{P}_u\big(\mathcal{X}_{v,v}^{k-1} > \mathcal{X}_{v,w}^{k-1} \;\big|\; \mathcal{X}_{v,w}^{k-1} > M\big)\,\mathbb{P}_u\big(\mathcal{X}_{v,w}^{k-1} > M\big) \\
&\quad + \mathbb{P}_u\big(\mathcal{X}_{v,v}^{k-1} > \mathcal{X}_{v,w}^{k-1} \;\big|\; \mathcal{X}_{v,w}^{k-1} \le M\big)\,\mathbb{P}_u\big(\mathcal{X}_{v,w}^{k-1} \le M\big) \\
&\le \mathcal{P}_1(M)\,\mathbb{P}_u\big(\mathcal{X}_{v,w}^{k-1} > M\big) \\
&\quad + \mathbb{P}_u\big(\mathcal{X}_{v,v}^{k-1} > \mathcal{X}_{v,w}^{k-1} \;\big|\; \mathcal{X}_{v,w}^{k-1} \le M\big)\,\mathbb{P}_u\big(\mathcal{X}_{v,w}^{k-1} \le M\big).
\end{aligned} \tag{3.97}$$

Pick $\epsilon > 0$ so small that, for $r > r_0(\epsilon)$,

$$\mathbb{P}_u\big(\mathcal{X}_{v,w}^{k-1} \le M\big) = \mathbb{P}_u\left(\mathcal{X}_{w,w}^{k-1} \le M\frac{\mathbb{E}_u[\mathcal{T}_v^{Q^{k-1}}]}{\mathbb{E}_u[\mathcal{T}_w^{Q^{k-1}}]}\right) \le \mathbb{P}_u\big(\mathcal{X}_{w,w}^{k-1} \le M\epsilon\big). \tag{3.98}$$

Letting $r \to \infty$ followed by $\epsilon \downarrow 0$, we get

$$\lim_{\epsilon\downarrow 0} \lim_{r\to\infty} \mathbb{P}_u\big(\mathcal{X}_{w,w}^{k-1} \le M\epsilon\big) = \lim_{\epsilon\downarrow 0}[1 - \mathcal{P}(M\epsilon)] = 0. \tag{3.99}$$

We can now let $M \to \infty$ and use (3.96)–(3.97) to arrive at

$$\lim_{r\to\infty} \mathbb{P}_u\big(\mathcal{T}_v^{Q^{k-1}} > \mathcal{T}_w^{Q^{k-1}}\big) = 0. \tag{3.100}$$

(II) $\beta = \frac{1}{\bar{d}_k - 1}$. As before, we know the law of the nucleation time for the fork of v and w. As shown in Chapter 2, with high probability as $r \to \infty$, $\mathcal{T}_w^{Q^{k-1}} / \mathbb{E}_u[\mathcal{T}_w^{Q^{k-1}}]$ tends to 1. Moreover, with high probability as $r \to \infty$, any nucleation time of a complete bipartite graph in the critical regime (including the fork of v) is smaller than the transition time of the same graph in the supercritical regime.

□

§3.6.3 Proof: most likely paths

Proof of Theorem 3.3.2. We prove the three statements separately.

(i) Assuming that the network does not follow the algorithm is equivalent to assuming that at some step k with $\beta \in (0, \bar{d}_k - 1]$ a node w that does not have a minimum degree is chosen instead of a node v with degree $\bar{d}_k$. The probability of a group of $d > \bar{d}_k$ nodes becoming simultaneously inactive before a group of $\bar{d}_k$ nodes is equivalent to the probability of w activating before v, which satisfies

$$\lim_{r\to\infty} \mathbb{P}_u\big(\mathcal{T}_w^{Q^{k-1}} < \mathcal{T}_v^{Q^{k-1}}\big) = 0 \tag{3.101}$$

by Lemma 3.2.11. Hence, with high probability as $r \to \infty$, nodes in V activate in a greedy way, as described by the algorithm. By Lemma 3.2.2, we also know that the nodes in U that deactivated remain inactive for the duration of the transition process. Consequently, they do not influence any future activation attempt of the nodes in V, whose activation therefore follows the algorithm. In the supercritical regime, we are only interested in the order of activation of the nodes until the first supercritical node, for which the above reasoning still holds.

(ii) Note that the queues Q^k depend on the sequence of indices $(Y_1, \dots, Y_{k-1})$ describing the order of the activating nodes in V. Indeed, we have seen in Section 3.4.3 that the queues change according to which nodes already activated. Moreover, for $k > 1$, also the probabilities $\frac{1}{n_k}$ depend on the sequence $(Y_1, \dots, Y_{k-1})$. The reader should keep this in mind while going through the proof. The proof evolves in three steps.

1. Denote the graph $G = ((U, V), E)$ by $G_1 = ((U_1, V_1), E_1)$. Write

$$\mathbb{E}_u[\mathcal{T}_G^{Q^0} \mathbb{1}_{\mathcal{E}(a^*)}] = \mathbb{E}_u[\mathcal{T}_{G_1}^{Q^0} \mathbb{1}_{\mathcal{E}(a^*)}] = \sum_{i_1 \colon v_{i_1} \in V_1} \mathbb{E}_u[\mathcal{T}_{G_1}^{Q^0} \mathbb{1}_{\mathcal{E}(a^*)} \mid Y_1 = i_1] \, \mathbb{P}(Y_1 = i_1). \tag{3.102}$$

By Lemma 3.6.3, when $\beta(\bar{d}_1 - 1) \leq 1$ not all the terms in the above sum have positive probability, only the ones corresponding to forks of minimum degree $\bar{d}_1$ (which all have the same probability). Recall that this probability is $\frac{1}{n_1}$. We can write the random variable $\mathcal{T}_{G_1}^{Q^0}$ as sum of three random variables

$$\mathcal{T}_{G_1}^{Q^0} = \bar{\tau}_1 + R_{U_2}^1 + \mathcal{T}_{G_2}^{Q^1}, \tag{3.103}$$

where $G_2 = ((U_2, V_2), E_2)$ is the subgraph with $U_2 = U_1 \setminus N(v_{Y_1})$, $V_2 = V_1 \setminus \{v_{Y_1}\}$ and $E_2 = E_1 \setminus \{(u, v) : u \in N(v_{Y_1})\}$, while $Q^1 = Q(\bar{\tau}_1)$ represents the updated queue lengths. The first variable represents the time it takes for the the first node to activate, the second variable represents the time it takes (after the activation of the first node) to reach the state with all the nodes in U_2 active (see Lemma 3.4.2), while the third variable represents the transition time of the remaining graph when we take the first activating node out. Note that, by Corollary 3.4.4, if we condition the network to follow a path generated by the algorithm with a specific first activating node, then we get

$$\mathbb{E}_u[\bar{\tau}_1 \mid Y_1 = i_1] = f_1 \, \mathbb{E}_u[\mathcal{T}_{v_{i_1}}^{Q^0}], \qquad r \to \infty, \tag{3.104}$$

where f_1 is the factor that arises from considering the minimum of random variables. Also the variable $\mathcal{T}_{G_2}^{Q^1}$ changes accordingly, but with an abuse of notation we may write it in the same way. Thus, with high probability as $r \to \infty$, by Lemma 3.4.2,

$$\begin{aligned}
\mathbb{E}_u\big[\mathcal{T}_{G_1}^{Q^0} \mathbb{1}_{\mathcal{E}(a^*)} \mid Y_1 = i_1\big] &= \mathbb{E}_u\big[(\bar{\tau}_1 + R_{U_2}^1 + \mathcal{T}_{G_2}^{Q^1}) \mid \mathcal{E}(a^*) \cap \{Y_1 = i_1\}\big] \\
&= \mathbb{E}_u[\bar{\tau}_1 \mathbb{1}_{\mathcal{E}(a^*)} \mid Y_1 = i_1] + o(1) \\
&\quad + \mathbb{E}_u\big[\mathcal{T}_{G_2}^{Q^1} \mathbb{1}_{\mathcal{E}(a^*)} \mid Y_1 = i_1\big] \\
&= f_1 \, \mathbb{E}_u\big[\mathcal{T}_{v_{i_1}}^{Q^0} \mathbb{1}_{\mathcal{E}(a^*)}\big] + o(1) \\
&\quad + \mathbb{E}_u\big[\mathcal{T}_{G_2}^{Q^1} \mathbb{1}_{\mathcal{E}(a^*)}\big], \qquad r \to \infty.
\end{aligned} \tag{3.105}$$

We want to analyze the latter in a recursive way. The k-th iteration gives

$$\mathbb{E}_u\big[\mathcal{T}_{G_k}^{Q^{k-1}} \mathbb{1}_{\mathcal{E}(a^*)}\big] = \sum_{i_k \colon v_{i_k} \in V_k} \mathbb{E}_u\big[\mathcal{T}_{G_k}^{Q^{k-1}} \mathbb{1}_{\mathcal{E}(a^*)} \mid Y_k = i_k\big] \, \mathbb{P}(Y_k = i_k). \tag{3.106}$$

2. We can again write the random variable $\mathcal{T}_{G_k}^{Q^{k-1}}$ as sum of three random variables

$$\mathcal{T}_{G_k}^{Q^{k-1}} = \bar{\tau}_k + R_{U_{k+1}}^k + \mathcal{T}_{G_{k+1}}^{Q^k}, \tag{3.107}$$

where $G_{k+1} = ((U_{k+1}, V_{k+1}), E_{k+1})$, $U_{k+1} = U_k \setminus g(v_{Y_k})$, $V_{k+1} = V_k \setminus \{v_{Y_k}\}$, $E_{k+1} = E_k \setminus \{(u, v) \colon u \in g(v_{Y_k})\}$, while $Q^k = Q(\sum_{j=1}^{k-1} \mathcal{T}_{v_{i_j}}^{Q^{j-1}})$. By Corollary 3.4.4, we again have that

$$\mathbb{E}_u[\bar{\tau}_k \mid Y_k = i_k] = f_k \, \mathbb{E}_u\big[\mathcal{T}_{v_{i_k}}^{Q^{k-1}}\big], \qquad r \to \infty, \tag{3.108}$$

and also the variable $\mathcal{T}_{G_{k+1}}^{Q^k}$ changes accordingly when it is conditioned (again, with an abuse of notation we write it in the same way). With high probability

as $r \to \infty$, the conditional expectation in (3.106) can be written as

$$\begin{aligned}\mathbb{E}_u\big[\mathcal{T}_{G_k}^{Q^{k-1}}\mathbb{1}_{\mathcal{E}(a^*)} \mid Y_k = i_k\big] &= \mathbb{E}_u\big[(\mathcal{T}_{v_{Y_k}}^{Q^{k-1}} + R_{U_{k+1}}^k + \mathcal{T}_{G_{k+1}}^{Q^k})\mathbb{1}_{\mathcal{E}(a^*)} \mid Y_k = i_k\big] \\ &= \mathbb{E}_u\big[\mathcal{T}_{v_{Y_k}}^{Q^{k-1}}\mathbb{1}_{\mathcal{E}(a^*)} \mid Y_k = i_k\big] + o(1) \\ &\quad + \mathbb{E}_u\big[\mathcal{T}_{G_{k+1}}^{Q^k}\mathbb{1}_{\mathcal{E}(a^*)} \mid Y_k = i_k\big] \\ &= f_k\,\mathbb{E}_u\big[\mathcal{T}_{v_{i_k}}^{Q^{k-1}}\mathbb{1}_{\mathcal{E}(a^*)}\big] + o(1) \\ &\quad + \mathbb{E}_u[\mathcal{T}_{G_{k+1}}^{Q^k}\mathbb{1}_{\mathcal{E}(a^*)}\big], \qquad r \to \infty.\end{aligned} \tag{3.109}$$

At each iteration the conditional expectation reduces to a sum of three terms: the first term represents the expected time it takes to switch the following node on (adjusted by a factor that keeps track of the fact that the node activates before the other nodes), the second term represents the expected time it takes (after the previous node activation) to reach the state with all the nodes remaining in U active, while the third term represents the mean transition time of the remaining network when we take the following activating node out.

3. Note that, for each $k = 1, \dots, N$, the graph G_{k+1} depends on the sequence of indices $(Y_1, \dots, Y_k)$. Moreover, we know that also the queue lengths Q^k depend on the indices $(Y_1, \dots, Y_{k-1})$. Thus, all the conditional expectations depend on the sequence of indices of activated nodes. By Lemma 3.6.3, the first iteration comes with a probability $\frac{1}{n_1}$ of choosing the first node activating, while each iteration with $k > 1$ comes with a probability $\frac{1}{n_k}$, also depending on the sequence $(Y_1, \dots, Y_{k-1})$. After $k = 2$ steps, using (3.106) and (3.109), with high probability as $r \to \infty$,

$$\begin{aligned}&\mathbb{E}_u\big[\mathcal{T}_{G_1}^{Q^0}\mathbb{1}_{\mathcal{E}(a^*)}\big] \\ &= \sum_{i_1:\, v_{i_1} \in V_1} \frac{1}{n_1}\left(f_1\,\mathbb{E}_u\big[\mathcal{T}_{v_{i_1}}^{Q^0}\mathbb{1}_{\mathcal{E}(a^*)}\big] + o(1) + \mathbb{E}_u\big[\mathcal{T}_{G_2}^{Q^1}\mathbb{1}_{\mathcal{E}(a^*)}\big]\right) \\ &= \sum_{i_1:\, v_{i_1} \in V_1} \frac{1}{n_1}\left(f_1\,\mathbb{E}_u\big[\mathcal{T}_{v_{i_1}}^{Q^0}\mathbb{1}_{\mathcal{E}(a^*)}\big] + o(1)\right. \\ &\qquad \left. + \sum_{i_2:\, v_{i_2} \in V_2} \frac{1}{n_2}\left(f_2\,\mathbb{E}_u\big[\mathcal{T}_{v_{i_2}}^{Q^1}\mathbb{1}_{\mathcal{E}(a^*)}\big] + o(1) + \mathbb{E}_u\big[\mathcal{T}_{G_3}^{Q^2}\mathbb{1}_{\mathcal{E}(a^*)}\big]\right)\right), \quad r \to \infty.\end{aligned} \tag{3.110}$$

After N steps, the last node in V activates and the conditional expectation

becomes

$$\mathbb{E}_u\big[\mathcal{T}_{G_N}^{Q^{N-1}}\mathbb{1}_{\mathcal{E}(a^*)}\big] = \sum_{i_N:\, v_{i_N}\in V_N} \frac{1}{n_N}\Big(f_N\,\mathbb{E}_u\big[\mathcal{T}_{v_{i_N}}^{Q^{N-1}}\mathbb{1}_{\mathcal{E}(a^*)}\big] + \mathbb{E}_u[R_{U_{N+1}}^N] + \mathbb{E}_u\big[\mathcal{T}_{G_{N+1}}^{Q^N}\mathbb{1}_{\mathcal{E}(a^*)}\big]\Big)$$
$$= \sum_{i_N:\, v_{i_N}\in V_N} \frac{1}{n_N} f_N\,\mathbb{E}_u\big[\mathcal{T}_{v_{i_N}}^{Q^{N-1}}\mathbb{1}_{\mathcal{E}(a^*)}\big], \qquad r\to\infty. \tag{3.111}$$

Indeed, as soon as the last node in V activates, we are actually done and we are not interested in what happens after. We can set $R_{U_{N+1}}^N = 0$ and we have $V_{N+1} = \emptyset$, which implies $\mathbb{E}_u[\mathcal{T}_{G_{N+1}}^{Q^N}] = 0$. Thus, we have arrived at (3.26).

(iii) The claim follows from steps analogous to the ones in (ii), given any path $a \in \mathcal{A}$ that the algorithm generates.

□

§3.6.4 Proof: mean of the transition time

Proof of Theorem 3.3.3. Recall that, in the subcritical and critical regimes, we are computing the mean transition time conditioned on the event that the nucleation follows a fixed path $a = (v_1, \dots, v_N) \in \mathcal{A}$. We again distinguish between the three regimes.

(I) $\beta \in (0, \frac{1}{d^*-1})$: subcritical regime. Every term in the sum is $o(r)$, which means that the significant terms are the ones with $\bar{d}_k = d^*$ only. The pre-factors of these terms are given by subcritical forks, and so

$$\mathbb{E}_u\big[\mathcal{T}_G^{Q^0}\big] = \sum_{k:\,\bar{d}_k=d^*} f_k\,\mathbb{E}_u\big[\mathcal{T}_{v_k}^{Q^{k-1}}\big] = \sum_{k:\,\bar{d}_k=d^*} f_k\,\frac{\mathbb{E}_u[Q_U^{k-1}]^{\beta(d^*-1)}}{d^*B^{-(d^*-1)}}\,[1+o(1)]$$
$$= \sum_{k:\,\bar{d}_k=d^*} f_k\,\frac{\gamma_U^{\beta(d^*-1)}}{d^*B^{-(d^*-1)}}\,r^{\beta(d^*-1)}\,[1+o(1)], \qquad r\to\infty, \tag{3.112}$$

with $f_k = \frac{1}{n_k}$. The last equality comes from (3.72) in Theorem 3.4.8.

(II) $\beta = \frac{1}{d^*-1}$: critical regime. Every term in the sum is $o(r)$, except the terms with $\bar{d}_k = d^*$, which is of order r. The significant terms are the ones with $\bar{d}_k = d^*$

only. The pre-factors of these terms are given by critical forks, and so

$$\mathbb{E}_u\big[\mathcal{T}_G^{Q^0}\big] = \sum_{k:\,\bar{d}_k=d^*} f_k\,\mathbb{E}_u\big[\mathcal{T}_{v_k}^{Q^{k-1}}\big] = \sum_{k:\,\bar{d}_k=d^*} f_k\,\frac{\mathbb{E}_u[Q_U^{k-1}]}{d^*B^{-(d^*-1)}+(c-\rho_U)}\,[1+o(1)]$$
$$= \sum_{k:\,\bar{d}_k=d^*} f_k\,\frac{\gamma_U^{(k-1)}}{d^*B^{-(d^*-1)}+(c-\rho_U)}\,r\,[1+o(1)], \qquad r\to\infty, \tag{3.113}$$

with $\gamma_U^{(k-1)}$ defined in (3.74) in Theorem 3.4.8 and

$$f_k = \frac{\bar{d}_k B^{-(\bar{d}_k-1)}+(c-\rho_U)}{n_k\bar{d}_k B^{-(\bar{d}_k-1)}+(c-\rho_U)}. \tag{3.114}$$

(III) $\beta\in(\frac{1}{d^*-1},\infty)$: supercritical regime. Denote by $v_{\rm sc}$ the first supercritical node. We know from (3.76) in Theorem 3.4.8 that, after $v_{\rm sc}$ activates, the queue lengths become negligible (order $o(r)$), and the mean transition time is given by the expected time it takes for them to hit zero, i.e.,

$$\mathbb{E}_u\big[\mathcal{T}_G^{Q^0}\big] = \frac{\gamma_U}{c-\rho_U}\,r\,[1+o(1)], \qquad r\to\infty. \tag{3.115}$$

□

§3.6.5 Proof: law of the transition time

Proof of Theorem 3.3.5. We again distinguish between the three regimes.

(I) $\beta\in(0,\frac{1}{d^*-1})$: subcritical regime. Recall that the significant terms in the sum for the mean transition time are those coming from nodes with degree $\bar{d}_k=d^*$ with $d^*<\frac{1}{\beta}+1$. There are $m^a_{\rm sub}$ such terms, where $m^a_{\rm sub}$ depends on the path $a\in\mathcal{A}$, and each term comes with a multiplicative factor f_k. We can write the transition time along path a divided by its mean as

$$\frac{\mathcal{T}_G^{Q^0}|A_a}{\mathbb{E}_u[\mathcal{T}_G^{Q^0}|A_a]} = \frac{\sum_{k=1}^N \bar{\tau}_k + \sum_{k=2}^N R_{U_k}^{k-1}}{\mathbb{E}_u[\mathcal{T}_G^{Q^0}|A_a]}$$
$$= \frac{\sum_{k':\ \bar{d}_{k'}=d^*}\bar{\tau}_{k'} + \sum_{k'':\ \bar{d}_{k''}<d^*}\bar{\tau}_{k''} + \sum_{k=2}^N R_{U_k}^{k-1}}{\mathbb{E}_u[\mathcal{T}_G^{Q^0}|A_a]}. \tag{3.116}$$

We know that the law of a sum of independent random variables has a density given by the convolution of their densities. Here the nucleation times and the return times can be considered as independent, since they only depend on the queue lengths, which remain close to the initial value in the subcritical regime.

There are three types of sums in the numerator of the last line of (3.116). The first type of sum is of the form $\bar{\tau}_{k'}/\mathbb{E}_u[\mathcal{T}_G^{Q^0}|A_a]$, with k' such that $\bar{d}_{k'}=d^*$. As

$r \to \infty$, these are the significant terms in the sum, since they are of the same order as the mean transition time. For each of them, i.e., for each k', we have

$$\begin{aligned}\lim_{r\to\infty} \mathbb{P}_u\left(\frac{\bar\tau_{k'}}{\mathbb{E}_u[\mathcal{T}_G^{Q^0}|A_a]} > x\right) &= \lim_{r\to\infty} \mathbb{P}_u\left(\frac{\bar\tau_{k'}}{\mathbb{E}_u[\bar\tau_{k'}]} > \frac{\mathbb{E}_u[\mathcal{T}_G^{Q^0}|A_a]}{\mathbb{E}_u[\bar\tau_{k'}]}\, x\right) \\ &= \exp\left(-\frac{\sum_{i:\ \bar d_i = d^*} \mathbb{E}_u[\bar\tau_i]}{\mathbb{E}_u[\bar\tau_{k'}]}\, x\right) \\ &= \exp\left(-\frac{\sum_{i:\ \bar d_i = d^*} f_i}{f_{k'}}\, x\right), \qquad x \in [0,\infty),\end{aligned} \tag{3.117}$$

where in the second step we use Proposition 3.4.6. We write the density as

$$\mathcal{P}_{\text{sub}}^{f_{k'}, S_{m^a_{\text{sub}}}}(x) = \frac{S_{m^a_{\text{sub}}}}{f_{k'}} \exp\left(-\frac{S_{m^a_{\text{sub}}}}{f_{k'}}\, x\right), \qquad x \in [0,\infty), \tag{3.118}$$

with

$$S_{m^a_{\text{sub}}} = \sum_{i:\ \bar d_i = d^*} f_i. \tag{3.119}$$

The second type of sum is of the form $\bar\tau_{k''}/\mathbb{E}_u[\mathcal{T}_G^{Q^0}|A_a]$, with k'' such that $\bar d_{k''} < d^*$. As $r \to \infty$, these are negligible, since they are of smaller order than the mean transition time. For each of them, i.e., for each k'', we have

$$\lim_{r\to\infty} \mathbb{P}_u\left(\frac{\bar\tau_{k''}}{\mathbb{E}_u[\mathcal{T}_G^{Q^0}|A_a]} > x\right) = \lim_{r\to\infty} \mathbb{P}_u\left(\frac{\bar\tau_{k''}}{\mathbb{E}_u[\bar\tau_{k''}]} > \frac{\mathbb{E}_u[\mathcal{T}_G^{Q^0}|A_a]}{\mathbb{E}_u[\bar\tau_{k''}]}\, x\right),\ x \in [0,\infty), \tag{3.120}$$

and the density is δ_0, the Dirac function at 0. The third type of sum is of the form $R_{U_k}^{k-1}/\mathbb{E}_u[\mathcal{T}_G^{Q^0}|A_a]$, with $k = 2, \dots, N$. As $r \to \infty$, these are also negligible, since they are $o(1)$ by Lemma 3.4.2, and hence their density is also δ_0.

The density of $\mathcal{T}_G^{Q^0}|A_a/\mathbb{E}_u[\mathcal{T}_G^{Q^0}|A_a]$ is given by the convolution of the densities of the three types of terms. Since δ_0 gives the identity for the convolution, we can write

$$\lim_{r\to\infty} \mathbb{P}_u\left(\frac{\mathcal{T}_G^{Q^0}}{\mathbb{E}_u[\mathcal{T}_G^{Q^0}|A_a]} > x \mid A_a\right) = \int_x^\infty \left(\circledast_{k'=1}^{m^a_{\text{sub}}} \mathcal{P}_{\text{sub}}^{f_{k'}, S_{m^a_{\text{sub}}}}\right)(y)dy, \quad x \in [0,\infty), \tag{3.121}$$

and we can rename the index k' by k.

(II) $\beta = \frac{1}{d^*-1}$: critical regime. For two reasons we do not know how to handle this regime: (a) We do not know the law of the next nucleation times because Proposition 3.4.6 only holds in the subcritical regime. (b) The next nucleation times are dependent random variables, and so convolution is no longer relevant.

(III) $\beta \in (\frac{1}{d^*-1}, \infty)$: supercritical regime. Recall that $T_U = \frac{\gamma_U}{c-\rho_U} r\,[1 + o(1)]$ as $r \to \infty$. The law of the transition time is given by $\mathcal{P}_3(x)$ from Theorem 2.1.6.

Indeed, the mean transition time is the expected time it takes for the queue lengths in U to hit zero and. With high probability as $r \to \infty$, the transition does not occur before or after its mean. Since

$$\lim_{r\to\infty} \mathbb{P}_u\Big(\mathcal{T}_G^{Q^0} > \mathbb{E}_u\big[\mathcal{T}_G^{Q^0}\big]\Big) = \lim_{r\to\infty} \mathbb{P}_u\big(\mathcal{T}_G^{Q^0} > T_U\big) = 0, \tag{3.122}$$

we can write

$$\lim_{r\to\infty} \mathbb{P}_u\left(\frac{\mathcal{T}_G^{Q^0}}{\mathbb{E}_u[\mathcal{T}_G^{Q^0}]} > x\right) = 0, \qquad x \in [1,\infty). \tag{3.123}$$

We also have

$$\lim_{r\to\infty} \mathbb{P}_u\left(\frac{\mathcal{T}_G^{Q^0}}{\mathbb{E}_u[\mathcal{T}_G^{Q^0}]} > x\right) = 1, \qquad x \in [0,1). \tag{3.124}$$

Hence the density is the Dirac function at 1.

□

§C Appendix: minimum of independent forks

In this appendix we compute the mean next nucleation time in the situation where the forks competing for nucleation are disjoint, i.e., they have no nodes in common. Recall that, in the subcritical regime, the nucleation time of a fork is given by an exponential random variable, while in the critical regime it is given by a "polynomial" random variable, in the sense that its law is truncated polynomial.

§C.1 Subcritical regime: exponential random variables

Let $X_1, \dots, X_n$ be i.i.d. exponential random variables with rate λ and let their minimum be $Z = \min\{X_1, \dots, X_n\}$. Then

$$\mathbb{P}(Z > t) = \mathbb{P}(X_1 > t, \dots, X_n > t) = \mathbb{P}(X_1 > t)^n = e^{-n\lambda t}. \tag{3.125}$$

Hence, Z is an exponential random variable with rate $n\lambda$, and we have

$$\mathbb{E}[Z] = \frac{1}{n\lambda} = \frac{1}{n}\,\mathbb{E}[X_1]. \tag{3.126}$$

If we consider $X_1, \dots, X_{n_k}$ to be the nucleation times of disjoint forks of degree $\bar{d}_k$, and Z to be the next nucleation time at step k, then we get

$$\mathbb{E}_u[\bar{\tau}_k] = f_k^{\mathrm{iid}}\,\mathbb{E}_u\big[\mathcal{T}_{v_k^\star}^{Q^{k-1}}\big], \qquad r \to \infty, \tag{3.127}$$

with $f_k^{\mathrm{iid}} = \frac{1}{n_k}$.

§C.2 Critical regime: polynomial random variables

Let $X_1, \dots, X_n$ be i.i.d. polynomial random variables such that

$$\mathbb{P}\Big(\frac{X_i}{\mathbb{E}[X_i]} > x\Big) = \begin{cases} (1-Cx)^{\frac{1-C}{C}}, & \text{if } x \in [0, \frac{1}{C}), \\ 0, & \text{if } x \in [\frac{1}{C}, \infty), \end{cases} \quad i = 1, \dots, n, \tag{3.128}$$

with

$$C = \frac{c - \rho_U}{\bar{d}_k B^{-(\bar{d}_k - 1)} + (c - \rho_U)}. \tag{3.129}$$

Let $Z = \min\{X_1, \dots, X_n\}$. Then, for $t = x\,\mathbb{E}[X_i]$,

$$\mathbb{P}(X_i > t) = \begin{cases} \big(1 - \frac{C}{\mathbb{E}[X_i]}t\big)^{\frac{1-C}{C}}, & \text{if } t \in [0, \frac{\mathbb{E}[X_i]}{C}), \\ 0, & \text{if } t \in [\frac{\mathbb{E}[X_i]}{C}, \infty), \end{cases} \quad i = 1, \dots, n. \tag{3.130}$$

Abbreviate $C = \frac{C_1}{C_1 + C_2}$, where $C_1 = c - \rho_U$ and $C_2 = \bar{d}_k B^{-(\bar{d}_k - 1)}$. Then the exponent $\frac{1-C}{C}$ becomes $\frac{C_2}{C_1}$. We have

$$\begin{aligned} \mathbb{P}(Z > t) &= \mathbb{P}(X_1 > t, \dots, X_n > t) \\ &= \mathbb{P}(X_1 > t)^n = \begin{cases} \big(1 - \frac{C}{\mathbb{E}[X_i]}t\big)^{n\frac{C_2}{C_1}}, & \text{if } t \in [0, \frac{\mathbb{E}[X_1]}{C}), \\ 0, & \text{if } t \in [\frac{\mathbb{E}[X_1]}{C}, \infty). \end{cases} \end{aligned} \tag{3.131}$$

The density function of Z is

$$f_z(t) = \frac{d}{dt}\big[1 - \mathbb{P}(Z > t)\big] = \begin{cases} \frac{C}{\mathbb{E}[X_1]} n \frac{C_2}{C_1} \big(1 - \frac{C}{\mathbb{E}[X_1]}t\big)^{n\frac{C_2}{C_1} - 1}, & \text{if } t \in [0, \frac{\mathbb{E}[X_1]}{C}), \\ 0, & \text{if } t \in [\frac{\mathbb{E}[X_1]}{C}, \infty). \end{cases} \tag{3.132}$$

Hence

$$\mathbb{E}[Z] = \int_0^{\frac{\mathbb{E}[X_1]}{C}} f_Z(t)\, t\, dt = \frac{C}{\mathbb{E}[X_1]} n \frac{C_2}{C_1} \int_0^{\frac{\mathbb{E}[X_1]}{C}} \Big(1 - \frac{C}{\mathbb{E}[X_1]}t\Big)^{n\frac{C_2}{C_1} - 1} t\, dt. \tag{3.133}$$

Substituting $u = 1 - \frac{C}{\mathbb{E}[X_1]}t$, we get

$$\begin{aligned} \mathbb{E}[Z] &= \frac{\mathbb{E}[X_1]}{C} n \frac{C_2}{C_1} \int_0^1 u^{n\frac{C_2}{C_1} - 1}(1-u)\, du \\ &= \frac{\mathbb{E}[X_1]}{C} n \frac{C_2}{C_1} \left[\int_0^1 u^{n\frac{C_2}{C_1} - 1}\, du - \int_0^1 u^{n\frac{C_2}{C_1}}\, du\right] = \frac{\mathbb{E}[X_1]}{C} n \frac{C_2}{C_1} \left[\frac{1}{n\frac{C_2}{C_1}} - \frac{1}{n\frac{C_2}{C_1} + 1}\right] \\ &= \frac{\mathbb{E}[X_1]}{C} n \frac{C_2}{C_1} \left[\frac{1}{n\frac{C_2}{C_1}(n\frac{C_2}{C_1} + 1)}\right] = \frac{\mathbb{E}[X_1]}{C} \left[\frac{1}{n\frac{C_2}{C_1} + 1}\right] = \mathbb{E}[X_1] \frac{C_1 + C_2}{nC_2 + C_1} \\ &= \frac{\bar{d}_k B^{-(\bar{d}_k - 1)} + (c - \rho_U)}{n\bar{d}_k B^{-(\bar{d}_k - 1)} + (c - \rho_U)}\, \mathbb{E}[X_1]. \end{aligned} \tag{3.134}$$

If we consider $X_1, \dots, X_{n_k}$ to be the nucleation times of disjoint forks of degree $\bar{d}_k$, and Z to be the next nucleation time at step k, then we get

$$\mathbb{E}_u[\bar{\tau}_k] = f_k^{\text{iid}}\, \mathbb{E}_u[\mathcal{T}_{v_k^*}^{Q^{k-1}}], \qquad r \to \infty, \tag{3.135}$$

with

$$f_k^{\text{iid}} = \frac{\bar{d}_k B^{-(\bar{d}_k - 1)} + (c - \rho_U)}{n_k \bar{d}_k B^{-(\bar{d}_k - 1)} + (c - \rho_U)}. \tag{3.136}$$

CHAPTER 4
Dynamic bipartite interference graphs

This chapter is based on:
M. Sfragara. *Adding edge dynamics to wireless random-access networks.* Preprint, 2020.

Abstract

We consider random-access networks with nodes representing servers with queues. The nodes can be either active or inactive: a node deactivates at unit rate, while it activates a rate that depends on its queue length, provided none of its neighbors is active. In order to model the effects of user mobility in wireless networks, we analyze dynamic interference graphs where the edges are allowed to appear and disappear over time. We focus on bipartite graphs, and study the transition time between the two states where one half of the network is active and the other half is inactive, in the limit as the queues become large. Depending on the speed of the dynamics, we are able to obtain a rough classification of the effects of the dynamics on the transition time.

§4.1 Introduction and main results

This chapter is a continuation of Chapters 2–3. We introduce an edge dynamics on the bipartite interference graph, by allowing edges to appear and disappear over time. This represents a natural basic model to capture the effects of user mobility in wireless networks.

In Section 4.1.1 we motivate our interest in adding edge dynamics to random-access network models. In Section 4.1.2 we describe the setting and the mathematical model of interest in this chapter by specifying edge dynamics. In Section 4.1.3 we state out main results for the mean transition time with dynamics. In Section 4.1.4 we explain the main idea behind our analysis and give an outline of the remainder of the chapter.

§4.1.1 Motivation and background

User mobility is one of the major features in wireless networks. Different mobility patterns can be distinguished (pedestrians, vehicles, aerial, dynamic medium, robot, and outer space motion) and mathematical models can be developed in order to generalize such patterns and analyze their characteristics. Understanding the effects of user mobility in wireless networks is crucial in order to design efficient protocols and improve the performance of the network.

For example, consider radio communication protocols, for which central radio stations are used as base-stations for transmitting radio signals. The radio landscape is partitioned into cells and in each cell a station serves the users in its vicinity. In such cellular networks the users may be either stationary or mobile. User mobility leads to problems of handover: when a user moves from one cell to another, the transmitting signal has to be handed over from one station to another in order to ensure continuity of service and seamless mobility. If not enough capacity is available in the adjacent cell, then the transmission might be interrupted. Imagine that a node transmitting to particular station moves away from its cell and reaches a cell where another station serves for transmissions. Although initially the node interferes with a specific group of nodes sharing the same initial station, after the node has moved it interferes with the nodes in the new cell sharing the new station. In a similar fashion, imagine a network where nodes represent transmitter-receiver pairs. The signal of a node interferes with the signals of the nodes in its vicinity. Hence the protocol allows only one of the interfering nodes to transmit alternately. When allowing node mobility, we get new groups of nodes interfering with each other depending on their vicinity. We are therefore dealing with a network whose interference graph changes over time.

To the best of our knowledge, random-access models with user mobility in the context of interference graphs have so far not been considered in the literature. All the studies we are aware of that have examined the impact of user mobility in wireless networks are concerned with handover mechanisms (see [89], [98]), so-called opportunistic scheduling algorithms (see [11], [101]), capacity issues in ad hoc and cellular networks (see [15], [56]), and flow-level performance (see [7], [8], [16], [96]).

In this chapter we investigate a dynamic version of the random-access protocols in order to try to capture some features of user mobility in wireless networks. A natural paradigm for constructing *dynamic interference graphs* would be to use geometric graphs, such as unit-disk graphs, with node mobility, where each node follows a random trajectory and experiences interference from all nodes within a certain distance. A feasible state of the interference graph would then be generated by a specific instance of the geometric graph. We follow a different approach and, with an explorative intention, we consider a model where edges are allowed to appear and disappear from the graph according to i.i.d. Poisson clocks placed on each edge. We find different results for the mean transition time depending on the speed of the dynamics. The evolution of the network is captured by a continuous-time Markov process that keeps track of how the state, the queue length and the number of active neighbors change for each node.

We focus on queue-based activation rates, in line with the models and the results in Chapters 2–3. This leads to two level of complexity, driven by the queue dependencies of the activation rates and by the edge dynamics. In the appendix we briefly consider a simplified version of the model where the activation rates are fixed.

§4.1.2 Setting

We refer to Section 1.1.5 for a general introduction to the mathematical model. In this chapter we add an extra dynamics to the model.

We are interested in analyzing the behavior of the network when we allow the interference graph to change over time. We mainly focus on the model with queue-based activation rates and assume these rates to satisfy Definitions 1.1.4 and 3.1.1.

Definition 4.1.1 (Dynamic interefence graphs).
We say that the interference graph is *dynamic* when the edges appear and disappear according to a continuous-time flip process. Consider the dynamic bipartite interference graph $G(\cdot) = (U \sqcup V, E(\cdot))$, where $U \sqcup V$ is the set of nodes, with $|U| = M$ and $|V| = N$, and $E(t)$ is the set of edges that are present between nodes in U and nodes in V at time t. The number of edges $|E(\cdot)|$ changes over time and can vary from a minimum of 0 to a maximum of MN. We set $G(0) = G$, where G is the initial bipartite graph. We denote by $G_{MN} = (U \sqcup V, E_{MN})$ the complete bipartite graph associated to (U, V) and, for every edge $e \in E_{MN}$, at time t we define the Bernoulli random variable $Y_e(t)$ as

$$Y_e(t) = \begin{cases} 0, & \text{if } e \notin E(t), \\ 1, & \text{if } e \in E(t). \end{cases} \tag{4.1}$$

In other words, $Y_e(t) = 0$ if edge e is not present in the graph at time t, while $Y_e(t) = 1$ if it is present. The *joint edge activity state* at time t is denoted by

$$Y(t) = \{Y_e(t)\}_{e \in E_{MN}} \tag{4.2}$$

and is an element of the state space

$$\mathcal{Y} = \left\{Y \in \{0,1\}^{U \times V}\right\}. \tag{4.3}$$

The degree of node v at time t is denoted by $d_v(t)$.

We model the dynamics of the graph in the following way. If an edge is not present, then it *appears* according to a Poisson clock with rate λ, independently of the other edges. If an edge is present, then it *disappears* according to a Poisson clock with rate λ, independently of the other edges. This is equivalent to having a system of i.i.d. Poisson clocks with rate λ on the edges and letting an edge change its state every time its clock ticks. In order to study how the edge dynamics affects the transition time, we consider Poisson clocks with rates $\lambda = \lambda(r)$ depending on the parameter r.

Throughout the chapter we use the notation $\prec$, $\succ$ to describe the asymptotic behavior in the limit $r \to \infty$. More precisely, $f(r) \prec g(r)$ means that $f(r) = o(g(r))$ as $r \to \infty$, and $f(r) \succ g(r)$ means that $g(r) = o(f(r))$ as $r \to \infty$.

Remark 4.1.2 (Rates on the edges).
We may allow different rates for the edges to change their state. Denote by $\lambda^+(r)$ and $\lambda^-(r)$ the rates at which edges appear and disappear, respectively. If these are of the same order, then we are in a situation similar to them being equal to $\lambda(r)$. If $\lambda^+(r) \to \infty$ and $\lambda^-(r) \prec \lambda^+(r)$, then, with high probability as $r \to \infty$, in time $o(1)$ the dynamics turns the initial graph into the complete bipartite graph with all the edges present. Analogously, if $\lambda^-(r) \to \infty$ and $\lambda^-(r) \succ \lambda^+(r)$, then, with high probability as $r \to \infty$, in time $o(1)$ the dynamics turns the initial graph into the empty graph with all the edges absent. Both these assumptions do not lead to interesting models. When $\lambda^+(r)$ and $\lambda^-(r)$ are of different order and do not tend to infinity, we have an intermediate situation where at any time t an edge is either present with high probability as $r \to \infty$ or absent with high probability as $r \to \infty$, but the total amount of time the edge has been absent or present, respectively, up to time t is not always negligible.

Remark 4.1.3 (Appearing edge).
When an edge disappears from the graph, the states of the nodes do not change. On the other hand, when an edge appears in the graph, it might appear between two active nodes. In this case, we assume that the active node in U deactivates, since the model does not allow two connected nodes to be simultaneously active. We could study alternative models, where the active node in V deactivates or where the deactivating node is chosen uniformly at random (or with certain probabilities). It is obvious that these alternative models slow down the transition and lead to a possible multiple counting of the time it takes for some nodes in V to activate.

§4.1.3 Main theorem

In Chapter 3 we analyzed the mean transition time for arbitrary bipartite graphs. We introduced a randomized algorithm that takes as input the graph and gives as output all the possible activation paths for nodes in V. We showed that, depending on the value of β, the transition exhibits a *subcritical regime*, a *critical regime* and a *supercritical regime*. Given a graph, the algorithm uniquely identifies the value

$$d^* = \max_{1 \le k \le N} \bar{d}_k, \tag{4.4}$$

which determines the leading order of the mean transition time and together with β determines the regime we are in. We were also able to identify the law of the transition time divided by its mean (in the subcritical and supercritical regime). The latter is beyond the scope of the present chapter, since understanding the effect of the edge dynamics is rather challenging. Our goal is to extend the results of Theorem 3.3.3 to dynamic bipartite graphs. We will distinguish between different types of dynamics and we will see how they affect the mean transition time. We denote by $\mathcal{T}_{G(\cdot)}^{Q^0}$ the transition time of the dynamic graph $G(\cdot)$ with initial queue lengths Q^0.

Theorem 4.1.4 (Mean transition time for dynamic bipartite graphs). *Consider the dynamic bipartite graph $G(\cdot) = ((U, V), E(\cdot))$ with the edge dynamics governed by $\lambda(r)$ and initial queue lengths Q^0.*

(FD) *If $\lambda(r) \to \infty$, then the dynamics is fast and, with high probability as $r \to \infty$, the transition time satisfies*

$$\mathbb{E}_u[\mathcal{T}_{G(\cdot)}^{Q^0}] \asymp \lambda(r)^{-1} = o(1), \qquad r \to \infty. \tag{4.5}$$

(RD) *If $\lambda(r) = C \in (0, \infty)$, then the dynamics is regular and, with high probability as $r \to \infty$, the transition time satisfies*

$$\mathbb{E}_u[\mathcal{T}_{G(\cdot)}^{Q^0}] \asymp \lambda(r)^{-1} = \mathcal{O}(1), \qquad r \to \infty. \tag{4.6}$$

(SD) *If $\lambda(r) \to 0$, then the dynamics is slow and the following cases occur.*

(SDc) *If $\lambda(r) \succeq r^{-(1 \wedge \beta(d^*-1))}$, then the dynamics is competitive and, with high probability as $r \to \infty$, the transition time satisfies*

$$\mathbb{E}_u[\mathcal{T}_{G(\cdot)}^{Q^0}] \asymp \lambda(r)^{-1}, \qquad r \to \infty. \tag{4.7}$$

More precisely, let $\lambda(r) = r^{-\alpha}$ with $0 < \alpha \leq 1 \wedge \beta(d^* - 1)$, and let $T_U(r)$ be the average time it takes for the queue lengths at nodes in U to hit zero.

(I) *$\beta \in (0, \frac{1}{d^*-1})$: subcritical regime. With high probability as $r \to \infty$,*

$$\mathbb{E}_u[\mathcal{T}_{G(\cdot)}^{Q^0}] \asymp r^{\alpha}\,[1 + o(1)], \qquad r \to \infty. \tag{4.8}$$

(II) *$\beta = \frac{1}{d^*-1}$: critical regime. With high probability as $r \to \infty$,*

$$\mathbb{E}_u[\mathcal{T}_{G(\cdot)}^{Q^0}] \asymp r^{\alpha}\,[1 + o(1)], \qquad r \to \infty. \tag{4.9}$$

In particular, when $\alpha = 1$, with positive probability,

$$\mathbb{E}_u[\mathcal{T}_{G(\cdot)}^{Q^0}] = T_U(r)\,[1 + o(1)], \qquad r \to \infty. \tag{4.10}$$

(III) *$\beta \in (\frac{1}{d^*-1}, \infty)$: supercritical regime. When $0 < \alpha \leq 1$, with high probability as $r \to \infty$,*

$$\mathbb{E}_u[\mathcal{T}_{G(\cdot)}^{Q^0}] \asymp r^{\alpha}\,[1 + o(1)], \qquad r \to \infty. \tag{4.11}$$

In particular, when $\alpha = 1$, with positive probability,

$$\mathbb{E}_u[\mathcal{T}^{Q^0}_{G(\cdot)}] = T_U(r)\,[1 + o(1)], \qquad r \to \infty. \tag{4.12}$$

When $\alpha > 1$, with high probability as $r \to \infty$,

$$\mathbb{E}_u[\mathcal{T}^{Q^0}_{G(\cdot)}] = T_U(r)\,[1 + o(1)], \qquad r \to \infty. \tag{4.13}$$

(SDnc) *If $\lambda(r) \prec r^{-(1\wedge\beta(d^*-1))}$, then the dynamics is non-competitive and, with high probability as $r \to \infty$, the transition time satisfies Theorem 3.3.3.*

Note that the order of the mean transition time depends on the speed of the dynamics. When the dynamics is fast (FD), the edges quickly appear and disappear, reaching in time $o(1)$ the state where nodes in V have no edges connecting them to U. Since nodes in V are aggressive, they eventually activate in time $o(1)$. When the dynamics is regular (RD), the situation is similar, but it takes time $\mathcal{O}(1)$ to reach the state where all the edges are simultaneously absent. When the dynamics is slow (SD), a node in V can also activate through the nucleation of its fork (recall Definition 3.1.2). In the case of competitive dynamics (SDc), the relation between the speed of the dynamics and the aggressiveness of the nodes in U plays a key role, while in the case of non-competitive dynamics (SDnc), the network behaves as if the edges were fixed at the initial configuration and there were no dynamics. Note that, in the cases of fast, regular and competitive dynamics, the order of the mean transition time is given by the reciprocal of the rate $\lambda(r)$.

§4.1.4 Discussion and outline

Intuition. A node in V can activate for two reasons. It can activate when its neighbors are simultaneously inactive or when there are no edges connecting it to nodes in U. Interpolation between these two situations gives rise to different cases, which mainly depend on the speed of the dynamics. In the case of competitive dynamics, we are able to distinguish between different behaviors for the mean transition time by analyzing the subcritical, critical and supercritical regimes separately. To summarize, with high probability as $r \to \infty$, the order of activation of nodes in V follows one of the paths generated by the algorithm until the edge dynamics of rate $\lambda(r)$ becomes competitive. The competition begins on time scale $\lambda(r)^{-1}$, the time scale on which all the remaining nodes in V activate, if there are any, and the transition occurs.

Pre-factor. In order to give precise asymptotics, including the pre-factor for the mean transition time, we must analyze a more complicated Markov process describing how the states of the nodes, the queue lengths and the states of the edges change over time. This is beyond the scope of the present chapter, but in Section 4.4 we give an overview of the main challenges.

Outline of the chapter. The remainder of this chapter is organized as follows. In Section 4.2 we discuss the main effects of the dynamics on the mean transition time and we explain how it can slow down or speed up the activation of each node in V.

In Section 4.3 we prove Theorem 4.1.4 by discussing the different types of dynamics separately. In Section 4.4 we describe the graph evolution and discuss what needs to be considered in order to compute the pre-factor of the mean transition time. In Appendix D we consider a model where the activation rates are fixed and not queue-dependent. We adapt results from this chapter and the previous chapters in order to study how the dynamics affects the transition time.

§4.2 The edge dynamics

In this section we analyze the effects that different types of dynamics have on the mean transition time of the network.

§4.2.1 Disconnection time

Recall from Section 3.1.1 that the nucleation time $\mathcal{T}_v^Q$ of the fork of a node $v \in V$ given the initial queue lengths Q is the time it takes for its neighbors to become simultaneously inactive, so that v can activate as soon as its clock ticks. Due to the dynamics, a node $v \in V$ does not necessarily activate through the nucleation of its fork, but it can also activate if at some point there are no edges connecting it to nodes in U. The dynamics, indeed, might sometimes bring the graph to a configuration where the degree of v is temporarily 0, so that v can activate as soon as its clock ticks in $o(1)$.

Definition 4.2.1 (Disconnection time).
Given $v \in V$, we call *disconnection time* of v the time it takes for v to be disconnected from U, i.e., to have all possible edges connecting it to U simultaneously absent. We denote the disconnection time of v by $D_v^{Q^0}$, where Q^0 indicates the initial queue lengths.

As introduced in Section 4.1.2, the dynamics affects the network by allowing the edges to appear and disappear according to a Poisson clock with rate $\lambda(r)$. The alternation between the states of each edge $e \in E_{MN}$ is described by an exponential random variable $S_e \simeq \mathrm{Exp}(\lambda(r))$ with mean $\mu(r) = \lambda(r)^{-1}$. Note that, with high probability as $r \to \infty$, S_e takes values of the order of its mean, i.e., $S_e \asymp \mu(r)$. Indeed, if we pick $x \prec \mu(r)$, then

$$\lim_{r\to\infty} \mathbb{P}(S_e \leq x) = \lim_{r\to\infty} 1 - e^{-\lambda(r)x} = 0, \tag{4.14}$$

and the same holds for $x \succ \mu(r)$. In other words, if an edge is absent at time t, then, with high probability as $r \to \infty$, it will take an amount of time of order $\mu(r)$ for the Poisson clock to tick and for the edge to become present. Vice versa, if an edge is present at time t, then it will take an amount of time of order $\mu(r)$ for the edge to become absent.

The arbitrary bipartite initial configuration of the graph plays an important role in understanding the transition time. Consider a node in $v \in V$ of initial degree $d_v(0) = d > 0$. Since $|U| = M$, there are M possible total edges connecting v to U.

We construct a continuous-time Markov chain $\mathcal{M}$ where each state k represents the set of configurations of the M edges in which k edges are present and $M-k$ edges are absent. State 0 corresponds to all edges being absent, state 1 corresponds to the M possible configurations with exactly one edge present, and so on (see Figure 4.1 below).

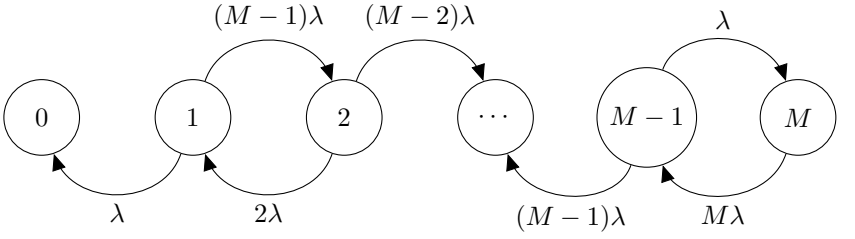


Figure 4.1: The Markov chain $\mathcal{M}$ describing how the edge dynamics changes the degree of a node in V. It is a birth-death process with M transient states and one absorbing state.

We consider state 0 as an absorbing state, since we are interested in computing the hitting times to state 0 starting from any other state. From state M we can only jump to state $M-1$, when one of the M present edges disappears, which happens with rate $M\lambda$. From each state $0 < k < M$ we jump to the neighboring states also with rate $M\lambda$. Indeed, as soon as the clock of one of the M possible edges ticks, we jump to the state $k+1$ if the edge was absent and becomes present, while we jump to the state $k-1$ if the edge was present and becomes absent. Hence, we jump from state k to state $k+1$ with probability $\frac{M-k}{M}$, while we jump from state k to state $k-1$ with probability $\frac{k}{M}$.

The transition rate matrix H of the Markov chain $\mathcal{M}$ is given by

$$H = \begin{array}{c} \\ 0 \\ 1 \\ 2 \\ \vdots \\ M-1 \\ M \end{array} \begin{array}{c} \begin{array}{cccccc} 0 & 1 & 2 & \cdots & M-1 & M \end{array} \\ \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \lambda & -M\lambda & (M-1)\lambda & 0 & 0 & 0 \\ 0 & 2\lambda & -M\lambda & \cdots & 0 & 0 \\ 0 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & 0 & \cdots & -M\lambda & \lambda \\ 0 & 0 & 0 & 0 & M\lambda & -M\lambda \end{pmatrix} \end{array} \tag{4.15}$$

and can be written as

$$H = \begin{pmatrix} 0 & \mathbf{0} \\ \mathbf{S}^0 & S \end{pmatrix}, \tag{4.16}$$

where S is an $M \times M$ matrix and $\mathbf{S}^0 = -S\mathbf{1}_M$, where $\mathbf{1}_M$ represents the M-dimensional column vector with every element being 1. Let

$$(a_0, \mathbf{a}) = (a_0, a_1, \ldots, a_M) \tag{4.17}$$

be the $M+1$ dimensional row vector describing the probability of starting in one of the $M+1$ states. Since $d_v(0) = d$, we have that the d-th entry of $\mathbf{a}$ equals 1 and all

the other entries equal 0. Computing the disconnection time of a node with initial degree d is equivalent to computing the hitting time of the Markov chain $\mathcal{M}$ to state 0 starting from state d.

Lemma 4.2.2 (Mean and law of the disconnection time).
Consider a node $v \in V$ of initial degree $d_v(0) = d > 0$, and let the edge dynamics be such that $\mathbb{E}_u[S_e] = \mu(r)$ for each $e \in E_{MN}$.

(i) The disconnection time $D_v^{Q^0}$ satisfies

$$\mathbb{E}_u[D_v^{Q^0}] = C_d\,\mu(r)\,[1 + o(1)], \qquad r \to \infty, \tag{4.18}$$

where $(C_1\,\mu(r), \dots, C_M\,\mu(r))$ is the solution of the linear system of equations

$$\begin{cases} x_1 &= \frac{1}{M}\frac{\mu(r)}{M} + \frac{M-1}{M}\left(\frac{\mu(r)}{M} + x_2\right) \\ x_2 &= \frac{2}{M}\left(\frac{\mu(r)}{M} + x_1\right) + \frac{M-2}{M}\left(\frac{\mu(r)}{M} + x_3\right) \\ \dots &= \dots \\ x_{M-1} &= \frac{M-1}{M}\left(\frac{\mu(r)}{M} + x_{M-2}\right) + \frac{1}{M}\left(\frac{\mu(r)}{M} + x_M\right) \\ x_M &= \frac{\mu(r)}{M} + x_{M-1}. \end{cases} \tag{4.19}$$

(ii) The law of the disconnection time $D_v^{Q^0}$ follows a phase-type distribution $\mathrm{PH}(\boldsymbol{a}, S)$ and is given by

$$\lim_{r\to\infty} \mathbb{P}_u(D_v^{Q^0} > x) = \boldsymbol{a}\exp(Sx)\boldsymbol{1}, \qquad x \in (0, \infty), \tag{4.20}$$

where $\boldsymbol{a}$ and S are as in (4.17) *and* (4.16)*, respectively. In particular, the above probability equals the sum of the entries in d-th row of the matrix* $\exp(Sx)$.

Proof. We prove the two statements separately.

(i) Consider the Markov chain $\mathcal{M}$ described above. We know that from each state $k > 0$, we jump to a neighboring state with rate $M\lambda$. The jump occurs exactly when the first of the M possible edges changes its state. This corresponds to the minimum of M i.i.d. exponential random variables, which is known to follow an exponential distribution with mean $\frac{\mu(r)}{M}$. If v has initial degree d, then we start from state d. We denote by x_k the mean hitting times of state 0 starting from state k. The above system of equations allows us to compute the mean disconnection time of v.

Since, the system of equations is linear in $\mu(r)$ and in the variables x_i's, its solution is linear in $\mu(r)$. Hence the mean disconnection time is of order $\mu(r)$.

(ii) The disconnection time of a node $v \in V$ of initial degree $d > 0$ is the hitting time of state 0 of the Markov chain $\mathcal{M}$ starting from state d. The distribution of the hitting time to the unique absorbing state, starting from any of the other

finite transient states, is said to be phase-type and is denoted by $\mathrm{PH}(\mathbf{a}, S)$, with $\mathbf{a}$ and S as in (4.17) and (4.16), respectively.

The distribution function of $D_v^{Q^0}$ is given by

$$\lim_{r\to\infty} \mathbb{P}_u(D_v^{Q^0} \le x) = \int_0^x \mathcal{P}(y)\,dy = 1 - \mathbf{a}\exp(Sx)\,\mathbf{1}, \qquad x \in (0,\infty), \quad (4.21)$$

where $\exp(\cdot)$ indicates the matrix exponential, and

$$\mathcal{P}(z) = \mathbf{a}\exp(Sz)\,\mathbf{S}^0, \qquad z \in (0,\infty), \quad (4.22)$$

with $\mathbf{S}^0$ as in (4.16). Since the vector $\mathbf{a}$ has its d-th entry equal to 1 and all the other entries equal to 0, we have that the product $\mathbf{a}\exp(Sx)\,\mathbf{1}$ equals the sum of the entries in the d-th row of the matrix $\exp(Sx)$.

□

Note that the results of Lemma 4.2.2 hold even without letting $r \to \infty$.

§4.2.2 Nucleation vs. dynamics

Without loss of generality, we may consider interference graphs with no isolated nodes in V, since after time $o(1)$ we would be in such a scenario anyway.

Lemma 4.2.3 (Isolated nodes).
Nodes in V with initial degree 0 activate in time $o(1)$ as $r \to \infty$.

Proof. Consider the situation where $\lambda(r) \prec g_V(0)$, i.e., the dynamics is slower than the average time it takes for the activation clock of nodes in V to tick. Then a node $v \in V$ with initial degree 0 activate as soon as its clock ticks, hence in time $o(1)$. Next, consider the situation where the dynamics is very fast, $\lambda(r) \succ g_V(0)$. Then a node $v \in V$ with initial degree 0 might be blocked by some active neighbors in U by the time its activation clock ticks for the first time. Recall that $|U| = M$ and note that there are 2^M possible configurations of edges connecting v to U. Each time the activation clock of v ticks, the probability of being in each of the possible configurations tends to the uniform probability $1/2^M$ as $r \to \infty$. Therefore, after a finite number of attempts, v eventually activates. Since each tick of the activation clock of v takes time $o(1)$, v activates in time $o(1)$. Lastly, consider the situation where $\lambda(r) \asymp g_V(0)$. If the activation clock of a node $v \in V$ with initial degree 0 ticks before any of its potential edges appear, then v activates in time $o(1)$. Otherwise, each subsequent activation attempt will not be successful unless the edge configuration is such that v has no neighbors. In other words, v can activate only when the Markov chain describing how its degree changes over time is in state 0. In this case, v activates with a probability that at time t is given by $\frac{g_V(t)}{g_V(t)+M\lambda(r)} > 0$ as $r \to \infty$. Since $\lambda(r)^{-1} = o(1)$, by using similar arguments as in the proof of Lemma 4.2.2, the time it takes for the Markov chain to return to state 0 when starting from state 0 is $o(1)$. Hence, v has the chance to activate with positive probability every period of time $o(1)$. Therefore, after a finite number of attempts, v eventually activates in time $o(1)$. □

We call *activation time* of $v \in V$ the time it takes for v to activate. Depending on the dynamics, this can be given either by its nucleation time $\mathcal{T}_v^{Q^0}$ or by its disconnection time $D_v^{Q^0}$. When the dynamics is fast enough, nodes in V eventually activate because their clocks tick and no edges connect them to nodes in U. On the other hand, when the dynamics is particularly slow, it is more likely for nodes in V to activate through the nucleation of its fork, and the network tends to behave as if the edges were frozen at the initial configuration. In between these two scenarios the dynamics is more interesting and, depending on its speed, we distinguish between different behaviors. Proposition 4.2.4 below describes the competition between the nucleation and the dynamics.

Proposition 4.2.4 (Nucleation vs. dynamics).
Let $v \in V$ be the node of minimum degree at time $t = 0$, with $d_v(0) = d > 0$.

(i) If $\lambda(r) \succ r^{-(1\wedge\beta(d-1))}$, then, with high probability as $r \to \infty$, the activation time of v is given by its disconnection time, i.e.,

$$\lim_{r\to\infty} \mathbb{P}_u(D_v^{Q^0} < \mathcal{T}_v^{Q^0}) = 1. \tag{4.23}$$

(ii) If $\lambda(r) \asymp r^{-(1\wedge\beta(d-1))}$, then the activation time of v is given either by its nucleation time with positive probability or by its disconnection time with positive probability.

(iii) If $\lambda(r) \prec r^{-(1\wedge\beta(d-1))}$, then, with high probability as $r \to \infty$, the activation time of v is given by its nucleation time, i.e.,

$$\lim_{r\to\infty} \mathbb{P}_u(\mathcal{T}_v^{Q^0} < D_v^{Q^0}) = 1. \tag{4.24}$$

Proof. Recall that $\mu(r) = \lambda(r)^{-1}$ and that the disconnection time $D_v^{Q^0}$ is given by a phase-type random variable with mean of order $\mu(r)$. Since phase-type random variables are constructed by convolutions of exponential random variables, we have that, with high probability as $r \to \infty$, $D_v^{Q^0}$ takes values of order $\mu(r)$. Recall also that, depending on the relation between β and d, the nucleation time $\mathcal{T}_v^{Q^0}$ is given by an exponential random variable with mean of order $r^{\beta(d-1)}$, by a polynomial random variable with mean of order r, or by $T_U(r)$, which is the average time it takes for the queue lengths at nodes in U to hit zero. Hence, with high probability as $r \to \infty$, $\mathcal{T}_v^{Q^0}$ takes values of order $r^{1\wedge\beta(d-1)}$. It is therefore immediate to distinguish between the three cases.

(i) Since $\mu(r) \prec r^{1\wedge\beta(d-1)}$, with high probability as $r \to \infty$, v activates due to absence of edges.

(ii) Since $\mu(r) \asymp r^{1\wedge\beta(d-1)}$, there is a competition between the nucleation time $\mathcal{T}_v^{Q^0}$ and the phase-type random variable $D_v^{Q^0}$. Depending on their parameters, each of them can occur before the other with positive probability.

(iii) Since $\mu(r) \succ r^{1\wedge\beta(d-1)}$, with high probability as $r \to \infty$, v activates through the nucleation of its fork.

□

§4.3 Proofs of the main results

In this section we prove Theorem 4.1.4 by analyzing the different types of dynamics separately.

§4.3.1 Proof: fast dynamics

Consider the fast dynamics (FD) where $\lambda(r) \to \infty$ as $r \to \infty$.

Proof of Theorem 4.1.4 (FD). With high probability as $r \to \infty$, for each edge the random intervals between clock ticks are of order $\lambda(r)^{-1} = o(1)$. By Lemma 4.2.2, the mean disconnection time of a node in V is of order $\lambda(r)^{-1}$. Moreover, by Proposition 4.2.4, with high probability as $r \to \infty$, each node activates due to absence of edges and not through the nucleation of its fork, and hence it activates in a time of order $\lambda(r)^{-1}$. In conclusion, with high probability as $r \to \infty$, the transition time of $G(\cdot)$ with initial queue lengths Q^0 satisfies

$$\mathbb{E}_u[\mathcal{T}^{Q^0}_{G(\cdot)}] \asymp \lambda(r)^{-1} = o(1), \qquad r \to \infty, \tag{4.25}$$

hence the claim is settled. □

§4.3.2 Proof: regular dynamics

Consider the regular dynamics (RD) where $\lambda(r) = C \in (0, \infty)$.

Proof of Theorem 4.1.4 (RD). With high probability as $r \to \infty$, for each edge the random intervals between clock ticks are of order $\lambda(r)^{-1} = \mathcal{O}(1)$. By Lemma 4.2.2, the mean disconnection time of a node in V is of order $\lambda(r)^{-1}$. Note that nodes in V of initial degree 1 can activate either because their only neighbor deactivates in $\mathcal{O}(1)$ or due to absence of edges with a mean disconnection time of order $\lambda(r)^{-1}$. Moreover, by Proposition 4.2.4, with high probability as $r \to \infty$, nodes in V of initial degree greater than 1 activate due to absence of edges in a time of order $\lambda(r)^{-1}$. In conclusion, with high probability as $r \to \infty$, the transition time of $G(\cdot)$ with initial queue lengths Q^0 satisfies

$$\mathbb{E}_u[\mathcal{T}^{Q^0}_{G(\cdot)}] \asymp \lambda(r)^{-1} = O(1), \qquad r \to \infty, \tag{4.26}$$

hence the claim is settled. □

§4.3.3 Proof: non-competitive dynamics

Consider the slow dynamics where $\lambda(r) \to 0$ as $r \to \infty$ with $\lambda(r) \prec r^{-(1 \wedge \beta(d^*-1))}$, called the non-competitive dynamics (SDnc). In this case the dynamics is so slow that it has no effect on the transition.

Proof of Theorem 4.1.4 (SDnc). The mean disconnection time of any node in V is of order larger than $r^{1\wedge\beta(d^*-1)}$. Hence each node in V activates through the nucleation of its fork, which is at most of order $r^{1\wedge\beta(d^*-1)}$. The dynamics is very slow, almost frozen, and so it does not affect the nucleation of the forks. Hence, with high probability as $r \to \infty$, the transition time of $G(\cdot)$ with initial queue lengths Q^0 satisfies Theorem 3.3.3 and the network behaves as if there were no dynamics. Hence the claim is settled. □

§4.3.4 Proof: competitive dynamics

Consider the slow dynamics (SD) where $\lambda(r) \to 0$ as $r \to \infty$ with $\lambda(r) = r^{-\alpha}$, with $0 < \alpha \leq 1 \wedge \beta(d^* - 1)$, called the competitive dynamics (SDc). This is the most interesting type of dynamics, since it competes with the fork nucleations. The activation of the nodes in V can occur both because of the absence of their edges and because of the nucleation of their forks. Recall the algorithm defined in Section 3.2 in Chapter 3.

Proof of Theorem 4.1.4 (SDc). Denote by $\hat{d}$ the largest integer such that $\beta(\hat{d}-1) < \alpha$. Let the algorithm generate all possible activation paths for nodes in V and denote this set by $\mathcal{A}$. Fix a path $a \in \mathcal{A}$. Consider the sequence of activating nodes along the path a up to the step in which the degree is larger than $\hat{d}$. Say that at step k we have $\bar{d}_k > \hat{d}$. Consider only the first $k-1$ steps. We indicate by $A_a(\alpha)$ the event that the network follows the path $a \in \mathcal{A}$ until time scale r^α. On time scale r^α the dynamics starts competing with the nucleation, and the order of activation of the remaining nodes described by the algorithm is not preserved anymore. In other words, the order of activation of nodes in V follows the order of activation of the path a only for the first $k-1$ nodes. With each of these $k-1$ nodes is associated a nucleation time of order less than or equal to $r^{1\wedge\beta(\hat{d}-1)}$. Hence, by Proposition 4.2.4, with high probability as $r \to \infty$, the activation time of these nodes is given by their nucleation time. We apply Proposition 4.2.4 to each iteration of the graph, each time by considering a node with minimum degree $\bar{d}_j$ for $j = 1, \ldots, k-1$. Indeed, we know from Lemma 4.2.2, that the mean disconnection time of a node is of order r^α. We treat the subcritical, critical and supercritical regimes separately.

(I) $\beta \in (0, \frac{1}{d^*-1})$: subcritical regime. We have $0 < \alpha \leq \beta(d^* - 1) < 1$. The activation time of the next activating node is of order r^α. It cannot be of smaller order since at step k we have $\bar{d}_k > \hat{d}$ by construction. It cannot be of higher order either since the disconnection time of any of the remaining nodes is of order r^α. After this activation, there might be nodes whose degree has decreased and whose nucleation time is of smaller order. When we sum the mean activation times of the nodes in V to compute the mean transition time, we see that these nodes will not contribute significantly as $r \to \infty$. All the remaining nodes are likely to activate in any possible order, but none of them will have an activation time of order larger than r^α. To know how many nodes contribute to the transition time with an activation time of order r^α, we need to have more control on how the degrees of the nodes evolve over time. To

conclude, the order of activation of nodes in V follows the path a as long as the nucleation times associated to the nodes are of order smaller than r^α. After that, the remaining nodes can activate with positive probability in any order with an activation time of order at most r^α. Hence, the transition time conditional on the event $A_a(\alpha)$ satisfies

$$\mathbb{E}_u[\mathcal{T}^{Q^0}_{G(\cdot)} \,|\, A_a(\alpha)] \asymp r^\alpha\,[1+o(1)], \qquad r\to\infty, \tag{4.27}$$

and we get

$$\mathbb{E}_u[\mathcal{T}^{Q^0}_{G(\cdot)}] \asymp r^\alpha\,[1+o(1)], \qquad r\to\infty. \tag{4.28}$$

(II) $\beta = \frac{1}{d^*-1}$: critical regime. For $0<\alpha<1$, the situation is the same as in the subcritical regime described above. For $\alpha=1$, the activation time of the next activating node is of order r. After this activation, all the remaining nodes are likely to activate in any possible order, but none of them will have an activation time of order larger than r. The order of activation of nodes in V follows the path a as long as the nucleation times associated to the nodes are of order smaller than r. After that, the remaining nodes can activate with positive probability in any order with an activation time of order at most r. Hence, the transition time conditional on the event $A_a(\alpha)$ satisfies

$$\mathbb{E}_u[\mathcal{T}^{Q^0}_{G(\cdot)} \,|\, A_a(\alpha)] \asymp r\,[1+o(1)], \qquad r\to\infty, \tag{4.29}$$

and we get

$$\mathbb{E}_u[\mathcal{T}^{Q^0}_{G(\cdot)}] \asymp r\,[1+o(1)], \qquad r\to\infty. \tag{4.30}$$

Note that if any of the nodes has an activation time of order r but larger than $T_U(r)$, then the transition time conditional on the event $A_a(\alpha)$ is the time it takes for the queue lengths at nodes in U to hit zero, which satisfies

$$\mathbb{E}_u[\mathcal{T}^{Q^0}_{G(\cdot)} \,|\, A_a(\alpha)] = T_U(r)\,[1+o(1)], \qquad r\to\infty. \tag{4.31}$$

Hence,

$$\mathbb{E}_u[\mathcal{T}^{Q^0}_{G(\cdot)}] = T_U(r)\,[1+o(1)], \qquad r\to\infty. \tag{4.32}$$

(III) $\beta \in (\frac{1}{d^*-1}, \infty)$: supercritical regime. For $0<\alpha<1$, the situation is the same as in the subcritical regime described above. For $\alpha=1$, the situation is the same as in the critical regime described above. For $\alpha>1$, the transition time conditional on the event $A_a(\alpha)$ is the time it takes for the queue lengths at nodes in U to hit zero, which satisfies

$$\mathbb{E}_u[\mathcal{T}^{Q^0}_{G(\cdot)} \,|\, A_a(\alpha)] = T_U(r)\,[1+o(1)], \qquad r\to\infty. \tag{4.33}$$

Hence,

$$\mathbb{E}_u[\mathcal{T}^{Q^0}_{G(\cdot)}] = T_U(r)\,[1+o(1)], \qquad r\to\infty. \tag{4.34}$$

□

Note that the order of the transition time does not depend on the path along which we compute it. The algorithm generates all possible activation paths of the nodes nucleating before time scale $\lambda(r)^{-1} = r^{\alpha}$. The remaining nodes can activate in any order depending on the dynamics. To compute the pre-factor of the mean transition time along these paths, we need to analyze in detail the Markov process describing the graph evolution, in particular, the degrees of the nodes changing over time. Our methods do not capture this detail and we are only able to state a result for the leading order term.

§4.4 The graph evolution

In this section we discuss the Markov process describing the graph evolution under the dynamics. Control on this process is the key to obtaining a more precise asymptotics for the mean transition time of the network.

§4.4.1 The graph evolution process

Consider a dynamics with rate $\lambda(r) = r^{-\alpha}$. We have seen in Proposition 4.2.4 that each node in V whose nucleation time is of smaller order than r^{α} activates through the nucleation of its fork. On time scale r^{α} the dynamics starts competing with the nucleation and the order of activation of the remaining nodes described by the algorithm is not preserved anymore. Note that the algorithm updates the graph at each iteration in order to keep track of the degree of the remaining nodes after each activation. When introducing the dynamics on the edges, we need information about the states of the nodes and the edges in the graph. We assume that the algorithm does not update the graph at each iteration anymore, but we focus on the number of active neighbors each node has.

Definition 4.4.1 (Active degree).
We define the *active degree* of a node as the number of its active neighbors. For $u \in U$, the active degree at time t is given by

$$\tilde{d}_u(t) = \big|\{v \in V : uv \in E(t), X_v(t) = 1\}\big|. \tag{4.35}$$

Analogously, for $v \in V$, the active degree at time t is given by

$$\tilde{d}_v(t) = \big|\{u \in U : uv \in E(t), X_u(t) = 1\}\big|. \tag{4.36}$$

Note that for a node to activate, its active degree must be 0. It is immediate to see that the active degree of a node cannot exceed its degree, i.e., for any $u \in U$ and $v \in V$,

$$\tilde{d}_u(t) \leq d_u(t), \qquad \tilde{d}_v(t) \leq d_v(t). \tag{4.37}$$

The main challenge in describing the graph evolution is that any of the remaining nodes could activate next with positive probability. The activation of a node due to absence of edges is captured by the scenario in which its active degree hits 0. The activation of a node through the nucleation of its fork depends on the aggressiveness of

the activation rates and on the number of active neighbors. Both types of activation are determined by the degree evolution. Assume, for example, that an edge between two active nodes appears. By our model assumptions (see Remark 4.1.3), the node in U deactivates, implying that the active degrees of its neighbors in V decrease by 1. If the mean nucleation time of the new fork of one of the neighbors is of order less than or equal to r^α, then this neighbor will be more likely to activate through the nucleation of its fork. The degree evolution induced by the dynamics affects both the disconnection and the nucleation times of the nodes.

The node activity process $(X(t), Q(t))_{t\geq 0}$ and the edge activity process $(Y(t))_{t\geq 0}$ form a continuous-time Markov process on

$$\mathcal{X} \times \mathbb{R}_{\geq 0} \times \mathcal{Y} \tag{4.38}$$

that describes the evolution of the graph under the effect of the dynamics. We refer to this process as the *graph evolution process*. Note that if we know which nodes are active and which edges are present, then we can recover the degree and the active degree of each node in the graph. Hence, understanding the graph evolution process is crucial to describe how the degrees of the nodes change over time and how nodes activate.

§4.4.2 Transitions

Consider a feasible state where some nodes are active and some edges are present. By feasible we mean that it respects the constraints given by the edges, for which two connected nodes cannot be active simulteneously. Recall that $|U| = M, |V| = N$ and $|E_{MN}| = MN$. Hence, an arbitrary feasible state at time t has h active nodes in U with $h = 0, \ldots, M$, k active nodes in V with $k = 0, \ldots, N$, and l present edges with $l = 0, \ldots, MN$. Consequently, there are $M - h$ inactive nodes in U, $N - k$ inactive nodes in V, and $MN - l$ absent edges. Note that the initial state u is described by $h = M$, $k = 0$ and $l = |E(0)|$, while the transition occurs as soon as state v is reached, for which $k = N$.

Clock ticks. The graph evolution is governed by different Poisson clocks ticking at various rates: the activation clocks, the deactivation clocks and the edge clocks. We analyze how the network evolves each time one of these clock ticks. Moreover, note that the queue lengths, hence the input process (recall Definition 1.1.3), also play a role, since the activation rates depend on them.

- The activation clock of a node $u \in U$ ticks at rate $g_U(Q_u(t))$ at time t. The probability of this clock being the first one to tick is given by

$$\frac{g_U(Q_u(t))}{Z}, \tag{4.39}$$

with

$$Z = \sum_{i=1}^{M-h} g_U(Q_i(t)) + \sum_{j=1}^{N-k} g_V(Q_j(t)) + h + k + MN\lambda(r). \tag{4.40}$$

The tick has two possible effects on the network. If the neighbors of u are all inactive, then u activates and the active degrees of all its neighbors increase by 1. If there is at least one active neighbor of u, then the activation attempt fails and nothing happens.

- The deactivation clock of a node $u \in U$ ticks at rate 1. The probability of this clock being the first one to tick is given by

$$\frac{1}{Z}. \tag{4.41}$$

Node u deactivates and the active degrees of all its neighbors decrease by 1.

- The activation clock of a node $v \in V$ ticks at rate $g_V(Q_v(t))$ at time t. The probability of this clock being the first one to tick is given by

$$\frac{g_V(Q_v(t))}{Z}. \tag{4.42}$$

The tick has two possible effects on the network. If the neighbors of v are all inactive, then v activates and the active degrees of all its neighbors increase by 1. If there is at least one active neighbor of v, then the activation attempt fails and nothing happens.

- The deactivation clock of a node $v \in V$ ticks at rate 1. The probability of this clock being the first one to tick is given by

$$\frac{1}{Z}. \tag{4.43}$$

Node v deactivates and the active degrees of all its neighbors decrease by 1.

- The activation clock of an edge $e \in E_{MN}$ ticks at rate $\lambda(r)$. The probability of this clock being the first one to tick is given by

$$\frac{\lambda(r)}{Z}. \tag{4.44}$$

Depending on which edge appears or disappears and on the nodes involved, the tick has different effects on the network, which are described below.

Edge appearing and disappearing. If we know the number of active nodes in U and V, then we can compute the probabilities of each of the following scenarios with simple combinatorial arguments. There are four possible scenarios in which an edge can appear.

($\circ\,\circ$) When an edge between two inactive nodes appears, their degrees increase by 1.

($\circ\,\bullet$) When an edge between an inactive node in U and an active node in V appears, the active degree of the node in U increases by 1 and the degree of the node in V increases by 1.

($\bullet\circ$) When an edge between an active node in U and an inactive node in V appears, the degree of the node in U increases by 1 and the active degree of the node in V increases by 1.

($\bullet\bullet$) When an edge between two active nodes appears, the node in U deactivates, its active degree increases by 1, the active degrees of all its neighbors in V decrease by 1 and the degree of the node in V increases by 1.

In a similar fashion, there are three possible scenarios in which an edge can disappear. Recall that there cannot be an edge between two active nodes.

($\circ\circ$) When an edge between two inactive nodes disappears, their degrees decrease by 1.

($\circ\bullet$) When an edge between an inactive node in U and an active node in V disappears, the active degree of the node in U decreases by 1 and the degree of the node in V decreases by 1.

($\bullet\circ$) When an edge between an active node in U and an inactive node in V disappears, the degree of the node in U decreases by 1 and the active degree of the node in V decreases by 1.

The transition time is related to the graph evolution process, since the activation times of the nodes in V depend on the activation rates, the speed of the dynamics and the degree evolution. The complicated nature of the process prevents us from deriving an explicit formula for the pre-factor of the mean transition time, which would require a better control on the precise asymptotics of each activation.

§D Appendix: a model with fixed activation rates

We have seen how the dynamics influences the mean transition time of wireless random-access models where the activation rates depend on the current queue lengths at the nodes. The model is quite challenging and deals with two levels of complexity, namely, the queue-based activation rates and the edge dynamics. Not much is known in the literature for random-access protocols with dynamic interference graph, even for models with *fixed activation rates*. In this section we adapt the theory built up in Chapters 2–3 to study the effect of the dynamics on such type of models. Assume that the activation rates are of the form

$$r_i(t) = \begin{cases} r^{\beta}, & \text{if } i \in U, \\ r^{\beta'}, & \text{if } i \in V, \end{cases} \tag{4.45}$$

with $\beta, \beta' \in (0, \infty)$ and $\beta' > \beta + 1$. We recall that we are interested in the transition time asymptotics as $r \to \infty$.

We start by adapting the results for complete bipartite graphs in Chapter 2 to the model with fixed activation rates. The following theorem is consistent with [59, Example 4.1].

Theorem D.1 (Complete bipartite graphs with fixed activation rates). *Consider the complete bipartite graph $G = ((U, V), E)$ with initial queue lengths Q^0 as in* (1.11)*. Suppose that* (4.45) *holds.*

(I) $\beta \in (0, \frac{1}{|U|-1})$*: subcritical regime. The transition time satisfies*

$$\mathbb{E}_u[\mathcal{T}_G^{Q^0}] = \frac{1}{|U|} r^{\beta(|U|-1)} [1 + o(1)], \qquad r \to \infty. \tag{4.46}$$

(II) $\beta = \frac{1}{|U|-1}$*: critical regime. The transition time satisfies*

$$\mathbb{E}_u[\mathcal{T}_G^{Q^0}] = \frac{1}{|U|} r [1 + o(1)], \qquad r \to \infty. \tag{4.47}$$

(III) $\beta \in (\frac{1}{|U|-1}, \infty)$*: supercritical regime. The transition time satisfies*

$$\mathbb{E}_u[\mathcal{T}_G^{Q^0}] = \frac{\gamma_U}{c - \rho_U} r [1 + o(1)], \qquad r \to \infty. \tag{4.48}$$

Proof. The claims follow from Sections 2.4.1–2.4.2 in Chapter 2. We compute the critical time scale and the mean transition time using fixed activation rates instead of time depending ones. In both the critical and subcritical regimes, the pre-factor turns out to be $\frac{1}{|U|}$ and the law is exponential. In the critical regime, we know that the queue lengths decrease significantly after a time of order r. However, this does not affect the transition time, since now the activation rates do not depend on the queue lengths. In the supercritical regime, we still have the same behavior as in the model with queue-dependent activation rates. Indeed, when the queue lengths at nodes in U hit zero, the nodes in U deactivate by assumption and the transition occurs. □

Next, we state a result for arbitrary bipartite graphs with fixed activation rates (analogue of Theorem 3.3.3 in Chapter 3). Note that the algorithm still plays a crucial role in determining the mean transition time.

Theorem D.2 (Arbitrary bipartite graphs with fixed activation rates). *Consider the bipartite graph $G = ((U, V), E)$ with initial queue lengths Q^0 as in* (1.11)*. Suppose that* (4.45) *holds. Let A_a be the event that the network follows the path $a \in \mathcal{A}$, among the paths generated by the algorithm.*

(I) $\beta \in (0, \frac{1}{d^*-1})$*: subcritical regime. The transition time satisfies*

$$\mathbb{E}_u[\mathcal{T}_G^{Q^0} | A_a] = \sum_{\substack{1 \le k \le N \\ k:\, \bar{d}_k = d^*}} \frac{1}{n_k d^*} r^{\beta(d^*-1)} [1 + o(1)], \qquad r \to \infty. \tag{4.49}$$

(II) $\beta = \frac{1}{d^*-1}$*: critical regime. Then the transition time satisfies*

$$\mathbb{E}_u[\mathcal{T}_G^{Q^0} | A_a] = \sum_{\substack{1 \le k \le N \\ k:\, \bar{d}_k = d^*}} \frac{1}{n_k d^*} r [1 + o(1)], \qquad r \to \infty. \tag{4.50}$$

The above result holds as long as the pre-factor is below the value $\frac{\gamma_U}{c-\rho_U}$*, which corresponds to the time it takes for the queue lengths at nodes in* U *to hit zero. Otherwise, the supercritical regime applies.*

(III) $\beta \in (\frac{1}{d^*-1}, \infty)$*: supercritical regime. The transition time satisfies*

$$\mathbb{E}_u[\mathcal{T}_G^{Q^0}] = \frac{\gamma_U}{c-\rho_U} r\,[1+o(1)], \qquad r \to \infty. \tag{4.51}$$

Proof. The claims follow from Theorem D.1 and the analysis of the algorithm and the next nucleation times in Sections 3.2 and 3.4.2 in Chapter 3. We derive the mean transition time along the paths generated by the algorithm by computing the next nucleation times at each step. In the subcritical regime, the nucleation times of nodes in V are all exponentially distributed and independent of each other. Indeed, the activation rates are the same, independently of the queue lengths decreasing over time. At each step k, the next nucleation time is the minimum of n_k i.i.d. exponential random variables, and hence its mean exhibits the term $f_k = \frac{1}{n_k}$ in the pre-factor. In the critical regime, the pre-factor of the mean transition time along each path must be below the value $\frac{\gamma_U}{c-\rho_U}$, otherwise the supercritical regime applies and the transition occurs because the queue lengths at nodes in U hit 0. If we assume that $\frac{\gamma_U}{c-\rho_U} > 1$, then the nucleation of a fork occurs before the queue lengths at nodes in U hit zero. We are able to derive the law of the transition time along each path for both the subcritical and critical regimes. Both are described by convolutions of the exponential laws of the next nucleation times of the activating nodes in V. In the supercritical regime, we have the same behavior as in the model with queue-dependent activation rates. □

Finally, we show that the results from Theorem 4.1.4 also hold when we consider a dynamic bipartite graph with fixed activation rates. We are able to compute the order of the mean transition time, while the pre-factor still depends on the graph evolution described in Section 4.4.

Theorem D.3 (Dynamic bipartite graphs with fixed activation rates). *Consider the dynamic bipartite graph* $G(\cdot) = ((U,V), E(\cdot))$ *with the edge dynamics governed by* $\lambda(r)$ *and initial queue lengths* Q^0*. Suppose that* (4.45) *holds. Then the results of Theorem 4.1.4 hold.*

Proof. The claim follows from Theorem D.2 and the intuition behind Proposition 4.2.4. The order of the mean transition time in the model with fixed activation rates is the same as in the model with queue-dependent activation rates. The dynamics competes with the nucleations of the nodes in the same way, depending on its speed. The different type of dynamics (fast, regular and slow) lead to the same results as in Theorem 4.1.4. □

PART II

SPECTRA OF INHOMOGENEOUS ERDŐS-RÉNYI RANDOM GRAPHS

CHAPTER 5

Spectral distribution of the adjacency and the Laplacian matrix

This chapter is based on:
A. Chakrabarty, R.S. Hazra, F. den Hollander, M. Sfragara. *Spectra of adjacency and Laplacian matrices of Inhomogeneous Erdős-Rényi random graphs.* Random Matrices: Theory and Applications, 2020.

Abstract

We consider inhomogeneous Erdős-Rényi random graphs G_N on N vertices in the non-sparse non-dense regime. The edge between the pair of vertices $\{i, j\}$ is retained with probability $\varepsilon_N f(\frac{i}{N}, \frac{j}{N})$, $1 \leq i \neq j \leq N$, independently of other edges, where $f\colon [0,1]^2 \to [0,\infty)$ is a continuous function such that $f(x,y) = f(y,x)$ for all $(x,y) \in [0,1]^2$. We study the empirical distribution of both the adjacency matrix A_N and the Laplacian matrix Δ_N associated with G_N, in the limit as $N \to \infty$ when $\lim_{N\to\infty} \varepsilon_N = 0$ and $\lim_{N\to\infty} N\varepsilon_N = \infty$. In particular, we show that the empirical spectral distributions of A_N and Δ_N, after appropriate scaling and centering, converge to deterministic limits weakly in probability. For the special case where $f(x,y) = r(x)r(y)$ with $r\colon [0,1] \to [0,\infty)$ a continuous function, we give an explicit characterization of the limiting distributions. Furthermore, we apply our results to constrained random graphs, Chung-Lu random graphs and social networks.

§5.1 Introduction and main results

In Section 5.1.1 we define the mathematical model. In Section 5.1.2 we state the existence of the limiting spectral distributions for the adjancency and Laplacian matrices after suitable scaling. In Section 5.1.3 we identify those limiting spectral distributions under the assumption that the connection probabilities have a multiplicative structure. In Section 5.1.4 we generalize our results to graphs where the connection probabilities are randomized. In Section 5.1.5 we anticipate some of the applications that we will discuss later and give an outline of the remainder of the chapter.

§5.1.1 Setting

We refer to Section 1.2.3 for a general introduction to spectra of Erdős-Rényi random graphs. We focus on *inhomogeneous Erdős-Rényi random graphs* and consider the non-dense non-sparse regime, where the degrees of the vertices diverge sublinearly with the size of the graph.

Let $f\colon [0,1]^2 \to [0,\infty)$ be a continuous function, satisfying

$$f(x,y) = f(y,x) \qquad \forall\, (x,y) \in [0,1]^2. \tag{5.1}$$

A sequence of positive real numbers $(\varepsilon_N\colon N \geq 1)$ is fixed that satisfies

$$\lim_{N\to\infty} \varepsilon_N = 0, \qquad \lim_{N\to\infty} N\varepsilon_N = \infty. \tag{5.2}$$

Consider the random graph G_N on the set of vertices $\{1,\dots,N\}$ where, for each (i,j) with $1 \leq i < j \leq N$, an edge is present between vertices i and j with probability

$$\varepsilon_N f(\tfrac{i}{N}, \tfrac{j}{N}), \tag{5.3}$$

independently of other pairs of vertices. In particular, G_N is an undirected graph with no self loops and no multiple edges. Boundedness of f ensures that $\varepsilon_N f(\frac{i}{N}, \frac{j}{N}) \leq 1$ for all $1 \leq i < j \leq N$ when N is large enough. If $f \equiv c$ with c a constant, then G_N is the Erdős-Rényi graph with edge retention probability $\varepsilon_N c$. For general f, G_N can be thought of as an inhomogeneous version of the Erdős-Rényi graph.

We next define our two main objects of interest. We refer to Section 1.2.2 for more details.

Definition 5.1.1 (Adjacency and Laplacian matrices).
The *adjacency matrix* of G_N is denoted by A_N and defined as in (1.15). Clearly, A_N is a symmetric random matrix whose diagonal entries are zero and whose upper triangular entries are independent Bernoulli random variables, i.e.,

$$A_N(i,j) \triangleq \mathrm{BER}\big(\varepsilon_N f\big(\tfrac{i}{N}, \tfrac{j}{N}\big)\big), \qquad 1 \leq i \neq j \leq N. \tag{5.4}$$

The *Laplacian matrix* of G_N is denoted Δ_N and defined as in (1.17).

§5.1.2 Existence of the limiting spectral distribution

We recall the definition of *empirical spectral distribution (ESD)* in (1.19). It is the probability measure that puts mass $1/N$ at every eigenvalue, respecting its algebraic multiplicity.

Our first theorem states the existence of the limiting spectral distribution of A_N after suitable scaling.

Theorem 5.1.2 (Existence of the limiting spectral distribution of A_N). *There exists a symmetric probability measure μ on $\mathbb{R}$ such that*

$$\lim_{N\to\infty} \mathrm{ESD}\big((N\varepsilon_N)^{-1/2}A_N\big) = \mu \quad \text{weakly in probability,} \tag{5.5}$$

and μ is compactly supported. Furthermore, if

$$\min_{0\le x,y\le 1} f(x,y) > 0, \tag{5.6}$$

then μ is absolutely continuous with respect to Lebesgue measure.

Our second theorem is the analogue of Theorem 5.1.2 with A_N replaced by Δ_N.

Theorem 5.1.3 (Existence of the limiting spectral distribution of Δ_N). *There exists a symmetric probability measure ν on $\mathbb{R}$ such that*

$$\lim_{N\to\infty} \mathrm{ESD}\big((N\varepsilon_N)^{-1/2}(\Delta_N - D_N)\big) = \nu \quad \text{weakly in probability,} \tag{5.7}$$

where

$$D_N = \mathrm{Diag}\left(\mathbb{E}\big[\Delta_N(1,1)\big], \dots, \mathbb{E}\big[\Delta_N(N,N)\big]\right). \tag{5.8}$$

Furthermore, if

$$f \not\equiv 0, \tag{5.9}$$

then the support of ν is unbounded.

The ESD of a random matrix is a random probability measure. Note that μ and ν are both deterministic, i.e., a law of large numbers is in force.

Theorems 5.1.2–5.1.3 are existential, in the sense that explicit descriptions of μ and ν are missing. We have some control on the Stieltjes transform of μ. In the proof of Theorem 5.1.2 (in Lemma 5.2.3) we will see that the ESD of $(N\varepsilon_N)^{-1/2}A_N$ has the same limit as the ESD of

$$\bar{A}_N(i,j) = \sqrt{\frac{1}{N} f\left(\frac{i}{N}, \frac{j}{N}\right)}\, G_{i\wedge j, i\vee j} \tag{5.10}$$

with $(G_{i,j} : 1 \le i \le j)$ a family of i.i.d. standard Gaussian random variables. Such random matrices are known in the literature as Wigner matrices with a variance profile (see, for example, [107], [129], [164], [185]). The limiting spectral distribution of $\bar{A}_N$ matches with the one of certain symmetric random matrices with dependent entries

(see [135] for details). It turns out that, by using the combinatorics of non-crossing partitions, we can derive a recursive equation for the Stieltjes transform of μ, i.e.,

$$G_\mu(z) = \int_{\mathbb{R}} \frac{1}{z-x}\,\mu(dx), \quad z \in \mathbb{C} \setminus \mathbb{R}. \tag{5.11}$$

It turns out that

$$G_\mu(z) = \int_0^1 \mathcal{H}(z,x)\,dx, \tag{5.12}$$

where $\mathcal{H}(z,x)$, $x \in [0,1]$, is the unique analytic solution of the integral equation

$$z\mathcal{H}(z,x) = 1 + \mathcal{H}(z,x)\int_0^1 \mathcal{H}(z,y)f(x,y)\,dy, \qquad x \in [0,1]. \tag{5.13}$$

The form of $\mathcal{H}(z,x)$ can also be expressed in terms of non-crossing partitions and the function $f(x,y)$ (see [134, Section 4.1] for details). We mention that the above measure is similar to the limiting measure in [196, Theorem 3.4]. There it is shown that a graphon sequence W_N can be associated with a Wigner matrix with a variance profile $(s_{i,j} : 1 \le i,j \le N)$. If the sequence of graphons W_N converges in the cut norm to W with $W(x,y) = f(x,y)$, then the limiting measure matches with μ.

The description of ν through its Stieltjes transform is hard to obtain, although, just like before, the ESD of $(N\varepsilon_N)^{-1/2}(\Delta_N - D_N)$ turns out to be the same as that of

$$\tilde{\Delta}_N = \bar{A}_N + Y_N, \tag{5.14}$$

where Y_N is a diagonal matrix of order N defined by

$$Y_N(i,i) = Z_i\sqrt{\frac{1}{N}\sum_{1\le j\le N,\, j\ne i} f\Big(\frac{i}{N},\frac{j}{N}\Big)}, \qquad 1 \le i \le N, \tag{5.15}$$

where $(Z_i\colon i \ge 1)$ is a family of i.i.d. standard normal random variables, independent of $(G_{i,j}\colon 1 \le i \le j)$. Suppose that Y_N is a deterministic diagonal matrix, embedded in $L^\infty[0,1]$ (as a step function). For the case where this function converges to a function h in the $\|\cdot\|_\infty$ norm, the limiting spectral distribution of $\bar{A}_N + Y_N$ was studied in [185] (see also [186, Theorem 22.7.2]). In our case, due to the presence in Y_N of Gaussian random variables (which have unbounded support) and the fact that the spectral norm of Y_N tends to infinity as $N \to \infty$, the existing results cannot be applied. One of the major contributions of our paper is to overcome this hurdle. Also, our proofs ensure that ν has a finite moment generating function (see (5.123) below) and unbounded support.

§5.1.3 Identification of the limiting spectral distribution

Our next theorem identifies μ and ν under the additional assumption that f has a *multiplicative structure*, i.e.,

$$f(x,y) = r(x)r(y), \qquad (x,y) \in [0,1]^2, \tag{5.16}$$

for some continuous function $r\colon [0,1] \to [0,\infty)$. The statement is based on the theory of (possibly unbounded) self-adjoint operators affiliated with a W^*-probability space. Recall Section 1.2.4 for an introduction to free probability theory. A few extra relevant definitions are given below. For details the reader is referred to [108, Section 5.2.3].

Definition 5.1.4 (Operators affiliated with a W^*-probability space). A C^*-algebra $\mathcal{A} \subset B(\mathcal{H})$, with $\mathcal{H}$ a Hilbert space, is a W^*-algebra when $\mathcal{A}$ is closed under the weak operator topology. If, in addition, τ is a state such that there exists a unit vector $\xi \in \mathcal{H}$ satisfying

$$\tau(a) = \langle a\xi, \xi \rangle \qquad \forall\, a \in \mathcal{H}, \tag{5.17}$$

then $(\mathcal{A}, \tau)$ is a W^*-probability space. In that case a densely defined self-adjoint (possibly unbounded) operator T on $\mathcal{H}$ is said to be *affiliated with* $\mathcal{A}$ if $h(T) \in \mathcal{A}$ for any bounded measurable function h defined on the spectrum of T, where $h(T)$ is defined by the spectral theorem. Finally, for an affiliated operator T, its *law* $\mathcal{L}(T)$ is the unique probability measure on $\mathbb{R}$ satisfying

$$\tau(h(T)) = \int_{\mathbb{R}} h(x)(\mathcal{L}(T))(dx) \tag{5.18}$$

for every bounded measurable $h\colon \mathbb{R} \to \mathbb{R}$.

The distribution of a single self-adjoint operator is defined above. For two or more self-adjoint operators $T_1, \ldots, T_n$, a description of their *joint distribution* is a specification of

$$\tau(h_1(T_{i_1}) \cdots h_k(T_{i_k})), \tag{5.19}$$

for all $k \geq 1$, all $i_1, \ldots, i_k \in \{1, \ldots, n\}$, and all bounded measurable functions $h_1, \ldots, h_k$ from $\mathbb{R}$ to itself. Once the above is specified, it is immediate to see that $\mathcal{L}(p(T_1, \ldots, T_k))$ can be calculated for any polynomial p in k variables such that $p(T_1, \ldots, T_k)$ is self-adjoint.

Definition 5.1.5 (Free independence of operators). Let $(\mathcal{A}, \tau)$ be a W^*-probability space and $a_1, a_2 \in \mathcal{A}$. Then a_1 and a_2 are *freely independent* if

$$\tau(p_1(a_{i_1}) \cdots p_n(a_{i_n})) = 0, \tag{5.20}$$

for all $n \geq 1$, all $i_1, \ldots, i_n \in \{1, 2\}$ with $i_j \neq i_{j+1}$, $j = 1, \ldots, n-1$, and all polynomials $p_1, \ldots, p_n$ in one variable satisfying

$$\tau(p_j(a_{i_j})) = 0, \qquad j = 1, \ldots, n. \tag{5.21}$$

For (possibly unbounded) operators $a_1, \ldots, a_k$ and $b_1, \ldots, b_m$ affiliated with $\mathcal{A}$, the collections $(a_1, \ldots, a_k)$ and $(b_1, \ldots, b_m)$ are *freely independent* if and only if

$$p(h_1(a_1), \ldots, h_k(a_k)) \quad \text{and} \quad q(g_1(b_1), \ldots, g_m(b_m)), \tag{5.22}$$

are freely independent for all bounded measurable $h_1, \ldots, h_k$ and $g_1, \ldots, g_m$, and all polynomials p and q in k and m non-commutative variables, respectively. It is immediate that the two operators in the above display are bounded, and hence belong to $\mathcal{A}$.

We are now in a position to state our next theorem.

Theorem 5.1.6 (Identification of the limiting spectral distribution). *If f is as in* (5.16)*, then*

$$\mu = \mathcal{L}\bigg(r^{1/2}(T_u)T_s r^{1/2}(T_u)\bigg), \tag{5.23}$$

and

$$\nu = \mathcal{L}\bigg(r^{1/2}(T_u)T_s r^{1/2}(T_u) + \alpha r^{1/4}(T_u)T_g r^{1/4}(T_u)\bigg), \tag{5.24}$$

where

$$\alpha = \bigg(\int_0^1 r(x)\,dx\bigg)^{1/2}. \tag{5.25}$$

Here, T_g and T_u are commuting self-adjoint operators affiliated with a W^*-probability space $(\mathcal{A}, \tau)$ such that, for bounded measurable functions h_1, h_2 from $\mathbb{R}$ to itself,

$$\tau(h_1(T_g)h_2(T_u)) = \bigg(\int_{\mathbb{R}} h_1(x)\phi(x)\,dx\bigg)\bigg(\int_0^1 h_2(u)\,du\bigg), \tag{5.26}$$

with ϕ the standard normal density. Furthermore, T_s has a standard semicircle distribution and is freely independent of (T_g, T_u).

The right-hand side of (5.23) is the same as the free multiplicative convolution of the standard semicircle law and the law of $r(U)$, where U is a standard uniform random variable.

The fact that T_g and T_u commute, together with (5.26), specifies their joint distribution. In fact, they are standard normal and standard uniform, respectively, independently of each other in the *classical sense*. Free independence of T_s and (T_g, T_u), plus the fact that the former follows the standard semicircle law, specifies the joint distribution of T_s, T_g, T_u.

In order to admit the unbounded operator T_g, a W^*-probability space is needed. If all the operators would have been bounded, then a C^*-probability space would have sufficed.

§5.1.4 Randomization

Theorem 5.1.2 can be generalized to the situation where the function f is random. Such a randomization helps us to address the applications listed in Section 5.4. Suppose that $(\varepsilon_N\colon N \geq 1)$ is a sequence of positive numbers satisfying (5.2). Suppose further that, for every $N \geq 1$, $(R_{Ni}\colon 1 \leq i \leq N)$ is a collection of non-negative random variables (defined on the same probability space) such that there is a deterministic $C < \infty$ for which

$$\sup_{N\geq 1} \max_{1\leq i\leq N} R_{Ni} \leq C \quad \text{almost surely.} \tag{5.27}$$

In addition, suppose that there is a probability measure μ_r on $\mathbb{R}$ such that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \delta_{R_{Ni}} = \mu_r \quad \text{weakly almost surely.} \tag{5.28}$$

The non-negativity of R_{Ni} and (5.27) ensure that μ_r is concentrated on $[0, C]$. Furthermore, the first line of (5.2) ensures that the additional assumption

$$\sup_{N\geq 1} \varepsilon_N \leq \frac{1}{C^2} \tag{5.29}$$

entails no loss of generality.

For fixed N and conditional on $(R_{N1}, \ldots, R_{NN})$, the random graph G_N is constructed as before, except that there is an edge between i and j with probability $\varepsilon_N R_{Ni} R_{Nj}$, which is at most 1 by (5.29) for all $1 \leq i < j \leq N$. In other words, G_N has two levels of randomness: one in the choice of $(R_{N1}, \ldots, R_{NN})$ and one in the choice of the set of edges. Once again, A_N is the adjacency matrix of G_N. The following is a *randomized* version of Theorem 5.1.2.

Theorem 5.1.7 (Limiting spectral distribution of A_N).
Under the assumptions (5.2) *and* (5.27)–(5.28)*,*

$$\lim_{N\to\infty} \mathrm{ESD}\big((N\varepsilon_N)^{-1/2} A_N\big) = \mu_r \boxtimes \mu_s \quad \textit{weakly in probability,} \tag{5.30}$$

where μ_s is the standard semicircle law.

§5.1.5 Applications and outline

As we will see in Section 5.4, our results can be applied in various ways. A first application consists in *constrained random graphs*. Given a sequence of positive integers, among the probability distributions for which the sequence of average degrees matches the given sequence, called the soft configuration model, the one that maximizes the entropy is the canonical Gibbs measure. It is known that, under a sparsity condition, the connection probabilities arising out of the canonical Gibbs measure asymptotically have a multiplicativestructure (see [187]). We show that our results on the adjacency matrix can be easily extended to cover such situations. Another important application consists in *Chung-Lu type random graph*, which are used to model *sociability patterns in networks*. We show how to use the rescaled empirical spectral distribution and free probability to statistically recover the underlying sociability distribution.

Outline of the chapter. The remainder of this chapter is organized as follows. In Section 5.2 a number of technical lemmas are proved. These serve as preparation for the proofs of our main theorems, which are given in Section 5.3. In Section 5.4, the above applications are discussed, organized into three propositions. Appendix E collects a few basic facts that are needed along the way.

§5.2 Preparatory approximations

The proofs of our main theorems rely on several preparatory approximations, which we organize in Lemmas 5.2.1–5.2.4 and 5.2.6 below. Along the way we need several basic results, which we collect in Appendix E.

§5.2.1 Centering

The first approximation is that the mean of each off-diagonal entry of A_N and Δ_N can be subtracted, with negligible perturbation in the respective empirical spectral distributions.

Lemma 5.2.1 (Centering).
Let A_N^0 and Δ_N^0 be $N \times N$ matrices defined by

$$\begin{aligned} A_N^0(i,j) &= (N\varepsilon_N)^{-1/2}\big(A_N(i,j) - \mathbb{E}[A_N(i,j)]\big), &(5.31)\\ \Delta_N^0(i,j) &= (N\varepsilon_N)^{-1/2}\big(\Delta_N(i,j) - \mathbb{E}[\Delta_N(i,j)]\big), &(5.32)\end{aligned}$$

for all $1 \le i,j \le N$. Then

$$\begin{aligned} &\lim_{N\to\infty} L\big(\mathrm{ESD}(A_N^0), \mathrm{ESD}((N\varepsilon_N)^{-1/2}A_N)\big) = 0 \quad \textit{in probability},\\ &\lim_{N\to\infty} L\big(\mathrm{ESD}(\Delta_N^0), \mathrm{ESD}((N\varepsilon_N)^{-1/2}(\Delta_N - D_N))\big) = 0 \quad \textit{in probability}, \end{aligned} \tag{5.33}$$

where $L(\eta_1,\eta_2)$ denotes the Lévy distance between the probability measures η_1 and η_2, and D_N is the diagonal matrix defined in (5.8).

Proof. An appeal to Lemma E.1 shows that

$$\begin{aligned} &L^3\big(\mathrm{ESD}(A_N^0), \mathrm{ESD}((N\varepsilon_N)^{-1/2}A_N)\big)\\ &\le \frac{1}{N^2\varepsilon_N}\sum_{i,j=1}^N \mathbb{E}^2[A_N(i,j)]\\ &= \frac{1}{N^2\varepsilon_N}\sum_{i\neq j}\varepsilon_N^2 f^2\left(\frac{i}{N},\frac{j}{N}\right)\\ &= \varepsilon_N\int_{[0,1]^2} f^2(x,y)\,dx\,dy\,[1+o(1)], \qquad N\to\infty. \end{aligned} \tag{5.34}$$

The first claim follows by recalling that $\varepsilon_N \to 0$. The proof the second claim is verbatim the same. □

§5.2.2 Gaussianisation

One of the crucial steps in studying the scaling properties of ESD is to replace each entry by a Gaussian random variable.

Lemma 5.2.2 (Gaussianisation).
Let $(G_{i,j}\colon 1 \le i \le j)$ be a family of i.i.d. standard Gaussian random variables. Define $N \times N$ matrices A_N^g and Δ_N^g by

$$A_N^g(i,j) = \begin{cases} \sqrt{\frac{1}{N} f\big(\frac{i}{N}, \frac{j}{N}\big)\big(1 - \varepsilon_N f\big(\frac{i}{N}, \frac{j}{N}\big)\big)}\, G_{i\wedge j, i\vee j}, & i \neq j, \\ 0, & i = j, \end{cases} \tag{5.35}$$

$$\Delta_N^g(i,j) = \begin{cases} A_N^g(i,j), & i \neq j, \\ -\sum_{k=1, k\neq i}^N A_N^g(i,k), & i = j. \end{cases} \tag{5.36}$$

Fix $z \in \mathbb{C} \setminus \mathbb{R}$ and a three times continuously differentiable function $h\colon \mathbb{R} \to \mathbb{R}$ such that

$$\max_{0 \le j \le 3} \sup_{x \in \mathbb{R}} |h^{(j)}(x)| < \infty. \tag{5.37}$$

For an $N \times N$ real symmetric matrix M, define

$$H_N(M) = \frac{1}{N} \operatorname{Tr}\big((M - zI_N)^{-1}\big), \tag{5.38}$$

where I_N is the identity matrix of order N. Then

$$\lim_{N\to\infty} \mathbb{E}\big[h\big(\Re H_N(A_N^g)\big) - h\big(\Re H_N(A_N^0)\big)\big] = 0, \tag{5.39}$$

$$\lim_{N\to\infty} \mathbb{E}\big[h\big(\Im H_N(A_N^g)\big) - h\big(\Im H_N(A_N^0)\big)\big] = 0, \tag{5.40}$$

and

$$\lim_{N\to\infty} \mathbb{E}\big[h\big(\Re H_N(\Delta_N^g)\big) - h\big(\Re H_N(\Delta_N^0)\big)\big] = 0, \tag{5.41}$$

$$\lim_{N\to\infty} \mathbb{E}\big[h\big(\Im H_N(\Delta_N^g)\big) - h\big(\Im H_N(\Delta_N^0)\big)\big] = 0, \tag{5.42}$$

where $\Re$ and $\Im$ denote the real and the imaginary part of a complex number, respectively.

Proof. We only prove (5.41). The proofs of the other claims are similar. We use ideas from [136]. Let $z = u + iv \in \mathbb{C}^+$ and $n = N(N-1)/2$. Define $\phi\colon \mathbb{R}^n \to \mathbb{C}$ as

$$\phi(x) = H_N(\Delta(x)) \tag{5.43}$$

where $\Delta(x)$ is the $N \times N$ symmetric Laplacian matrix given by

$$\Delta(x)(i,j) = \begin{cases} -\sum_{k=1,k\neq i}^N x_{i,k}, & i = j, \\ x_{i\wedge j, i\vee j}, & i \neq j. \end{cases} \tag{5.44}$$

Note that $\partial\Delta(x)/\partial x_{ij}$ is the $N \times N$ matrix that has -1 at the i-th and j-th diagonal and 1 at (i,j)-th and (j,i)-th entry. The following identities were derived in [136,

Section 2]:

$$\begin{aligned}\frac{\partial\phi}{\partial x_{i,j}} &= -N^{-1}\operatorname{Tr}\left(\frac{\partial\Delta}{\partial x_{i,j}}K^2\right),\\ \frac{\partial^2\phi}{\partial x_{i,j}^2} &= 2N^{-1}\operatorname{Tr}\left(\frac{\partial\Delta}{\partial x_{i,j}}K\frac{\partial\Delta}{\partial x_{i,j}}K^2\right),\\ \frac{\partial^3\phi}{\partial x_{i,j}^3} &= -6N^{-1}\operatorname{Tr}\left(\frac{\partial\Delta}{\partial x_{i,j}}K\frac{\partial\Delta}{\partial x_{i,j}}K\frac{\partial\Delta}{\partial x_{i,j}}K^2\right),\end{aligned} \tag{5.45}$$

where $K(x) = (\Delta(x) - zI)^{-1}$. Now using these identities we get

$$\left\|\frac{\partial\phi}{\partial x_{ij}}\right\|_\infty \le \frac{4}{|\Im z|^2}\frac{1}{N}, \qquad \left\|\frac{\partial^2\phi}{\partial x_{ij}^2}\right\|_\infty \le \frac{8}{|\Im z|^3}\frac{1}{N}, \qquad \left\|\frac{\partial^3\phi}{\partial x_{ij}^3}\right\|_\infty \le \frac{48}{|\Im z|^4}\frac{1}{N}. \tag{5.46}$$

If we define

$$\lambda_2(\phi) = \sup\left\{\left\|\frac{\partial\phi}{\partial x_{i,j}}\right\|_\infty^2,\ \left\|\frac{\partial^2\phi}{\partial x_{i,j}^2}\right\|_\infty\right\}, \tag{5.47}$$

$$\lambda_3(\phi) = \sup\left\{\left\|\frac{\partial\phi}{\partial x_{i,j}}\right\|_\infty^3, \left\|\frac{\partial^2\phi}{\partial x_{i,j}^2}\right\|_\infty^2, \left\|\frac{\partial^3\phi}{\partial x_{i,j}^3}\right\|_\infty\right\}, \tag{5.48}$$

then there exist constants C_2 and C_3 depending on $\Im z$ such that $\lambda_2(\phi) \le C_2N^{-1}$ and $\lambda_3(\phi) \le C_3N^{-1}$. Hence, using $\lambda_r(\Re\phi) \le \lambda_r(\phi)$ and

$$U = \Re\big(H_N(\Delta_N^0)\big), \qquad V = \Im\big(H_N(\Delta_N^q)\big), \tag{5.49}$$

we have from [136, Theorem 1.1]

$$\begin{aligned}&\big|\mathbb{E}[h(U)] - \mathbb{E}[h(V)]\big|\\ &\le C_1(h)\lambda_2(\phi)\sum_{1\le i\ne j\le N}\big(\mathbb{E}[A_N^0(i,j)^2; |A_N^0(i,j)| > K]\\ &\qquad\qquad + \mathbb{E}[A_N^q(i,j)^2; |A_N^q(i,j)| > K]\big)\\ &\quad + C_2(h)\frac{\lambda_3(\phi)}{(N\varepsilon_N)^{3/2}}\sum_{i\ne j}\big(\mathbb{E}[A_N^0(i,j); |A_N^0(i,j)| > K]\\ &\qquad\qquad + \mathbb{E}[A_N^q(i,j)^3; |A_N^q(i,j)| > K]\big).\end{aligned} \tag{5.50}$$

Using the fact that $\varepsilon_N \to 0$, we have that $\mathbb{E}[A_N^0(i,j)^4] = \mathrm{O}(N^{-2}\varepsilon_N^{-1})$. Also

$$\mathbb{P}(|A_N^0(i,j)| > K) \le \mathrm{O}(N^{-1}). \tag{5.51}$$

So, by the Cauchy-Schwartz inequality and the above bounds, we have

$$\mathbb{E}[A_N^0(i,j)^2; |A_N^0(i,j)| > K] \le \mathrm{O}\big(\varepsilon_N^{-1/2}N^{-3/2}\big). \tag{5.52}$$

Since $N\varepsilon_N \to \infty$, we have

$$\lambda_2(\phi)\sum_{1\le i\ne j\le N}\mathbb{E}[A_N^0(i,j)^2; |A_N^0(i,j)| > K] \le CN^{-1/2}\varepsilon_N^{-1/2}, \tag{5.53}$$

which tends to 0 as $N \to \infty$. Similarly, we have

$$\lambda_3(\phi) \sum_{i \neq j} \mathbb{E}[A_N^0(i,j)^3; |A_N^0(i,j)| > K] \leq \frac{C}{N^{5/2} \varepsilon_N^{3/2}} N^2 \varepsilon_N, \tag{5.54}$$

which also tends to 0 as $N \to \infty$. Using Gaussian tail bounds, we can also show that the other two terms in (5.50) tend to 0 as $N \to \infty$, which settles (5.41). In order to prove (5.42), a similar computation can be done for the imaginary part in (5.49). The proofs of (5.39) and (5.40) are analogous (and, in fact, closer to the argument in [136]). □

§5.2.3 Leading order variance

Next, we show that another minor tweak to the entries of A_N^g and Δ_N^g results in a negligible perturbation.

Lemma 5.2.3 (Leading order variance).
Define an $N \times N$ matrix A_N by

$$\bar{A}_N(i,j) \quad = \quad \sqrt{\frac{1}{N} f\left(\frac{i}{N}, \frac{j}{N}\right)}\, G_{i \wedge j, i \vee j}, \qquad 1 \leq i,j \leq N, \tag{5.55}$$

and let

$$\bar{\Delta}_N = \bar{A}_N - X_N, \tag{5.56}$$

where X_N is a diagonal matrix of order N defined by

$$X_N(i,i) = \sum_{k=1, k \neq i}^{N} \bar{A}_N(i,k), \qquad 1 \leq i \leq N. \tag{5.57}$$

Then

$$\lim_{N \to \infty} L\big(\mathrm{ESD}(A_N^g), \mathrm{ESD}(\bar{A}_N)\big) \quad = \quad 0 \quad \textit{in probability}, \tag{5.58}$$

$$\lim_{N \to \infty} L\big(\mathrm{ESD}(\Delta_N^g), \mathrm{ESD}(\bar{\Delta}_N)\big) \quad = \quad 0 \quad \textit{in probability}. \tag{5.59}$$

Proof. To prove (5.59), yet another application of Lemma E.1 implies that

$$\begin{aligned}
&\mathbb{E}\big[L^3\big(\mathrm{ESD}(\Delta_N^g), \mathrm{ESD}(\bar{\Delta}_N)\big)\big] \\
&\le \frac{1}{N}\mathbb{E}\big[\operatorname{Tr}\big((\Delta_N^g - \bar{\Delta}_N)^2\big)\big] \\
&= \frac{1}{N}\sum_{1\le i\ne j\le N} \mathrm{Var}(\bar{A}_N(i,j) - A_N^g(i,j)) \\
&\quad + \frac{1}{N}\sum_{i=1}^N \mathrm{Var}\bigg(\sum_{j=1, j\ne i}^N (\bar{A}_N(i,j) - A_N^g(i,j))\bigg) + \frac{1}{N^2}\sum_{i=1}^N f\bigg(\frac{i}{N}, \frac{i}{N}\bigg) \qquad (5.60)\\
&= \frac{4}{N^2}\sum_{1\le i<j\le N} f\bigg(\frac{i}{N}, \frac{j}{N}\bigg)\bigg(1 - \sqrt{1 - \varepsilon_N f\bigg(\frac{i}{N}, \frac{j}{N}\bigg)}\bigg)^2 \\
&\quad + \frac{1}{N^2}\sum_{i=1}^N f\bigg(\frac{i}{N}, \frac{i}{N}\bigg),
\end{aligned}$$

which tends to 0 as $N \to \infty$ because f is bounded. Thus, (5.59) follows. The proof of (5.58) is similar. □

§5.2.4 Decoupling

The (diagonal) entries of X_N are nothing but the row sums of $\bar{A}_N$. However, the correlation between an entry of $\bar{A}_N$ and that of X_N is small. The following decoupling lemma shows that it does not hurt when the entries of X_N are replaced by a mean-zero Gaussian random variable of the same variance that is independent of $\bar{A}_N$.

Lemma 5.2.4 (Decoupling).
Let $(Z_i\colon i \ge 1)$ be a family of i.i.d. standard normal random variables, independent of $(G_{i,j}\colon 1 \le i \le j)$. Define a diagonal matrix Y_N of order N by

$$Y_N(i,i) = Z_i\sqrt{\frac{1}{N}\sum_{j=1, j\ne i}^N f\bigg(\frac{i}{N}, \frac{j}{N}\bigg)}, \qquad 1 \le i \le N, \qquad (5.61)$$

and let

$$\tilde{\Delta}_N = \bar{A}_N + Y_N. \qquad (5.62)$$

Then, for every $k \in \mathbb{N}$,

$$\lim_{N\to\infty}\frac{1}{N}\mathbb{E}\big[\operatorname{Tr}\big((\tilde{\Delta}_N)^{2k} - (\bar{\Delta}_N)^{2k}\big)\big] = 0, \qquad (5.63)$$

and

$$\lim_{N\to\infty}\frac{1}{N^2}\mathbb{E}\big[\operatorname{Tr}^2\big((\tilde{\Delta}_N)^k\big) - \operatorname{Tr}^2\big((\bar{\Delta}_N)^k\big)\big] = 0. \qquad (5.64)$$

Proof. Without loss of generality we may assume that $f \leq 1$. For $N \geq 1$, define the $N \times N$ matrices $\bar{M}_N$ and $\tilde{M}_N$ by

$$\bar{M}_N(i,j) = \begin{cases} N^{-1/2} G_{i\wedge j, i\vee j}, & i \neq j, \\ N^{-1/2} G_{i,i} - \sum_{k=1,\, k\neq i}^{N} \bar{M}_N(i,k), & i = j, \end{cases} \tag{5.65}$$

and

$$\tilde{M}_N(i,j) = \begin{cases} \bar{M}_N(i,j), & i \neq j, \\ N^{-1/2} G_{i,i} + Z_i \sqrt{\frac{N-1}{N}}, & i = j. \end{cases} \tag{5.66}$$

Note that, in the special case where f is identically 1, $\bar{M}_N$ and $\tilde{M}_N$ are identical to $\bar{\Delta}_N$ and $\tilde{\Delta}_N$, respectively. For $k \in \mathbb{N}$ and Π a partition of $\{1, \ldots, 2k\}$, let

$$\Psi(\Pi, N) = \Big\{ i \in \{1, \ldots, N\}^{2k} \colon\ i_u = i_v \iff u, v \text{ belong to the same block of } \Pi \Big\}. \tag{5.67}$$

For fixed Π and N, an immediate application of Wick's formula shows that, for all $i, j \in \Psi(\Pi, N)$,

$$\mathbb{E}\Big[\prod_{u=1}^{2k} \bar{M}_N(i_u, i_{u+1}) \Big] = \mathbb{E}\Big[\prod_{u=1}^{2k} \bar{M}_N(j_u, j_{u+1}) \Big], \tag{5.68}$$

with the convention that $i_{2k+1} \equiv i_1$, and

$$\mathbb{E}\Big[\prod_{u=1}^{2k} \tilde{M}_N(i_u, i_{u+1}) \Big] = \mathbb{E}\Big[\prod_{u=1}^{2k} \tilde{M}_N(j_u, j_{u+1}) \Big], \tag{5.69}$$

Therefore, for any $i \in \Psi(\Pi, N)$, we can unambiguously define

$$\psi(\Pi, N) = \mathbb{E}\Big[\prod_{u=1}^{2k} \bar{M}_N(i_u, i_{u+1}) \Big] - \mathbb{E}\Big[\prod_{u=1}^{2k} \tilde{M}_N(i_u, i_{u+1}) \Big]. \tag{5.70}$$

As shown in [128, Lemma 4.12], for a fixed Π,

$$\lim_{N\to\infty} N^{-1} |\psi(\Pi, N)||\Psi(\Pi, N)| = 0. \tag{5.71}$$

An immediate observation is that, for all $1 \leq i, j, i', j' \leq N$,

$$\mathrm{Cov}\big(\tilde{M}_N(i,j), \tilde{M}_N(i',j')\big) = 0 \quad \text{if} \quad (i \wedge j, i \vee j) \neq (i' \wedge j', i' \vee j'), \tag{5.72}$$

and likewise for $\tilde{\Delta}_N$. Furthermore,

$$\mathrm{Var}\big(\tilde{M}_N(i,j)\big) = \mathrm{Var}\big(\bar{M}_N(i,j)\big), \qquad 1 \leq i, j \leq N, \tag{5.73}$$

and likewise for $\tilde{\Delta}_N$ and $\bar{M}_N$. For $N \geq 1$ and $1 \leq i, j, i', j' \leq N$, define

$$\eta_N(i,j,i',j') = \begin{cases} \frac{\mathrm{Cov}\left(\bar{\Delta}_N(i,j), \bar{\Delta}_N(i',j')\right)}{\mathrm{Cov}\left(\bar{M}_N(i,j), \bar{M}_N(i',j')\right)}, & \text{if the denominator is non-zero,} \\ 0, & \text{otherwise.} \end{cases} \tag{5.74}$$

It is easy to check that the assumption $f \leq 1$ ensures that $|\eta_N(i,j,i',j')| \leq 1$. Therefore, for all N and $1 \leq i,j,i',j' \leq N$,

$$\begin{aligned}\mathrm{Cov}\big(\bar{\Delta}_N(i,j),\bar{\Delta}_N(i',j')\big) &= \eta_N(i,j,i',j')\mathrm{Cov}\big(\bar{M}_N(i,j),\bar{M}_N(i',j')\big),\\ \mathrm{Cov}\big(\tilde{\Delta}_N(i,j),\tilde{\Delta}_N(i',j')\big) &= \eta_N(i,j,i',j')\mathrm{Cov}\big(\tilde{M}_N(i,j),\tilde{M}_N(i',j')\big).\end{aligned}$$

For fixed Π, N and $i \in \Psi(\Pi,N)$, by an appeal to Wick's formula the above implies that there exists a $\xi(i,N) \in [-1,1]$ such that

$$\mathbb{E}\Big[\prod_{u=1}^{2k}\bar{\Delta}_N(i_u,i_{u+1})\Big] - \mathbb{E}\Big[\prod_{u=1}^{2k}\tilde{\Delta}_N(i_u,i_{u+1})\Big] = \xi(i,N)\psi(\Pi,N), \tag{5.75}$$

and therefore, by (5.71),

$$\begin{aligned}&\sum_{i\in\Psi(\Pi,N)}\Big|\mathbb{E}\Big[\prod_{u=1}^{2k}\bar{\Delta}_N(i_u,i_{u+1})\Big] - \mathbb{E}\Big[\prod_{u=1}^{2k}\tilde{\Delta}_N(i_u,i_{u+1})\Big]\Big|\\ &\quad= \sum_{i\in\Psi(\Pi,N)}|\xi(i,N)||\psi(\Pi,N)| \leq |\psi(\Pi,N)||\Psi(\Pi,N)| = o(N), \qquad N\to\infty.\end{aligned} \tag{5.76}$$

Since this holds for every partition Π of $\{1,\ldots,2k\}$, (5.63) follows. The proof of (5.64) follows along similar lines. □

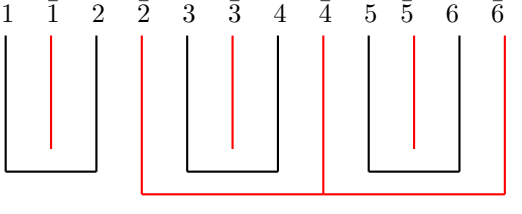
§5.2.5 Combinatorics from free probability

The final preparation is a general result from random matrix theory. To state this, the following notions from the theory of free probability are borrowed. We refer to Section 1.2.4 for an introduction to free probability theory and to [186] for more details.

Definition 5.2.5 (Kreweras complement).
For an even positive integer k, $NC_2(k)$ is the set of non-crossing pair partitions of $\{1,\ldots,k\}$. For $\sigma \in NC_2(k)$, its *Kreweras complement* $K(\sigma)$ is the maximal non-crossing partition $\bar{\sigma}$ of $\{\bar{1},\ldots,\bar{k}\}$, such that $\sigma \cup \bar{\sigma}$ is a non-crossing partition of $\{1,\bar{1},\ldots,k,\bar{k}\}$. For example,

$$\begin{aligned}K(\{(1,4),(2,3)\}) &= \{(1,3),(2),(4)\},\\ K(\{(1,2),(3,4),(5,6)\}) &= \{(1),(2,4,6),(3),(5)\}.\end{aligned} \tag{5.77}$$

The second example is illustrated as



For $\sigma \in NC_2(k)$ and $N \geq 1$, define

$$S(\sigma, N) = \big\{ i \in \{1, \dots, N\}^k \colon\ i_u = i_v \iff u, v \text{ belong to the same block of } K(\sigma) \big\} \tag{5.78}$$

and

$$C(k, N) = \{1, \dots, N\}^k \setminus \Big(\bigcup_{\sigma \in NC_2(k)} S(\sigma, N) \Big). \tag{5.79}$$

In other words, $S(\sigma, N)$ is the same as $\Psi(K(\sigma), N)$ defined in (5.67).

Lemma 5.2.6 (Trace of product of random matrices).
Suppose that, for each $N \geq 1$, $W_{N,1}, \dots, W_{N,k}$ are $N \times N$ real (and possibly asymmetric) random matrices, where k is a positive even number. Suppose further that, for each $u = 1, \dots, k$,

$$\max_{1 \leq i,j \leq N} \mathbb{E}\big[W_{N,u}(i,j)^k \big] = O(N^{-k/2}) \tag{5.80}$$

and

$$\lim_{N \to \infty} \mathbb{E}\bigg[\Big(\frac{1}{N} \sum_{i \in C(k,N)} P_i \Big)^2 \bigg] = 0, \tag{5.81}$$

and that, for every $\sigma \in NC_2(k)$, there exists a deterministic and finite $\beta(\sigma)$ such that

$$\lim_{N \to \infty} \mathbb{E}\Big(\frac{1}{N} \sum_{i \in S(\sigma,N)} P_i \Big) = \beta(\sigma), \tag{5.82}$$

$$\lim_{N \to \infty} \mathbb{E}\bigg[\Big(\frac{1}{N} \sum_{i \in S(\sigma,N)} P_i \Big)^2 \bigg] = \beta(\sigma)^2, \tag{5.83}$$

where

$$P_i = W_{N,1}(i_1, i_2) \cdots W_{N,k-1}(i_{k-1}, i_k) W_{N,k}(i_k, i_1), \qquad i \in \{1, \dots, N\}^k. \tag{5.84}$$

Furthermore, let $V_1, V_2, \dots$ be i.i.d. random variables drawn from some distribution with all moments finite, independent of $(W_{N,j} \colon\ N \geq 1,\ 1 \leq j \leq k)$, and let

$$U_N = \mathrm{Diag}(V_1, \dots, V_N), \qquad N \geq 1. \tag{5.85}$$

Then, for all choices of $n_1, \dots, n_k \geq 0$,

$$\lim_{N \to \infty} \frac{1}{N} \operatorname{Tr}\big(U_N^{n_1} W_{N,1} \cdots U_N^{n_k} W_{N,k} \big) = c \quad \text{in } L^2 \tag{5.86}$$

for some deterministic $c \in \mathbb{R}$ (depending on $k, n_1, \dots, n_k$).

Proof. The fact that the sets $S(\sigma, N)$ are disjoint for different $\sigma \in NC_2(k)$ allows us to write

$$\operatorname{Tr}\big(U_N^{n_1} W_{N,1} \cdots U_N^{n_k} W_{N,k} \big) = \sum_{\sigma \in NC_2(k)} \sum_{i \in S(\sigma,N)} \tilde{P}_i + \sum_{i \in C(k,N)} \tilde{P}_i, \tag{5.87}$$

where

$$\tilde{P}_i = \prod_{j=1}^{k} \big(V_{i_j}^{n_j} W_{N,j}(i_j, i_{j+1})\big), \qquad i \in \{1, \dots, N\}^k. \tag{5.88}$$

In order to show that the second sum in the right-hand side is negligible after scaling by N, the independence of $(V_1, V_2, \dots)$ and $(W_{N,j} \colon N \geq 1,\ 1 \leq j \leq k)$, together with the fact that the common distribution of the former has finite moments, implies that

$$\mathbb{E}\bigg[\bigg(\frac{1}{N} \sum_{i \in C(k,N)} \tilde{P}_i\bigg)^2\bigg] \leq K N^{-2} \sum_{i,j \in C(k,N)} \mathbb{E}[P_i P_j],$$

where K is a finite constant. Assumption (5.81) shows that

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i \in C(k,N)} \tilde{P}_i = 0 \quad \text{ in } L^2. \tag{5.89}$$

In order to complete the proof, it suffices to show that for every $\sigma \in NC_2(k)$ there exists a $\theta(\sigma) \in \mathbb{R}$ with

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i \in S(\sigma,N)} \tilde{P}_i = \theta(\sigma) \qquad \text{ in } L^2. \tag{5.90}$$

To that end, fix $\sigma \in NC_2(k)$ and note that, for $i \in S(\sigma, N)$,

$$\mathbb{E}[\tilde{P}_i] = \mathbb{E}[P_i] \mathbb{E}\bigg[\prod_{j=1}^{k} V_{i_j}^{n_j}\bigg] = \mathbb{E}[P_i] \prod_{u \in K(\sigma)} \mathbb{E}\Big[V_1^{\sum_{j \in u} n_j}\Big], \tag{5.91}$$

the product in the last line being taken over every block u of $K(\sigma)$. Putting

$$\theta(\sigma) = \beta(\sigma) \prod_{u \in K(\sigma)} \mathbb{E}\Big[V_1^{\sum_{j \in u} n_j}\Big], \tag{5.92}$$

we see that (5.82) gives

$$\lim_{N \to \infty} \mathbb{E}\bigg[\frac{1}{N} \sum_{i \in S(\sigma,N)} \tilde{P}_i\bigg] = \theta(\sigma). \tag{5.93}$$

Let us call $i, j \in \mathbb{N}^k$ "disjoint" if no coordinate of i matches any coordinate of j, i.e.,

$$\min_{1 \leq u,v \leq k} |i_u - j_v| \geq 1. \tag{5.94}$$

Since $K(\sigma)$ has exactly $\frac{1}{2}k + 1$ blocks, (5.80) implies that

$$\lim_{N \to \infty} N^{-2} \sum_{\substack{i,j \in S(\sigma,N) \\ i,j \text{ not disjoint}}} \mathbb{E}[\tilde{P}_i \tilde{P}_j] = 0. \tag{5.95}$$

If $i, j \in S(\sigma, N)$ are disjoint, then it is immediate that

$$\mathbb{E}[\tilde{P}_i \tilde{P}_j] = \Big(\prod_{u \in K(\sigma)} \mathbb{E}\big[V_1^{\sum_{j \in u} n_j}\big] \Big)^2 \mathbb{E}[P_i P_j]. \tag{5.96}$$

The above two displays, in conjunction with (5.83), show that

$$\lim_{N \to \infty} \mathbb{E}\bigg[\bigg(\frac{1}{N} \sum_{i \in S(\sigma, N)} \tilde{P}_i \bigg)^2 \bigg] = \theta(\sigma)^2. \tag{5.97}$$

This, along with (5.93), establishes (5.90), from which the proof follows. □

§5.3 Proofs of the main results

In this section we prove the theorems in Section 5.1. In Section 5.3.1 we prove Theorems 5.1.2–5.1.3 on the existence of the limiting spectral distributions of A_N and Δ_N. In Section 5.3.2 we identify those distributions by proving Theorem 5.1.6. In Section 5.3.3 we prove Theorem 5.1.7.

§5.3.1 Proof: existence

Proof of Theorem 5.1.2. From [129, Theorem 2.1] we know that, as $N \to \infty$,

$$\lim_{N \to \infty} \mathrm{ESD}(\bar{A}_N) = \mu \quad \text{weakly in probability,} \tag{5.98}$$

for a compactly supported symmetric probability measure μ. Lemma 5.2.3 immediately tells us that

$$\lim_{N \to \infty} \mathrm{ESD}(A_N^g) = \mu \quad \text{weakly in probability,} \tag{5.99}$$

and hence for h and H_N as in Lemma 5.2.2,

$$\lim_{N \to \infty} \mathbb{E}\big[h\big(\Re H_N(A_N^g)\big)\big] = h\bigg(\Re \int_{\mathbb{R}} \frac{1}{x - z}\, \mu(dx) \bigg). \tag{5.100}$$

The claim in (5.39) shows that A_N^g can be replaced by A_N^0 in the above display. Since the right-hand side is deterministic and the above holds for any h satisfying the hypothesis of Lemma 5.2.2, it follows that

$$\lim_{N \to \infty} \Re H_N(A_N^0) = \Re \int_{\mathbb{R}} \frac{1}{x - z}\, \mu(dx) \quad \text{in probability.} \tag{5.101}$$

A similar argument works for the imaginary part, which shows that

$$\lim_{N \to \infty} \mathrm{ESD}(A_N^0) = \mu \quad \text{weakly in probability.} \tag{5.102}$$

Lemma 5.2.1 completes the proof of (5.5).

Finally, if f is bounded away from 0, then the combination of [129, Lemma 3.1] and [120, Corollary 2] implies that μ is absolutely continuous with respect to the Lebesgue measure (see also [131]). □

A close inspection of the proof reveals that it suffices to assume that f is bounded and Riemann integrable instead of continuous. In other words, if f is symmetric and bounded, and its set of discontinuities has Lebesgue measure zero, then the result holds. However, continuity will be used later in (5.107) in the proof of Theorem 5.1.3. Furthermore, if $\varepsilon_N = 1$ for all N, then

$$\lim_{N\to\infty} \mathrm{ESD}\big(N^{-1/2}(A_N - \mathbb{E}(A_N))\big) = \mu_{\sqrt{f(1-f)}} \quad \text{weakly in probability,} \tag{5.103}$$

where the right-hand side is the probability measure obtained after replacing f with $\sqrt{f(1-f)}$ in [129, Theorem 2.1].

Proof of Theorem 5.1.3. The proof comes in three steps.

1 (Riemann approximation). For $N \geq 1$, define the $N \times N$ diagonal matrix Q_N by

$$Q_N(i,i) = F(i/N)Z_i, \qquad 1 \leq i \leq N, \tag{5.104}$$

where

$$F(x) = \left(\int_0^1 f(x,y)\,dy\right)^{1/2}, \qquad x \in [0,1], \tag{5.105}$$

and $(Z_i\colon\ i \geq 1)$ is as in Lemma 5.2.4. Lemma E.2 implies that

$$\left|\left(\frac{1}{N}\operatorname{Tr}\left((\tilde{\Delta}_N)^k\right)\right)^{1/k} - \left(\frac{1}{N}\operatorname{Tr}\left((\bar{A}_N + Q_N)^k\right)\right)^{1/k}\right| \leq \left(\frac{1}{N}\operatorname{Tr}\left((Y_N - Q_N)^k\right)\right)^{1/k}. \tag{5.106}$$

Since, f being continuous,

$$\begin{aligned}&\mathbb{E}\big[N^{-2}\operatorname{Tr}^2\left((Y_N - Q_N)^k\right)\big]\\ &= O(1)\sup_{x\in[0,1]}\left[F(x) - \left(\frac{1}{N}\sum_{j=1, j\neq[Nx]/N}^{N} f\left(x,\frac{j}{N}\right)\right)^{1/2}\right]^{2k}, \qquad N\to\infty,\end{aligned} \tag{5.107}$$

and it tends to 0 as $N \to \infty$, we get that, for every even k,

$$\left(\frac{1}{N}\operatorname{Tr}\left((\tilde{\Delta}_N)^k\right)\right)^{1/k} - \left(\frac{1}{N}\operatorname{Tr}\left((\bar{A}_N + Q_N)^k\right)\right)^{1/k} \tag{5.108}$$

tends to 0 in L^{2k} as $N \to \infty$.

Our next step is to show that, for every even integer k,

$$\lim_{N\to\infty}\frac{1}{N}\operatorname{Tr}\left((\bar{A}_N + Q_N)^k\right) = \gamma_k \quad \text{in } L^2 \tag{5.109}$$

for some $\gamma_k \in \mathbb{R}$. The above will follow once we show that, for all $m \geq 1$ and $n_1, \ldots, n_m \geq 0$,

$$\lim_{N\to\infty}\frac{1}{N}\operatorname{Tr}\left(Q_N^{n_1}\bar{A}_N\cdots Q_N^{n_m}\bar{A}_N\right) = \theta \quad \text{in } L^2 \tag{5.110}$$

for some $\theta \in \mathbb{R}$ (depending on $m, n_1, \ldots, n_m$). To that end, define the diagonal matrices U_N and B_N by

$$U_N(i,i) = Z_i \quad \text{and} \quad B_N(i,i) = F(i/N), \tag{5.111}$$

for $i = 1, \ldots, N$. Observe that

$$Q_N = B_N U_N = U_N B_N, \tag{5.112}$$

and hence the left-hand side of (5.110) is the same as

$$\frac{1}{N} \operatorname{Tr}\big(U_N^{n_1} W_{N,1} \cdots U_N^{n_m} W_{N,m}\big), \tag{5.113}$$

where

$$W_{N,j} = B_N^{n_j} \bar{A}_N, \qquad j = 1, \ldots, m. \tag{5.114}$$

In order to apply Lemma 5.2.6 we need to verify its hypotheses.

2 (Verification of the hypotheses). Our next claim is that $W_{N,1}, \ldots, W_{N,m}$ satisfy (5.80)–(5.83). To that end, observe that for $N \geq 1$ and $j = 1, \ldots, m$,

$$W_{N,j}(u,v) = F^{n_j}\left(\frac{u}{N}\right) f^{1/2}\left(\frac{u}{N}, \frac{v}{N}\right) N^{-1/2} G_{u\wedge v,\, u\vee v}, \qquad 1 \leq u, v \leq N. \tag{5.115}$$

Let

$$H_j(x,y) = F^{n_j}(x) f^{1/2}(x,y), \qquad (x,y) \in [0,1]^2. \tag{5.116}$$

Fix a partition Π of $\{1, \ldots, m\}$. Recall the notation $\Psi(\Pi, N)$ introduced in the proof of Lemma 5.2.4. Clearly, for every $i \in \Psi(\Pi, N)$,

$$\mathbb{E}\Big[\prod_{j=1}^{m} W_{N,j}(i_j, i_{j+1})\Big] = N^{-m/2} \psi(\Pi) \Big(\prod_{j=1}^{m} H_j\Big(\frac{i_j}{N}, \frac{i_{j+1}}{N}\Big)\Big), \tag{5.117}$$

where

$$\psi(\Pi) = \mathbb{E}\Big[\prod_{j=1}^{m} G_{i_j \wedge i_{j+1}, i_j \vee i_{j+1}}\Big], \tag{5.118}$$

which does not depend on $i \in \Psi(\Pi, N)$. The standard arguments leading to a proof via the method of moments of the Wigner semicircle law show that

$$\begin{aligned} &\lim_{N\to\infty} N^{-m/2+1}\, \psi(\Pi)\, |\Psi(\Pi, N)| \\ &\quad = \begin{cases} 1, & \text{if } m \text{ is even, and } \Pi = K(\sigma) \text{ for some } \sigma \in NC_2(m), \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{5.119}$$

Assume for the moment that m is even, and let $\sigma \in NC_2(m)$. It is known that $K(\sigma)$ has $m/2 + 1$ blocks. Define a function $\mathcal{L}_\sigma : \{1, \ldots, m\} \to \{1, \ldots, \frac{1}{2}m + 1\}$ such that

$\mathcal{L}_\sigma(j) = \mathcal{L}_\sigma(k)$ if and only if j, k are in the same block of $K(\sigma)$. It follows that for $\Pi = K(\sigma)$,

$$\begin{aligned}\lim_{N\to\infty} \frac{1}{N} \sum_{i\in\Psi(\Pi,N)} \mathbb{E}\Big[\prod_{j=1}^{m} W_{N,j}(i_j, i_{j+1})\Big] \\ = \int_{[0,1]^{(m/2)+1}} \prod_{(u,v)\in\sigma, u<v} H_u\big(x_{\mathcal{L}_\sigma(u)}, x_{\mathcal{L}_\sigma(v)}\big)\, dx_1 \cdots dx_{(m/2)+1}.\end{aligned} \tag{5.120}$$

This shows that hypothesis (5.82) holds. The hypotheses (5.81) and (5.83) follow similarly by an analogue of the standard arguments, while (5.80) is trivial.

Thus, $W_{N,1}, \ldots, W_{N,m}$ and U_N satisfy the hypotheses of Lemma 5.2.6. The claim of that lemma shows that the random variable in (5.113) converges in L^2 to a finite deterministic constant as $N \to \infty$, i.e., (5.110) holds. This in turn proves (5.109), which in conjunction with (5.108) shows that

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{Tr}\big((\tilde{\Delta}_N)^k\big) = \gamma_k \quad \text{in } L^2. \tag{5.121}$$

Lemma 5.2.4 asserts that

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{Tr}\big((\bar{\Delta}_N)^k\big) = \gamma_k \quad \text{in } L^2, \tag{5.122}$$

and hence also in probability.

3 (Uniqueness of the limiting measure). Equation (5.109) ensures that there exists a symmetric probability measure on $\mathbb{R}$ whose k-th moment is γ_k for every even integer k. Our next claim is that such a measure is unique, i.e., $(\gamma_k\colon\ k \geq 1)$ determines the measure. It is not obvious how to check Carleman's condition, and therefore we argue as follows. It suffices to exhibit a probability measure ν whose odd moments are zero and whose k-th moment is γ_k for even k such that

$$\int_{\mathbb{R}} e^{tx} \nu(dx) < \infty \qquad \forall\, t \in \mathbb{R}. \tag{5.123}$$

To do so we bring in the notion of a non-commutative probability space (NCP), which is defined in Appendix E. For $K > 0$ and $N \geq 1$, define

$$U_{NK} = \operatorname{Diag}\big(Z_1\mathbf{1}(|Z_1| \leq K), \ldots, Z_1\mathbf{1}(|Z_N| \leq K)\big), \tag{5.124}$$

and

$$Q_{NK} = B_N U_{NK}. \tag{5.125}$$

The arguments leading to (5.110) can be easily tweaked to show that, for fixed $K > 0$ and a fixed polynomial p in two non-commuting variables,

$$\lim_{N\to\infty} \frac{1}{N} \mathbb{E}\big[\operatorname{Tr}\big(p(\bar{A}_N, Q_{NK})\big)\big] \tag{5.126}$$

exists. Lemma E.4 implies that there exist self-adjoint elements q and a in a tracial NCP $(\mathcal{A}, \phi)$ such that the above limit equals $\phi\left[p\left(a, q\right)\right]$ for every polynomial p in two non-commuting variables. Hence

$$\lim_{N\to\infty} \mathrm{EESD}\big[p(\bar{A}_N, Q_{NK})\big] = \mathcal{L}(p\left(a, q\right)) \quad \text{in distibution,} \tag{5.127}$$

for any symmetric polynomial p, where EESD denotes the expectation of ESD. Theorem 5.1.2 implies that the limiting spectral distribution of $\bar{A}_N$, which is $\mathcal{L}(a)$ by (5.127), is compactly supported, and hence a is a bounded element. The spectrum of q is clearly a subset of $[-K, K]$. The second claim in Lemma E.4 allows us to assume that $(\mathcal{A}, \phi)$ is a W^*-probability space.

Let

$$\nu_K = \mathcal{L}(a + q). \tag{5.128}$$

If C is a finite constant such that

$$-C\mathbf{1} \le a \le C\mathbf{1}, \tag{5.129}$$

then clearly

$$a + q \le C\mathbf{1} + q. \tag{5.130}$$

Applying the method of moments to Q_{NK}, we find by an appeal to (5.127) that the law of q is the same as the law of

$$F(V)Z_1\mathbf{1}(|Z_1| \le K),$$

where V is standard uniform independently of Z_1, and F is as in (5.105). Under the assumption that $f \le 1$, which represents no loss of generality,

$$\int_{\mathbb{R}} e^{tx}\big(\mathcal{L}(q)\big)(dx) \le e^{t^2/2}, \qquad t \in \mathbb{R}. \tag{5.131}$$

By [118, Corollary 3.3] applied to (5.130), it follows that

$$\int_{\mathbb{R}} e^{tx}\nu_K(dx) \le \int_{\mathbb{R}} e^{tx}\big(\mathcal{L}(C\mathbf{1} + q)\big)(dx) \le \exp\big(\tfrac{1}{2}t^2 + tC\big), \qquad t > 0. \tag{5.132}$$

Lemma E.1 applied to $\bar{A}_N + Q_{NK_1}$ and $\bar{A}_N + Q_{NK_1}$ shows that

$$\sup_{N\ge 1} L\big(\mathrm{EESD}[\bar{A}_N + Q_{NK_1}], \mathrm{EESD}[\bar{A}_N + Q_{NK_2}]\big) \tag{5.133}$$

is small for large K_1 and K_2. Thus, $(\nu_K\colon\ K > 0)$ is Cauchy in the Lévy metric, and hence there exists a probability measure ν such that

$$\lim_{K\to\infty} \nu_K = \nu. \tag{5.134}$$

This, along with (5.132), establishes that

$$\int_{\mathbb{R}} e^{tx}\nu(dx) \le \exp\big(\tfrac{1}{2}t^2 + tC\big), \qquad t > 0, \tag{5.135}$$

and

$$\lim_{K\to\infty}\int_{\mathbb{R}} x^k \nu_K(dx) = \int_{\mathbb{R}} x^k \nu(dx), \qquad k \geq 1. \tag{5.136}$$

Clearly,

$$\int_{\mathbb{R}} x^k \nu_K(dx) = \lim_{N\to\infty} N^{-1}\mathbb{E}\big[\operatorname{Tr}\big((\bar{A}_N + Q_{NK})^k\big)\big]. \tag{5.137}$$

Therefore, by keeping track of the limit in (5.126), we can show (with some effort) that

$$\lim_{K\to\infty}\int_{\mathbb{R}} x^k \nu_K(dx) = \begin{cases}\gamma_k, & k \text{ even},\\ 0, & k \text{ odd}.\end{cases} \tag{5.138}$$

Thus, ν has the desired moments. By extending (5.135) to the case $t < 0$, we see that (5.123) follows. Thus, ν is the only symmetric probability measure whose even moments are (γ_k).

Equation (5.122) and the claim proved above show that

$$\lim_{N\to\infty} \operatorname{ESD}(\bar{\Delta}_N) = \nu \quad \text{ weakly in probability.} \tag{5.139}$$

Hence Lemmas 5.2.1–5.2.3 imply that

$$\lim_{N\to\infty} \operatorname{ESD}\big((N\varepsilon_N)^{-1/2}(\Delta_N - D_N)\big) = \nu \quad \text{ weakly in probability.} \tag{5.140}$$

as in the proof of Theorem 5.1.2.

It remains to show that if f is not identically zero, then the support of ν is unbounded. To that end, recall that (5.109), together with the fact that ν is the only symmetric probability measure whose even moments are (γ_k), establish that

$$\lim_{N\to\infty} \operatorname{ESD}(\bar{A}_N + Q_N) = \nu \quad \text{ weakly in probability,} \tag{5.141}$$

where $\bar{A}_N$ and Q_N are as in (5.55) and (5.104), respectively. Fix $0 < p < 1/2$, and for any $N \times N$ real symmetric matrix Σ, enumerate its eigenvalues in descending order by $\lambda_1(\Sigma), \ldots, \lambda_N(\Sigma)$. Weyl's inequality (see [189, Equation (1.54)]) implies that

$$\lambda_{2\lceil Np\rceil - 1}(Q_N) \leq \lambda_{\lceil Np\rceil}(\bar{A}_N + Q_N) + \lambda_{\lceil Np\rceil}(-\bar{A}_N), \tag{5.142}$$

where $\lceil x\rceil$ denotes the smallest integer larger than or equal to x. Therefore

$$\begin{aligned}\limsup_{N\to\infty} \lambda_{\lceil Np\rceil}(\bar{A}_N + Q_N) &\geq \limsup_{N\to\infty} \lambda_{2\lceil Np\rceil-1}(Q_N) - \liminf_{N\to\infty} \lambda_{\lceil Np\rceil}(-\bar{A}_N)\\ &\geq \limsup_{N\to\infty} \lambda_{2\lceil Np\rceil-1}(Q_N) - C,\end{aligned} \tag{5.143}$$

where C is as in (5.129). Letting $p \to 0$ and appealing to Lemma E.6, we find that

$$\begin{aligned}\sup(\operatorname{Supp}(\nu)) &= \lim_{p\to 0}\limsup_{N\to\infty} \lambda_{\lceil Np\rceil}(\bar{A}_N + Q_N)\\ &\geq \lim_{p\to 0}\limsup_{N\to\infty} \lambda_{2\lceil Np\rceil-1}(Q_N) - C = \infty,\end{aligned} \tag{5.144}$$

where the last line uses the fact that, as $N \to \infty$, $\operatorname{ESD}(Q_N)$ converges weakly in probability to the distribution of $F(V)Z_1$, the support of which is unbounded because f is not identically zero. □

§5.3.2 Proof: identification

Proof of Theorem 5.1.6. Let $(G_{i,j}\colon\ 1 \le i \le j)$ and $(Z_i\colon\ i \ge 1)$ be as in Lemma 5.2.4. For $N \ge 1$, define the $N \times N$ matrices

$$\begin{aligned} G_N(i,j) &= N^{-1/2} G_{i\wedge j, i\vee j}, \qquad 1 \le i,j \le N, && (5.145)\\ R_N &= \mathrm{Diag}\left(\sqrt{r(1/N)}, \ldots, \sqrt{r(1)}\right), && (5.146)\\ U_N &= \mathrm{Diag}(Z_1, \ldots, Z_N). && (5.147) \end{aligned}$$

The notation U_N is exactly as in the proof of Theorem 5.1.3. Let $\bar{A}_N$ and Q_N be as in (5.55) and (5.104), respectively. Observe that, under the assumption (5.16),

$$\bar{A}_N = R_N G_N R_N, \tag{5.148}$$

and

$$Q_N = \alpha R_N^{1/2} U_N R_N^{1/2}, \tag{5.149}$$

where α is as defined in the statement of Theorem 5.1.6. Proceeding as in the proofs of Theorems 5.1.2–5.1.3, wee see that it suffices to show that

$$\lim_{N\to\infty} \mathrm{ESD}\left(R_N G_N R_N\right) = \mathcal{L}\big(r^{1/2}(T_u) T_s T^{1/2}(T_u)\big) \quad \text{weakly in probability} \tag{5.150}$$

and

$$\begin{aligned} &\lim_{N\to\infty} \mathrm{ESD}\big(R_N G_N R_N + \alpha R_N^{1/2} U_N R_N^{1/2}\big)\\ &\quad = \mathcal{L}\big(r^{1/2}(T_u) T_s T^{1/2}(T_u) + \alpha r^{1/4}(T_u) T_g r^{1/4}(T_u)\big) \quad \text{weakly in probability,} \end{aligned} \tag{5.151}$$

where T_s, T_g, T_u are as in the statement. Define U_{NK} to be the "truncated" version of U_N, for a fixed $K > 0$, as in the proof of Theorem 5.1.3. Both (5.150) and (5.151) will follow once we show that

$$\lim_{N\to\infty} \frac{1}{N} \mathrm{Tr}\left(p\big(R_N^{1/2}, U_{NK}, G_N\big)\right) = \tau\big(p(T_r, T_g', T_s)\big) \quad \text{in probability,} \tag{5.152}$$

where $T_r = r^{1/4}(T_u)$ and $T_g' = T_g \mathbb{1}_{\{|T_g| \le K\}}$, for any symmetric polynomial p in three non-commuting variables. It is a well known fact that, for all $k \ge 1$,

$$\lim_{N\to\infty} \frac{1}{N} \mathrm{Tr}(G_N^k) = \tau(T_s^k) \quad \text{in probability.} \tag{5.153}$$

Since R_N and U_{NK} are diagonal matrices, they commute. This, in conjunction with the strong law of large numbers, implies that, for any $k \ge 1$, $m_1, \ldots, m_k$ and $n_1, \ldots, n_k \ge 0$,

$$\begin{aligned} &\lim_{N\to\infty} \frac{1}{N} \mathrm{Tr}\left(R_N^{m_1} U_{NK}^{n_1} \cdots R_N^{m_k} U_{NK}^{n_k}\right)\\ &\quad = \int_0^1 r^{(m_1+\ldots+m_k)/4}(u)\, du \int_{-K}^{K} (2\pi)^{-1/2} x^{n_1+\ldots+n_k} e^{-x^2/2}\, dx \quad \text{almost surely} \end{aligned} \tag{5.154}$$

The above, in conjunction with (5.26) and the fact that T_g and T_r commute, implies that

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{Tr}\big(p\big(R_N^{1/2}, U_{NK}\big)\big) = \tau\big(p(T_r, T_g')\big) \quad \text{almost surely} \tag{5.155}$$

for any polynomial p in two variables.

Thus, all that remains to show is the asymptotic free independence of T_s and (T_r, T_g'), which is precisely the claim of Lemma E.5, i.e., (5.153) and (5.155) imply (5.152). Applying (5.152) to $p(x,y,z) = x^2zx^2$ and $p(x,y,z) = x^2zx^2 + \alpha xyx$, we get the truncated versions of (5.150) and (5.151), respectively. Yet another application of Lemma E.1 allows us to let $K \to \infty$, obtaining (5.150) and (5.151). This completes the proof of (5.23) and (5.24). □

§5.3.3 Proof: randomization

Proof of Theorem 5.1.7. As before, Lemma 5.2.1 and (5.2) imply that the mean of the entries of A_N can be subtracted at the cost of a negligible perturbation of the ESD. The inequalities (5.2) and (5.27) ensure that the Gaussianization as in Lemma 5.2.2 goes through by conditioning on $R_{N1}, \dots, R_{NN}$. That is, if $(G_{ij}\colon 1 \le i \le j)$ is a collection of i.i.d. standard normal random variables that are independent of $(R_{Ni}\colon 1 \le i \le N, N \ge 1)$, W_N^g is an $N \times N$ matrix defined by

$$W_N^g(i,j) = G_{i\wedge j, i\vee j}, 1 \le i,j \le N, \tag{5.156}$$

and

$$\Theta_N = \operatorname{Diag}\big(\sqrt{R_{N1}}, \dots, \sqrt{R_{NN}}\big), \tag{5.157}$$

then the ESD of $A_N/\sqrt{N\varepsilon_N}$ is close to that of $\Theta_N W_N^g \Theta_N/\sqrt{N}$.

The assumptions (5.27) and (5.28) imply that, for $k \ge 1$,

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{Tr}\big(\Theta_N^{2k}\big) = \int_{\mathbb{R}} x^k \mu_r(dx) \quad \text{almost surely.} \tag{5.158}$$

Finally, Lemma E.5 together with (5.27) shows the asymptotic free independence of Θ_N and W_N^g, that is,

$$\lim_{N\to\infty} \operatorname{ESD}\big(N^{-1/2}\Theta_N W_n^g \Theta_N\big) = \mu_r \boxtimes \mu_s \quad \text{weakly in probability.} \tag{5.159}$$

This completes the proof. □

§5.4 Applications

In this section we discuss three applications, explained in Sections 5.4.1–5.4.3.

§5.4.1 Constrained random graphs

Let $\mathcal{S}_N$ be the set of all simple graphs on N vertices. Suppose that we fix the degrees of the vertices, namely, vertex i has degree k_i^*. Here, $k^* = (k_i^*\colon 1 \le i \le N)$ is a

sequence of positive integers of which we only require that they are graphical, i.e., there is at least one simple graph with these degrees. The so-called *canonical ensemble* P_N is the unique probability distribution on S_N with the following two properties.

(I) The *average degree* of vertex i, defined by $\sum_{G\in\mathcal{S}_N} k_i(G)P_N(G)$, equals k_i^* for all $1 \le i \le N$.

(II) The *entropy* of P_N, defined by $-\sum_{G\in\mathcal{S}_N} P_N(G)\log P_N(G)$, is maximal.

The name canonical ensemble comes from Gibbs theory in equilibrium statistical physics. The probability distribution P_N describes a random graph of which we have no prior information other than the average degrees, and is called the *soft configuration model*. It is known that, because of property (II), P_N takes the form (see [169])

$$P_N(G) = \frac{1}{Z_N(\theta^*)} \exp\Big[-\sum_{i=1}^{N} \theta_i^* k_i(G)\Big], \qquad G \in \mathcal{S}_N, \tag{5.160}$$

where $\theta^* = (\theta_i^*\colon 1 \le i \le N)$ is a sequence of real-valued Lagrange multipliers that must be chosen in such a way that property (I) is satisfied. The normalization constant $Z_N(\theta^*)$, which depends on θ^*, is called the partition function in Gibbs theory.

The gradients of the constraints in property (I) are linearly independent vectors and the matching of property (I) uniquely fixes θ^*. It turns out that

$$P_N(G) = \prod_{1\le i<j\le N} (p_{ij}^*)^{A_N[G](i,j)} (1-p_{ij}^*)^{1-A_N[G](i,j)}, \qquad G \in \mathcal{S}_N, \tag{5.161}$$

where $A_N[G]$ is the adjacency matrix of G, and p_{ij}^* represent a reparameterisation of the Lagrange multipliers, namely,

$$p_{ij}^* = \frac{x_i^* x_j^*}{1 + x_i^* x_j^*}, \qquad 1 \le i \ne j \le N, \tag{5.162}$$

with $x_i^* = e^{-\theta_i^*}$ (see [187] for more details). Thus, we see that P_N is nothing other than an inhomogeneous Erdős-Rényi random graph where the probability that vertices i and j are connected by an edge equals p_{ij}^*. In order to match property (I), these probabilities must satisfy

$$k_i^* = \sum_{j=1, j\ne i}^{N} p_{ij}^*, \qquad 1 \le i \le N, \tag{5.163}$$

which constitutes a set of N equations for the N unknowns $x_1^*, \dots, x_N^*$.

In order to state the next result, we need to make some assumptions on the sequence $(k_{Ni}^*\colon 1 \le i \le N)$. For the sake of notational simplification, the dependence on N will be suppressed from the notation.

Proposition 5.4.1 (Theorem 5.1.7 for constrained random graphs). *Let $(k_i^*\colon 1 \le i \le N)$ be a graphical sequence of positive integers. Define*

$$m_N = \max_{1\le \ell \le N} k_\ell^*. \tag{5.164}$$

Assume that

$$\lim_{N\to\infty} m_N = \infty, \qquad \lim_{N\to\infty} m_N/\sqrt{N} = 0, \tag{5.165}$$

and

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \delta_{k_i^*/m_N} = \mu_r \quad \textit{weakly}, \tag{5.166}$$

for some probability measure μ_r. Let x_i^* and p_{ij}^* be determined by (5.162) *and* (5.163). *Let A_N be the adjacency matrix of an inhomogeneous Erdős-Rényi random graph on N vertices, with p_{ij}^* the probability of an edge being present between vertices i and j for $1 \le i \ne j \le N$. Then*

$$\lim_{N\to\infty} \mathrm{ESD}\big((N\varepsilon_N)^{-1/2} A_N\big) = \mu_r \boxtimes \mu_s \quad \textit{weakly in probability}. \tag{5.167}$$

Proof. Abbreviate

$$\sigma_N = \sum_{\ell=1}^{N} k_\ell^*. \tag{5.168}$$

In [187] it is shown that

$$\max_{1\le \ell\le N} x_\ell^* = o(1), \qquad N\to\infty, \tag{5.169}$$

in which case (5.162) and (5.163) give

$$x_i^* = \frac{k_i^*}{\sqrt{\sigma_N}}\,[1+o(1)] \quad \text{and} \quad p_{ij}^* = \frac{k_i^* k_j^*}{\sqrt{\sigma_N}}\,[1+o(1)], \qquad N\to\infty, \tag{5.170}$$

with the error term uniform in $1 \le i \ne j \le N$. Pick

$$\varepsilon_N = \frac{m_N^2}{\sigma_N}. \tag{5.171}$$

It follows from (5.165) that

$$\lim_{N\to\infty} \varepsilon_N = 0, \qquad \lim_{N\to\infty} N\varepsilon_N = \infty. \tag{5.172}$$

As in the proof of Theorem 5.1.7, Lemmas 5.2.1–5.2.2 imply that the upper triangular entries of A_N can be replaced by independent mean-zero normal random variables. In other words, if $(G_{ij}\colon\ 1 \le i \le j)$ are i.i.d. standard normal, and A_N^g is the random matrix defined by

$$A_N^g(i,j) = \sqrt{p_{ij}^*}\, G_{i\wedge j, i\vee j}, \qquad 1 \le i,j \le N, \tag{5.173}$$

with $p_{ii}^* = 0$ for all $1 \le i \le N$, then $\mathrm{ESD}\big((N\varepsilon_N)^{-1/2}A_N\big)$ and $\mathrm{ESD}\big((N\varepsilon_N)^{-1/2}A_N^g\big)$ are asymptotically close. The second part of (5.170) implies that

$$\sqrt{p_{ij}^*} = \sqrt{\varepsilon_N \frac{k_i^* k_j^*}{m_N^2}}\,[1+o(1)], \qquad N\to\infty, \tag{5.174}$$

uniformly in $1 \le i \ne j \le N$, and hence

$$\sum_{i,j=1}^{N} \left(\sqrt{p_{ij}^*} - \sqrt{\varepsilon_N \, \frac{k_i^* k_j^*}{m_N^2}} \right)^2 = o(N^2 \varepsilon_N), \qquad N \to \infty. \tag{5.175}$$

In other words, if $\tilde{A}_N$ is defined by

$$\tilde{A}_N(i,j) = \sqrt{\frac{k_i^* k_j^*}{m_N^2}} \, G_{i \wedge j, i \vee j}, \qquad 1 \le i, j \le N, \tag{5.176}$$

then

$$\lim_{N \to \infty} \frac{1}{N} \mathbb{E}\big[\operatorname{Tr} \big((N\varepsilon_N)^{-1/2} A_N^g - N^{-1/2} \tilde{A}_N \big)^2 \big] = 0. \tag{5.177}$$

Lemma E.1 implies that

$$\lim_{N \to \infty} L\big(\operatorname{ESD}\big((N\varepsilon_N)^{-1/2} A_N^g \big), \operatorname{ESD}\big(N^{-1/2} \tilde{A}_N \big) \big) = 0 \quad \text{in probability.} \tag{5.178}$$

Finally, by an appeal to Lemma E.5, (5.166) implies that

$$\lim_{N \to \infty} \operatorname{ESD}\big(N^{-1/2} \tilde{A}_N \big) = \mu_r \boxtimes \mu_s \quad \text{weakly in probability,} \tag{5.179}$$

where μ_s is the standard semicircle law. Hence

$$\lim_{N \to \infty} \operatorname{ESD}\big((N\varepsilon_N)^{-1/2} A_N \big) = \mu_r \boxtimes \mu_s \quad \text{weakly in probability,} \tag{5.180}$$

and this completes the proof. □

Remark 5.4.2 (Example).
We look at a concrete example of a graphical sequence $(k_i^* \colon 1 \le i \le N)$ satisfying (5.165)–(5.166). For $N \ge 1$, let

$$k_i^* = \lfloor i^{1/3} \rfloor, \qquad 1 \le i \le N, \tag{5.181}$$

where $\lfloor x \rfloor$ denotes the greatest integer smaller than or equal to x. Then [165, Theorem 7.12] implies that $(k_i^* \colon 1 \le i \le N)$ is graphical for N large enough. Since $m_N = \lfloor N^{1/3} \rfloor$, it is immediate that (5.165) holds and that

$$\lim_{N \to \infty} \left(\frac{1}{N} \sum_{i=1}^{N} \delta_{k_i^*/m_N} \right)(\cdot) = \mathbb{P}\big(U^{1/3} \in \cdot \big) \quad \text{weakly,} \tag{5.182}$$

with U a standard uniform random variable.

§5.4.2 Chung-Lu graphs

The following random graph introduced by [140] is similar to the one discussed in Section 5.4.1. For $N \ge 1$, let $(d_{Ni} \colon 1 \le i \le N)$ be a sequence of positive real numbers. Abbreviate

$$m_N = \max_{1 \le i \le N} d_{Ni}, \qquad \sigma_N = \sum_{i=1}^{N} d_{Ni}. \tag{5.183}$$

Assume that

$$\lim_{N\to\infty} \frac{m_N^2}{\sigma_N} = 0, \qquad \lim_{N\to\infty} N\frac{m_N^2}{\sigma_N} = \infty, \tag{5.184}$$

and

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \delta_{d_{Ni}/m_N} = \mu_r \quad \text{weakly} \tag{5.185}$$

for some measure μ_r on $\mathbb{R}$. Consider an inhomogeneous Erdős-Rényi graph on N vertices where an edge exists between i and j, $1 \le i \ne j \le N$, with probability $d_{Ni}d_{Nj}/\sigma_N$. Such graph is called a Chung-Lu graph. If A_N denotes its adjacency matrix, then the following result follows from Theorem 5.1.7.

Proposition 5.4.3 (Theorem 5.1.7 for Chung-Lu graphs).
Under the hypotheses mentioned above,

$$\lim_{N\to\infty} \mathrm{ESD}\big((N\varepsilon_N)^{-1/2}A_N\big) = \mu_r \boxtimes \mu_s \quad \textit{weakly in probability,} \tag{5.186}$$

where

$$\varepsilon_N = \frac{m_N^2}{\sigma_N} \tag{5.187}$$

and μ_s is the standard semicircle law.

§5.4.3 Social networks

Consider a community consisting of N individuals. Data is available on whether the i-th individual and the j-th individual are acquainted, for every pair (i, j) with $1 \le i, j \le N$. Based on this data, the *sociability pattern* of the community has to be inferred statistically. Examples arise in social networks and collaboration networks.

The above situation can be modeled in several ways, one being the following. Denote by ρ the *sociability distribution* of the community, which is a compactly supported probability measure on $[0, \infty)$. Let $(R_i)_{1\le i\le N}$ be i.i.d. random variables drawn from ρ. Think of R_i as the *sociability index* of the i-th individual. Fix $\varepsilon_N > 0$ such that $\varepsilon_N m^2 \le 1$, where m is the supremum of the support of ρ, so that

$$0 \le \varepsilon_N R_i R_j \le 1, \qquad 1 \le i \ne j \le N. \tag{5.188}$$

Suppose that, conditional on $(R_i)_{1\le i\le N}$, the i-th and the j-th individual are acquainted with probability $\varepsilon_N R_i R_j$. In other words, the graph in which the vertices are individuals and the edges are mutual acquaintances is an inhomogeneous Erdős-Rényi random graph G_N with random connection parameters that are controlled by ρ. The data that is available is the adjacency matrix A_N of this graph. The goal is to draw information about ρ from this data. This statistical inference problem boils down to estimating ρ from A_N. Without loss of generality we assume that ρ is standardized, i.e.,

$$\int_0^\infty x\rho(dx) = 1. \tag{5.189}$$

Proposition 5.4.4 (Theorem 5.1.7 for social networks).
Under the assumptions $N^{-1} \ll \varepsilon_N \ll 1$ *and* (5.189),

$$\lim_{N\to\infty} \mathrm{ESD}\left(\sqrt{\frac{N}{\mathrm{Tr}(A_N^2)}}\, A_N\right) = \rho \boxtimes \mu_s \qquad \textit{weakly in probability,} \tag{5.190}$$

where μ_s *is the standard semicircle law.*

Proof. It is immediate that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \delta_{R_i} = \rho \quad \text{weakly almost surely.} \tag{5.191}$$

Theorem 5.1.7 implies that if $N^{-1} \ll \varepsilon_N \ll 1$, then

$$\lim_{N\to\infty} \mathrm{ESD}\big((N\varepsilon_N)^{-1/2} A_N\big) = \rho \boxtimes \mu_s \qquad \text{weakly in probability.} \tag{5.192}$$

Since $A_N(i,j)$ is either 0 or 1,

$$\mathbb{E}\big[\,\mathrm{Tr}(A_N^2)\big] = \sum_{i,j=1}^{N} \mathbb{E}[A_N(i,j)] = \sum_{1\le i\ne j\le N} \varepsilon_N \mathbb{E}[R_i R_j] = \varepsilon_N N(N-1), \tag{5.193}$$

where the last equality follows from (5.189). Consequently,

$$\lim_{N\to\infty} \frac{1}{N^2\varepsilon_N} \mathbb{E}\big[\,\mathrm{Tr}(A_N^2)\big] = 1. \tag{5.194}$$

The fact that the variance equals the sum of the expectation of the conditional variance and the variance of the conditional expectation, implies that

$$\begin{aligned}
\mathrm{Var}\big(\,\mathrm{Tr}(A_N^2)\big) &= \mathrm{Var}\left(2 \sum_{1\le i<j\le N} A_N(i,j)\right) \\
&= 4E\left(\sum_{1\le i<j\le N} \varepsilon_N R_i R_j (1 - \varepsilon_N R_i R_j)\right) + 4\mathrm{Var}\left(\sum_{1\le i<j\le N} \varepsilon_N R_i R_j\right) \\
&= O(N^2\varepsilon_N) + 4\varepsilon_N^2 \sum_{1\le i<j\le N} \sum_{1\le k<l\le N} \mathrm{Cov}(R_i R_j, R_k R_l) \\
&= O(N^3\varepsilon_N^2), \qquad N\to\infty,
\end{aligned} \tag{5.195}$$

where the last line follows from the observation that if i,j,k,l are distinct, then $\mathrm{Cov}(R_iR_j, R_kR_l)$ vanishes. Hence,

$$\lim_{N\to\infty} \mathrm{Var}\left(\frac{1}{N^2\varepsilon_N}\,\mathrm{Tr}(A_N^2)\right) = 0. \tag{5.196}$$

The above in conjunction with (5.194) shows that

$$\lim_{N\to\infty} \frac{1}{N^2\varepsilon_N}\,\mathrm{Tr}(A_N^2) = 1 \quad \text{in probability.} \tag{5.197}$$

This, together with (5.192), completes the proof. □

Thus, $\rho \boxtimes \mu_s$ can in principle be statistically estimated from A_N. Subsequently, ρ can be computed because the moments of $\rho \boxtimes \mu_s$ are functions of the moments of ρ, as shown below. We know from [183, Equation (14.5)] that, for $n \geq 1$,

$$\int_{\mathbb{R}} x^{2n} \, \rho \boxtimes \mu_s(dx) = \sum_{\sigma \in NC_2(2n)} \prod_{j=1}^{n+1} \int_{\mathbb{R}} x^{l_j(\sigma)} \rho(dx), \tag{5.198}$$

where $l_1(\sigma), \ldots, l_{n+1}(\sigma)$ are block sizes of $K(\sigma)$, the Kreweras complement of σ. With the help of the above, the n-th moment of ρ can be written in terms of the $2n$-th moment of $\rho \boxtimes \mu_s$, and the first $n-1$ moments of ρ. Therefore, the moments of ρ can be recursively computed from those of $\rho \boxtimes \mu_s$. Since ρ is compactly supported, it can be computed from its moments.

§E Appendix: basic facts

The following is [114, Corollary A.41], and is also a corollary of the Hoffman-Wielandt inequality.

Lemma E.1 (Lévy distance between empirical spectral distributions).
If L denotes the Lévy distance between two probability measures, then for $N \times N$ symmetric matrices A and B,

$$L^3\big(\mathrm{ESD}(A), \mathrm{ESD}(B)\big) \leq \frac{1}{N} \operatorname{Tr}\big((A-B)^2\big). \tag{5.199}$$

The following is a consequence of the Minkowski and k-Hoffman-Wielandt inequalities. The latter can be found in Exercise 1.3.6 of [189].

Lemma E.2 (Difference of traces).
For real symmetric matrices A and B of the same order, and an even positive integer k,

$$\big|\operatorname{Tr}^{1/k}(A^k) - \operatorname{Tr}^{1/k}(B^k)\big| \leq \operatorname{Tr}^{1/k}\big((A-B)^k\big). \tag{5.200}$$

Definition E.3 (Non-commutative probability space).
A *non-commutative probability space (NCP)* $(\mathcal{A}, \phi)$ is a unital $*$-algebra $\mathcal{A}$ equipped with a linear functional $\phi\colon \mathcal{A} \to \mathbb{C}$ that is unital, i.e.,

$$\phi(\mathbf{1}) = 1, \tag{5.201}$$

and positive, i.e.,

$$\phi(a^* a) \geq 0 \qquad \forall\, a \in \mathcal{A}. \tag{5.202}$$

An NCP $(\mathcal{A}, \phi)$ is *tracial* if

$$\phi(ab) = \phi(ba), \qquad a, b \in \mathcal{A}. \tag{5.203}$$

Lemma E.4 (Limit of polynomials in an NCP).
Suppose that, for every $n \in \mathbb{N}$, $(\mathcal{A}_n, \phi_n)$ is a tracial NCP, and there exist self-adjoint $a_{n1}, \ldots, a_{nk} \in \mathcal{A}_n$ such that, for every polynomial p in k non-commuting variables,

$$\lim_{n \to \infty} \phi_n\big(p(a_{n1}, \ldots, a_{nk})\big) = \alpha_p \in \mathbb{C}. \tag{5.204}$$

Then there exists a tracial NCP $(\mathcal{A}_\infty, \phi_\infty)$ and self-adjoint $a_{\infty 1}, \ldots, a_{\infty k} \in \mathcal{A}_\infty$ such that, for every polynomial p in k non-commuting variables,

$$\phi_\infty\big(p(a_{\infty 1}, \ldots, a_{\infty k})\big) = \alpha_p. \tag{5.205}$$

Furthermore, if

$$\sup_{1 \le i \le k,\, j \ge 1} \left(\phi_\infty\big(a_{\infty i}^{2j}\big)\right)^{1/2j} < \infty, \tag{5.206}$$

then $(\mathcal{A}_\infty, \phi_\infty)$ can be embedded into a W^*-probability space.

Proof. Let

$$\mathcal{A}_\infty = \mathbb{C}\langle X_1, \ldots, X_k \rangle, \tag{5.207}$$

the set of all polynomials in k non-commuting variables. For a monomial

$$p = \alpha X_{i_1} \ldots X_{i_m}, \tag{5.208}$$

define

$$p^* = \overline{\alpha} X_{i_m} \ldots X_{i_1}. \tag{5.209}$$

This defines the $*$-operation on the whole of $\mathcal{A}$. Let

$$\phi_\infty(p) = \alpha_p \qquad \forall\, p \in \mathcal{A}_\infty. \tag{5.210}$$

It is immediate from (5.204) that ϕ_∞ is positive and unital, i.e., $(\mathcal{A}_\infty, \phi_\infty)$ is an NCP. The desired conclusions are ensured by defining

$$a_{\infty 1} = X_1, \ \ldots, \ a_{\infty k} = X_k. \tag{5.211}$$

Finally, (5.206) implies that $a_{\infty,1}, \ldots, a_{\infty,k}$ are bounded. Hence, by going from polynomials to continuous functions with the help of the Bolzano-Weierstrass theorem, we can embed $(\mathcal{A}_\infty, \phi_\infty)$ into a W^*-probability space. □

The next lemma follows from [182, Theorem 4.20] (which is due to Voiculescu) and the discussion immediately following it.

Lemma E.5 (Polynomials and independence in an NCP).
Suppose that W_N is an $N \times N$ scaled standard Gaussian Wigner matrix, i.e., a symmetric matrix whose upper triangular entries are i.i.d. normal with mean zero and variance $1/N$. Let D_N^1 and D_N^2 be (possibly random) $N \times N$ symmetric matrices such that there exists a deterministic C satisfying

$$\sup_{N \ge 1, i = 1,2} \|D_N^i\| \le C < \infty, \tag{5.212}$$

where $\|\cdot\|$ denotes the usual matrix norm (which for a symmetric matrix is the same as the largest absolute value of its eigenvalues). Furthermore, assume that there is a W^*-probability space $(\mathcal{A}, \tau)$ in which there are self-adjoint elements d_1 and d_2 such that, for any polynomial p in two variables, it

$$\lim_{N \to \infty} \frac{1}{N} \operatorname{Tr}\big(p(D_N^1, D_N^2)\big) = \tau\big(p(d_1, d_2)\big) \quad \textit{almost surely.} \tag{5.213}$$

Finally, suppose that (D_N^1, D_N^2) is independent of W_N. Then there exists a self-adjoint element s in $\mathcal{A}$ (possibly after expansion) that has the standard semicircle distribution and is freely independent of (d_1, d_2), and is such that

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{Tr}\big(p(W_N, D_N^1, D_N^2)\big) = \tau\big(p(s, d_1, d_2)\big) \quad \text{almost surely} \tag{5.214}$$

for any polynomial p in three variables.

Lemma E.6 (Support of the limiting measure of random variables). *Suppose that for all $n \geq 1$, $Z_{n1} \geq \ldots \geq Z_{nn}$ are random variables such that*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} \delta_{Z_{nj}} = \mu \quad \text{weakly in probability,} \tag{5.215}$$

for some probability measure μ on $\mathbb{R}$, where δ_x is the probability measure that puts mass 1 at x. Then,

$$\lim_{p\to 0} \limsup_{n\to\infty} Z_{n\,\lceil np\rceil} = \sup(\operatorname{Supp}(\mu)) \quad \text{almost surely,} \tag{5.216}$$

where $\lceil x\rceil$ denotes the smallest integer larger than or equal to x.

Proof. Our first claim is that if $x \in \mathbb{R}$ and $0 < p < 1$ are such that

$$\mu((-\infty, x)) < 1 - p, \tag{5.217}$$

then

$$\limsup_{n\to\infty} Z_{n\,\lceil np\rceil} \geq x \quad \text{almost surely.} \tag{5.218}$$

To see why, fix p, x as above and $\varepsilon > 0$ such that $\mu(\{x - \varepsilon\}) = 0$, and note that the hypothesis implies that

$$\lim_{n\to\infty} \frac{1}{n} \big|\{1 \leq j \leq n : Z_{nj} \leq x - \varepsilon\}\big| = \mu((-\infty, x - \varepsilon]) \quad \text{in probability.} \tag{5.219}$$

Therefore,

$$\begin{aligned} &P\left(\limsup_{n\to\infty} Z_{n\,\lceil np\rceil} \leq x - \varepsilon\right) \\ &\leq P\left(\frac{1}{n}\big|\{1 \leq j \leq n : Z_{nj} \leq x - \varepsilon\}\big| \geq 1 - \frac{1}{n}\lceil np\rceil \text{ for large } n\right) \\ &\leq \limsup_{n\to\infty} P\left(\frac{1}{n}\big|\{1 \leq j \leq n : Z_{nj} \leq x - \varepsilon\}\big| \geq 1 - \frac{1}{n}\lceil np\rceil\right) = 0, \end{aligned} \tag{5.220}$$

where the last step follows from (5.219) and the observation that

$$\lim_{n\to\infty} 1 - \frac{1}{n}\lceil np\rceil = 1 - p > \mu((-\infty, x)) \geq \mu((-\infty, x - \varepsilon]). \tag{5.221}$$

Since $\varepsilon > 0$ can be chosen to be arbitrarily small such that $\mu(\{x - \varepsilon\}) = 0$, (5.218) follows.

It is immediate to see that $\limsup_{n\to\infty} Z_{n\,\lceil np\rceil}$ is monotone in p, and hence the almost sure limit exists as $p \to 0$. Furthermore,

$$\lim_{p\to 0} \limsup_{n\to\infty} Z_{n\,\lceil np\rceil} \le \alpha \quad \text{almost surely,} \tag{5.222}$$

where

$$\alpha = \sup(\mathrm{Supp}(\mu)). \tag{5.223}$$

To complete the proof, choose x_k such that $x_k \to \alpha$ and $x_k < \alpha$. Since α is the right end point of the support of μ, it follows that

$$\mu((-\infty, x_k)) < 1. \tag{5.224}$$

Choosing

$$0 < p_k < [1 - \mu((-\infty, x_k))] \wedge \frac{1}{k}, \qquad k \ge 1, \tag{5.225}$$

we see that (5.218) implies

$$\limsup_{n\to\infty} Z_{n\,\lceil np_k\rceil} \ge x_k \quad \text{almost surely.} \tag{5.226}$$

Therefore, since $x_k \to \alpha$,

$$\liminf_{k\to\infty} \limsup_{n\to\infty} Z_{n\,\lceil np_k\rceil} \ge \alpha \quad \text{almost surely.} \tag{5.227}$$

Since $p_k \to 0$, the left-hand side above equals that of (5.222). □

CHAPTER 6

Largest eigenvalue of the adjacency matrix

This chapter is based on:
A. Chakrabarty, R.S. Hazra, F. den Hollander, M. Sfragara. *Large deviation principle for the maximal eigenvalue of inhomogeneous Erdős-Rényi random graphs.* [arXiv:2008.08367], 2020.

Abstract

We consider inhomogeneous Erdős-Rényi random graphs G_N on N vertices in the dense regime. The edge between the pair of vertices $\{i, j\}$ is retained with probability $r(\frac{i}{N}, \frac{j}{N})$, $1 \leq i \neq j \leq N$, independently of other edges, where $r\colon [0,1]^2 \to (0,1)$ is a symmetric function that plays the role of a reference graphon. Let λ_N be the largest eigenvalue of the adjacency matrix of G_N. It is known that λ_N/N satisfies a large deviation principle as $N \to \infty$. The associated rate function ψ_r is given by a variational formula that involves the rate function I_r of a large deviation principle on graphon space. We analyze this variational formula in order to identify the properties of ψ_r, specially when the reference graphon is of rank 1.

§6.1 Introduction and main results

In Section 6.1.1 we define the mathematical model and we state the large deviation principle (LDP) for inhomogeneous Erdős Rényi random graphs. In Section 6.1.2 we present some facts about graphon operators. In Section 6.1.3 we state the LDP for the largest eigenvalue of the adjacency matrix, together with some properties of the rate function. Moreover, under the assumption that the connection probabilities have a multiplicative structure, we identify the scaling behavior of the rate function around its minimum and its end points. In Section 6.1.4 we briefly discuss these results and give an outline of the remainder of the chapter.

§6.1.1 Setting

We refer to Section 1.2.3 for a general introduction to spectra of Erdős-Rényi random graphs. We focus on *inhomogeneous Erdős-Rényi random graphs* and consider the dense regime, where the degrees of the vertices diverge linearly with the size of the graph.

Recall Section 1.2.5 for an introduction to graphon theory. Let $r \in \mathcal{W}$ be a *reference graphon* satisfying

$$\exists\, \eta > 0\colon \qquad \eta \leq r(x,y) \leq 1-\eta \qquad \forall\, (x,y) \in [0,1]^2. \tag{6.1}$$

Fix $N \in \mathbb{N}$ and consider the random graph G_N with vertex set $[N] = \{1,\dots,N\}$ where the pair of vertices $i,j \in [N]$, $i \neq j$, is connected by an edge with probability $r(\frac{i}{N}, \frac{j}{N})$, independently of other pairs of vertices. Write $\mathbb{P}_N$ to denote the law of G_N. Use the same symbol for the law on $\mathcal{W}$ induced by the map that associates with the graph G_N its graphon h^{G_N}, defined by

$$h^{G_N}(x,y) = \begin{cases} 1, & \text{if there is an edge between vertex } \lceil Nx \rceil \text{ and vertex } \lceil Ny \rceil, \\ 0, & \text{otherwise.} \end{cases} \tag{6.2}$$

Recall the equivalence relation $\sim$ on $\mathcal{W}$ defined in Section 1.2.5 and write $\widetilde{\mathbb{P}}_N$ to denote the law of $\widetilde{h}^{G_N}$.

The following LDP is proved in [145] and is an extension of the celebrated LDP for homogeneous Erdős-Rényi random graphs derived in [138]. Further properties of the rate function were derived in [179].

Theorem 6.1.1 (LDP for inhomogeneous Erdős-Rényi random graphs). *Subject to* (6.1)*, the sequence* $(\widetilde{\mathbb{P}}_N)_{N\in\mathbb{N}}$ *satisfies the LDP on* $(\widetilde{\mathcal{W}}, \delta_\square)$ *with rate* $\binom{N}{2}$*, i.e.,*

$$\begin{aligned} \limsup_{N\to\infty} \frac{1}{\binom{N}{2}} \log \widetilde{\mathbb{P}}_N(\mathcal{C}) &\leq -\inf_{\widetilde{h}\in\mathcal{C}} J_r(\widetilde{h}) \qquad \forall\, \mathcal{C} \subset \widetilde{\mathcal{W}} \text{ closed}, \\ \liminf_{N\to\infty} \frac{1}{\binom{N}{2}} \log \widetilde{\mathbb{P}}_N(\mathcal{O}) &\geq -\inf_{\widetilde{h}\in\mathcal{O}} J_r(\widetilde{h}) \qquad \forall\, \mathcal{O} \subset \widetilde{\mathcal{W}} \text{ open}, \end{aligned} \tag{6.3}$$

where the rate function $J_r\colon \widetilde{\mathcal{W}} \to \mathbb{R}$ *is given by*

$$J_r(\widetilde{h}) = \inf_{\phi\in\mathcal{M}} I_r(h^\phi), \tag{6.4}$$

where h is any representative of $\widetilde{h}$ and

$$I_r(h) = \int_{[0,1]^2} \mathcal{R}\big(h(x,y) \mid r(x,y)\big)\, dx\, dy, \qquad h \in \mathcal{W}, \tag{6.5}$$

with

$$\mathcal{R}\big(a \mid b\big) = a \log \tfrac{a}{b} + (1-a) \log \tfrac{1-a}{1-b} \tag{6.6}$$

the relative entropy of two Bernoulli distributions with success probabilities $a \in [0,1]$, $b \in (0,1)$ (with the convention $0 \log 0 = 0$).

It is clear that J_r is a good rate function, i.e., $J_r \not\equiv \infty$ and J_r has compact level sets. Note that (6.4) differs from the expression in [145], where the rate function is the lower semi-continuous envelope of $I_r(h)$. However, it was shown in [180] that, under the integrability conditions $\log r, \log(1-r) \in L^1([0,1]^2)$, the two rate functions are equivalent, since $J_r(\widetilde{h})$ is lower semi-continuous on $\widetilde{\mathcal{W}}$. Clearly, these integrability conditions are implied by (6.1).

§6.1.2 Graphon operators

With $h \in \mathcal{W}$ we associate a *graphon operator* acting on $L^2([0,1])$, defined as the linear integral operator

$$(T_h u)(x) = \int_{[0,1]} h(x,y) u(y)\, dy, \qquad x \in [0,1], \tag{6.7}$$

with $u \in L^2([0,1])$. The operator norm of T_h is defined as

$$\|T_h\| = \sup_{\substack{u \in L^2([0,1]) \\ \|u\|_2 = 1}} \|T_h u\|_2, \tag{6.8}$$

where $\|\cdot\|_2$ denotes the L^2-norm. Given a graphon $h \in \mathcal{W}$, we have that $\|T_h\| \leq \|h\|_2$. Hence, a graphon sequence converging in the L^2-norm also converges in the operator norm.

The product of two graphons $h_1, h_2 \in \mathcal{W}$ is defined as

$$(h_1 h_2)(x,y) = \int_{[0,1]} h_1(x,z) h_2(z,y)\, dz, \qquad (x,y) \in [0,1]^2, \tag{6.9}$$

and the n-th power of a graphon $h \in \mathcal{W}$ as

$$h^n(x,y) = \int_{[0,1]^{n-1}} h(x,z_1) \cdots h(z_{n-1},y)\, dz_1 \cdots dz_{n-1}, \qquad (x,y) \in [0,1]^2,\ n \in \mathbb{N}. \tag{6.10}$$

Definition 6.1.2 (Eigenvalues and eigenfunctions).
The number $\mu \in \mathbb{R}$ is said to be an *eigenvalue* of the graphon operator T_h if there exists a non-zero function $u \in L^2([0,1])$ such that

$$(T_h u)(x) = \mu u(x), \qquad x \in [0,1]. \tag{6.11}$$

The function u is said to be an *eigenfunction* associated with μ.

Proposition 6.1.3 (Properties of the graphon operator).
For any $h \in \mathcal{W}$ the following statements hold.

(i) The graphon operator T_h is self-adjoint, bounded and continuous.

(ii) The graphon operator T_h is diagonalisable and has countably many eigenvalues, all of which are real and can be ordered as $\mu_1 \geq \mu_2 \geq \ldots \geq 0$. Moreover, there exists a collection of eigenfunctions which form an orthonormal basis of $L^2([0,1])$.

(iii) The largest eigenvalue μ_1 of the graphon operator T_h is strictly positive and has an associated eigenfunction u_1 satisfying $u_1(x) > 0$ for all $x \in [0,1]$. Moreover, $\mu_1 = \|T_h\|$, i.e., the largest eigenvalue equals the operator norm.

Proof. The claim is a special case of [184, Theorem 7.3] (when the compact Hermitian operators considered there are taken to be the graphon operators). See also [143, Theorem 19.2] and [147, Appendix A]. □

§6.1.3 Main theorems

Let λ_N be the largest eigenvalue of the adjacency matrix A_N of G_N. Write $\mathbb{P}_N^*$ to denote the law of λ_N/N.

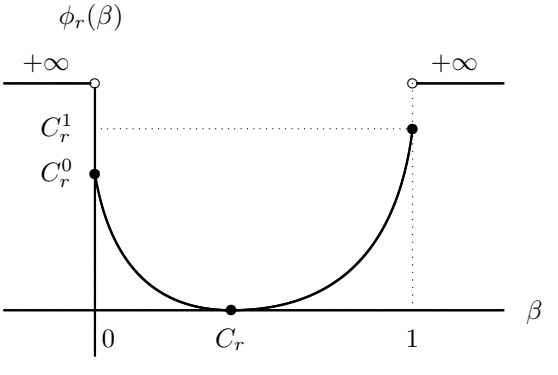


Figure 6.1: Graph of $\beta \mapsto \psi_r(\beta)$.

Theorem 6.1.4 (LDP for the largest eigenvalue).
Subject to (6.1)*, the sequence $(\mathbb{P}_N^*)_{N\in\mathbb{N}}$ satisfies the LDP on $\mathbb{R}$ with rate $\binom{N}{2}$ and with rate function*

$$\psi_r(\beta) = \inf_{\substack{\widetilde{h}\in\widetilde{\mathcal{W}} \\ \|T_{\widetilde{h}}\|=\beta}} J_r(\widetilde{h}) = \inf_{\substack{h\in\mathcal{W} \\ \|T_h\|=\beta}} I_r(h), \qquad \beta \in \mathbb{R}. \tag{6.12}$$

Proof. Note that $\lambda_N/N = \|T_{h^{G_N}}\|$, where h is any representative of $\widetilde{h}$ (we use the fact that $\|T_{\widetilde{h}}\| = \|T_{h^\phi}\|$ for all $\phi \in \mathcal{M}$). Also note that $\widetilde{h} \mapsto \|T_{\widetilde{h}}\|$ is a bounded and continuous function on $\widetilde{\mathcal{W}}$ [137, Exercises 6.1–6.2, Lemma 6.2]. Hence the claim follows from Theorem 6.1.1 via the contraction principle (see [167, Chapter 3]). □

Put

$$C_r = \|T_r\|. \tag{6.13}$$

When $\beta = C_r$, the graphon h that minimizes $I_r(h)$ such that $\|T_h\| = C_r$ is the reference graphon $h = r$ almost everywhere, for which $I_r(r) = 0$ and no large deviation occurs. When $\beta > C_r$, we are looking for graphons h with a larger operator norm. The large deviation cannot go above 1, which is represented by the constant graphon $h \equiv 1$, for which $I_r(1) = C_r^1$. Similarly, when $\beta < C_r$, we are looking for graphons h with a smaller operator norm. The large deviation cannot go below 0, which is represented by the constant graphon $h \equiv 0$, for which $I_r(0) = C_r^0$ (see Figure 6.1).

Theorem 6.1.5 (Properties of the rate function).
Subject to (6.1)*, the rate function in* (6.12) *satisfies the following.*

(i) The rate function ψ_r is continuous and unimodal on $[0,1]$, with a unique zero at C_r.

(ii) The rate function ψ_r is strictly decreasing on $[0, C_r]$ and strictly increasing on $[C_r, 1]$.

(iii) For every $\beta \in [0,1]$, the set of minimisers of the variational formula for $\psi_r(\beta)$ is non-empty and compact in $\widetilde{\mathcal{W}}$.

If the reference graphon r is of rank 1, i.e.,

$$r(x,y) = \nu(x)\,\nu(y), \qquad (x,y) \in [0,1]^2, \tag{6.14}$$

for some $\nu\colon\ [0,1] \to [0,1]$ that is bounded away from 0 and 1, then we are able to say more. Define

$$m_k = \int_{[0,1]} \nu(x)^k \, dx, \qquad k \in \mathbb{N}. \tag{6.15}$$

Note that $C_r = m_2$. Abbreviate

$$B_r = \int_{[0,1]^2} r(x,y)^3 (1 - r(x,y)) \, dx \, dy, \tag{6.16}$$

and note that $B_r = m_3^2 - m_4^2$. Further abbreviate

$$N_r^1 = \int_{[0,1]^2} \frac{1 - r(x,y)}{r(x,y)} \, dx \, dy, \qquad N_r^0 = \int_{[0,1]^2} \frac{r(x,y)}{1 - r(x,y)} \, dx \, dy. \tag{6.17}$$

Recall from Section 1.2.5 that $\mathcal{M}$ is the set of Lebesgue measure-preserving bijective maps $\phi\colon\ [0,1] \to [0,1]$.

Theorem 6.1.6 (Scaling of the rate function).
Let ψ_r be the rate function in (6.12)*.*

(i) Subject to (6.1) *and* (6.14)*,*

$$\psi_r(\beta) = K_r \, (\beta - C_r)^2 \, [1 + o(1)], \qquad \beta \to C_r, \tag{6.18}$$

with

$$K_r = \frac{C_r^2}{2B_r} = \frac{m_2^2}{2(m_3^2 - m_4^2)}. \tag{6.19}$$

(ii) Subject to (6.1),

$$C_r^1 - \psi_r(\beta) = (1-\beta)\Big(\log\frac{N_r^1}{1-\beta} + 1 + o(1)\Big), \qquad \beta \to 1. \tag{6.20}$$

(iii) Subject to (6.1),

$$C_r^0 - \psi_r(\beta) = \beta\Big(\log\frac{N_r^0}{\beta} + 1 + o(1)\Big), \qquad \beta \to 0. \tag{6.21}$$

Theorem 6.1.7 (Scaling of the minimisers).
Let $h_\beta \in \mathcal{W}$ be any minimiser of the second infimum in (6.12).

(i) Subject to (6.1) *and* (6.14),

$$\lim_{\beta\to C_r} (\beta - C_r)^{-1} \|h_\beta - r - (\beta - C_r)\Delta\|_2 = 0, \tag{6.22}$$

with

$$\Delta(x,y) = \frac{C_r}{B_r}\, r(x,y)^2(1-r(x,y)), \qquad (x,y) \in [0,1]^2. \tag{6.23}$$

(ii) Subject to (6.1),

$$\lim_{\beta\to 1} (1-\beta)^{-1} \|1 - h_\beta - (1-\beta)\Delta\|_2 = 0, \tag{6.24}$$

with

$$\Delta(x,y) = \frac{1}{N_r^1}\,\frac{1-r(x,y)}{r(x,y)}, \qquad (x,y) \in [0,1]^2. \tag{6.25}$$

(iii) Subject to (6.1),

$$\lim_{\beta\to 0} \beta^{-1} \|h_\beta - \beta\Delta\|_2 = 0, \tag{6.26}$$

with

$$\Delta(x,y) = \frac{1}{N_r^0}\,\frac{r(x,y)}{1-r(x,y)}, \qquad (x,y) \in [0,1]^2. \tag{6.27}$$

§6.1.4 Discussion and outline

Theorems. Theorem 6.1.5 confirms the picture of ψ_r drawn in Figure 6.1. It remains open whether or not ψ_r is convex. We do not expect ψ_r to be analytic, because bifurcations may occur in the set of minimisers of ψ_r as β is varied. Theorem 6.1.6 identifies the scaling of ψ_r around its minimum and near its end points, provided r is of rank 1. The inverse curvature $1/K_r$ equals the variance in the central limit theorem derived in [130]. This is in line with the standard folkore of large deviation theory. Theorem 6.1.7 identifies the corresponding scaling of the minimiser h_β of ψ_r. Interestingly, the scaling corrections are not rank 1. It remains open to determine what happens near C_r when r is not of rank 1 (see Appendix F).

Conditions. It would be interesting to investigate to what extent the condition on the reference graphon in (6.1) can be weakened to some form of integrability condition.

Especially for the upper bound in the LDP this is delicate, because the proof in [145] is based on block-graphon approximation (see [180]).

Outline of the chapter. The remainder of this chapter is organized as follows. In Section 6.2 we derive an expansion for the operator norm of a graphon around any graphon of rank 1. In Section 6.3 we prove our main theorems. In Appendix F we show how the expansion around reference graphons can be extended to finite rank.

§6.2 Expansion around rank-one graphons

In this section we show how we can expand the operator norm of a graphon around any graphon of rank 1. This prepares for the perturbation analysis in Sections 6.3.2–6.3.4.

Lemma 6.2.1 (Rank-one expansion).
Consider a graphon $\bar{h} \in \mathcal{W}$ of rank 1 such that $\bar{h}(x,y) = \bar{\nu}(x)\bar{\nu}(y)$, $(x,y) \in [0,1]^2$. For any $h \in \mathcal{W}$ such that $\|T_{h-\bar{h}}\| < \|T_h\|$, the operator norm $\mu = \|T_h\|$ is a solution of the equation

$$\mu = \sum_{n \in \mathbb{N}_0} \frac{1}{\mu^n} \mathcal{F}_n(h, \bar{h}), \tag{6.28}$$

where

$$\mathcal{F}_n(h, \bar{h}) = \int_{[0,1]^2} \bar{\nu}(x)(h - \bar{h})^n(x,y)\bar{\nu}(y)\, dx\, dy. \tag{6.29}$$

Proof. By Proposition 6.1.3, we have

$$T_h u = \mu u, \tag{6.30}$$

where μ equals both the norm and the largest eigenvalue of T_h, and u is the eigenfunction associated with μ. Put $g = h - \bar{h}$ and we have $(\mu - T_g)u = T_{\bar{h}}u$. This gives

$$u = (\mu - T_g)^{-1}\bar{\nu}\langle \bar{\nu}, u\rangle, \tag{6.31}$$

where we use that $\mu - T_g$ is invertible because $\|T_g\| = \|T_{h-\bar{h}}\| < \|T_h\|$. Hence, taking the inner product of u with $\bar{\nu}$ and observing that $\langle \bar{\nu}, u\rangle \neq 0$, we get

$$\langle \bar{\nu}, u\rangle = \langle \bar{\nu}, u\rangle \langle \bar{\nu}, (\mu - T_g)^{-1}\bar{\nu}\rangle, \tag{6.32}$$

which gives

$$\mu = \langle \bar{\nu}, (1 - T_g/\mu)^{-1}\bar{\nu}\rangle. \tag{6.33}$$

We can expand the above to get

$$\begin{aligned} \mu &= \left\langle \bar{\nu}, \sum_{n \in \mathbb{N}_0} \left(\frac{T_g}{\mu}\right)^n \bar{\nu} \right\rangle \\ &= \sum_{n \in \mathbb{N}_0} \frac{1}{\mu^n} \int_{[0,1]^{n+1}} \bar{\nu}(x_0) g(x_0, x_1) \cdots g(x_{n-1}, x_n) \bar{\nu}(x_n)\, dx_0\, dx_1 \cdots dx_n \\ &= \sum_{n \in \mathbb{N}_0} \frac{1}{\mu^n} \mathcal{F}_n(h, \bar{h}), \end{aligned} \tag{6.34}$$

and this completes the proof. □

Subject to (6.14), it follows from Lemma 6.2.1 with $h = \bar{h} = r$ that

$$C_r = \|T_r\| = m_2, \tag{6.35}$$

because only the term with $n = 0$ survives in the expansion.

Remark 6.2.2 (Higher rank).
The expansion around reference graphons of rank 1 can be extended to finite rank. We provide the details in Appendix F. In this chapter we focus on rank 1, for which Lemma 6.2.1 allows us to analyse the behavior of $\psi_r(\beta)$ near the values $\beta = C_r$, $\beta = 1$ and $\beta = 0$. Note that both the graphons $h = r$ and $h \equiv 1$ are of rank 1.

§6.3 Proofs of the main results

In this section we prove the theorems in Section 6.1.3. In Section 6.3.1 we prove Theorem 6.1.5. In the last three sections we prove Theorems 6.1.6–6.1.7 by analyzing graphon perturbations around the minimum of the rate function and near its end points. The proofs of the theorems rely on the variational formula in (6.12). Since the largest eigenvalue is invariant under relabeling of the vertices, we can work directly with I_r in (6.5) without worrying about the equivalence classes.

§6.3.1 Proof: properties of the rate function

Proof of Theorem 6.1.5. We follow [137, Chapter 6]. Even though this monograph deals with constant reference graphons only, most arguments carry over to r satisfying (6.1). Define

$$\psi_r^+(\beta) = \inf_{\substack{h\in\mathcal{W}\\ \|T_h\|\geq\beta}} I_r(h), \quad \psi_r^-(\beta) = \inf_{\substack{h\in\mathcal{W}\\ \|T_h\|\leq\beta}} I_r(h), \qquad \beta \in \mathbb{R}. \tag{6.36}$$

(i) Because $h \mapsto \|T_h\|$ is a nice graph parameter, in the sense of [137, Definition 6.1], it follows that $\beta \mapsto \psi_r^+(\beta)$ is non-decreasing and continuous, while $\beta \mapsto \psi_r^-(\beta)$ is non-increasing and continuous (see [137, Proposition 6.1]). The proof requires the fact that $\|f_n - f\|_2 \to 0$ implies $I_r(f_n) \to I_r(f)$ and that $I_r(f)$ is lower semi-continuous on $\mathcal{W}$. The continuity and unimodality of ψ_r follow from the proof of (iii). Moreover, since $I_r(h) = 0$ if and only if $h = r$ almost everywhere, it is immediate that C_r is the unique zero of ψ_r.

(ii) The proof is by contradiction. Suppose that $\beta \mapsto \psi_r^+(\beta)$ is not strictly increasing on $[C_r, 1]$. Then there exist $\beta_1, \beta_2 \in [C_r, 1]$, $\beta_1 < \beta_2$, such that ψ_r^+ is constant on $[\beta_1, \beta_2]$. Consequently, there exist minimisers $h_{\beta_1}^{\phi_1}, h_{\beta_2}^{\phi_2}$, $\phi_1, \phi_2 \in \mathcal{M}$, satisfying $r \leq h_{\beta_1}^{\phi_1} \leq h_{\beta_2}^{\phi_2}$, such that $I_r(h_{\beta_1}^{\phi_1}) = I_r(h_{\beta_2}^{\phi_2})$ and $\|T_{h_{\beta_1}^{\phi_1}}\| = \beta_1 < \beta_2 = \|T_{h_{\beta_2}^{\phi_2}}\|$. However, since $a \mapsto \mathcal{R}(a \mid b)$ is strictly increasing on $[b, 1]$ (recall (6.5)), it follows that $h_{\beta_1}^{\phi_1} = h_{\beta_2}^{\phi_2}$ almost everywhere. This in turn implies that $\|T_{h_{\beta_1}^{\phi_1}}\| = \|T_{h_{\beta_2}^{\phi_2}}\|$, which is a contradiction. A similar argument shows that $\beta \mapsto \psi_r^-(\beta)$ cannot have a flat piece on $[0, C_r]$.

(iii) The variational formulas in (6.36) achieve minimisers. In fact, the sets of minimiser are non-empty compact subsets of $\widetilde{\mathcal{W}}$ (see [137, Theorem 6.2]). In addition, all minimisers h of $\phi_r^+(h)$ satisfy $h \geq r$ almost everywhere, while all minimisers h of ϕ_r^- satisfy $h \leq r$ almost everywhere (see [137, Lemma 6.3]). Moreover, because

$$\begin{aligned} h_1 \geq h_2 \geq r \quad &\Longrightarrow \quad \|T_{h_1}\| \geq \|T_{h_2}\|,\ I_r(h_1) \geq I_r(h_2), \\ h_1 \leq h_2 \leq r \quad &\Longrightarrow \quad \|T_{h_1}\| \leq \|T_{h_2}\|,\ I_r(h_1) \leq I_r(h_2), \end{aligned} \tag{6.37}$$

(use that $a \mapsto \mathcal{R}(a \mid b)$ is unimodal on $[0,1]$ with unique zero at b), it follows that both variational formulas achieve minimisers with norm equal to β, and so

$$\psi_r(\beta) = \begin{cases} \psi_r^+(\beta), & \beta \geq C_r, \\ \psi_r^-(\beta), & \beta \leq C_r. \end{cases} \tag{6.38}$$

□

§6.3.2 Proof: perturbation around the minimum

Note that when $\beta = C_r$, the infimum in (6.12) is attained at $h = r$ and $\psi_r(C_r) = 0$. Take $\beta = C_r + \epsilon$ with $\epsilon > 0$ small, and assume that the infimum is attained by a graphon of the form $h = r + \Delta_\epsilon$, where $\Delta_\epsilon \colon [0,1]^2 \to \mathbb{R}$ represents a perturbation of the graphon r. Note that $r + \Delta_\epsilon \in \mathcal{W}$, hence we are dealing with a perturbation Δ_ϵ which is symmetric and bounded. We compare

$$\psi_r(C_r + \epsilon) = \inf_{\substack{\Delta_\epsilon \colon [0,1]^2 \to \mathbb{R} \\ r + \Delta_\epsilon \in \mathcal{W} \\ \|T_{r+\Delta_\epsilon}\| = C_r + \epsilon}} I_r(r + \Delta_\epsilon) \tag{6.39}$$

with $\psi_r(C_r) = 0$ by computing the difference

$$\delta_r(\epsilon) = \psi_r(C_r + \epsilon) - \psi_r(C_r) = \psi_r(C_r + \epsilon) \tag{6.40}$$

and studying its behavior as $\epsilon \to 0$. Since $r(x,y) = \nu(x)\nu(y)$, $(x,y) \in [0,1]^2$, we can use Lemma 6.2.1 to control the norm of $T_h = T_{r+\Delta_\epsilon}$. Pick $\bar{h} = r$ and $h = r + \Delta_\epsilon$ in (6.28) such that $\|\Delta_\epsilon\|_2 \to 0$ as $\epsilon \to 0$. Note that $\|T_{\Delta_\epsilon}\| \leq \|\Delta_\epsilon\|_2 < C_r$ for ϵ small enough. Hence, writing out the expansion for the norm, we get

$$\|T_{r+\Delta_\epsilon}\| = C_r + \sum_{n \in \mathbb{N}} \frac{1}{\|T_{r+\Delta_\epsilon}\|^n} \mathcal{F}_n(r + \Delta_\epsilon, r). \tag{6.41}$$

Since $\|T_{r+\Delta_\epsilon}\| = C_r + \epsilon$, we have

$$C_r + \epsilon = C_r + \frac{\langle \nu, \Delta_\epsilon \nu \rangle}{C_r + \epsilon} + \sum_{n \in \mathbb{N} \setminus \{1\}} \frac{1}{(C_r + \epsilon)^n} \langle \nu, \Delta_\epsilon^n \nu \rangle \tag{6.42}$$

with $\langle \nu, \Delta_\epsilon \nu \rangle = \int_{[0,1]^2} r \Delta_\epsilon$. So

$$\epsilon(C_r + \epsilon) = \int_{[0,1]^2} r \Delta_\epsilon + \sum_{n \in \mathbb{N} \setminus \{1\}} \frac{1}{(C_r + \epsilon)^{n-1}} \langle \nu, \Delta_\epsilon^n \nu \rangle. \tag{6.43}$$

Since ν is bounded, using the generalized Hölder's inequality (see [179, Theorem 3.1]) we get

$$|\langle \nu, \Delta_\epsilon^n \nu \rangle| \le \|\Delta_\epsilon\|_2^n. \tag{6.44}$$

Since $\|\Delta_\epsilon\|_2 \to 0$ as $\epsilon \to 0$, we can choose ϵ small enough such that $\|\Delta_\epsilon\|_2 < \frac{1}{2}(C_r+\epsilon)$, which gives

$$\sum_{n\in\mathbb{N}\setminus\{1\}} \frac{1}{(C_r+\epsilon)^{n-1}} \langle \nu, \Delta_\epsilon^n \nu \rangle = O\big(\|\Delta_\epsilon\|_2^2\big). \tag{6.45}$$

The constraint $\|r+\Delta_\epsilon\| = C_r + \epsilon$ therefore reads

$$\int_{[0,1]^2} r\Delta_\epsilon = \epsilon C_r + \epsilon^2 + O\big(\|\Delta_\epsilon\|_2^2\big). \tag{6.46}$$

Observe that if $\Delta_\epsilon = \epsilon\Delta$ for some function $\Delta \in L^2([0,1]^2)$, then

$$\int_{[0,1]^2} r\Delta = C_r\,[1+o(1)]. \tag{6.47}$$

Small perturbation on a given region. In what follows we use the standard notation $o(\cdot)$, $O(\cdot)$, $\asymp$ to describe the asymptotic behavior in the limit as $\epsilon \to 0$. We first show that it is enough to consider Δ_ϵ of the form $\epsilon\Delta$ for some $\Delta \in L^2([0,1]^2)$, because these perturbations contribute to the minimum cost.

Lemma 6.3.1 (Order of minimal cost).
Let $\Delta_\epsilon : [0,1]^2 \to \mathbb{R}$ be such that $r+\Delta_\epsilon \in \mathcal{W}$ and $\|T_{r+\Delta_\epsilon}\| = C_r + \epsilon$. Then

$$I_r(r+\Delta_\epsilon) \ge 2\epsilon^2. \tag{6.48}$$

Moreover, if $\Delta_\epsilon = \epsilon\Delta$, then

$$I_r(r+\epsilon\Delta) = 2\epsilon^2 \int_{[0,1]^2} \frac{\Delta(x,y)^2}{4r(x,y)(1-r(x,y))}\,dx\,dy\,[1+o(1)], \qquad \epsilon \to 0. \tag{6.49}$$

Proof. Fix $b \in [0,1]$ and abbreviate (recall (6.6))

$$\chi(a) = \mathcal{R}(a \mid b) = a\log\frac{a}{b} + (1-a)\log\frac{1-a}{1-b}, \qquad a \in [0,1]. \tag{6.50}$$

Note that

$$\chi(b) = \chi'(b) = 0, \quad \chi''(a) \ge 4, \qquad a \in [0,1]. \tag{6.51}$$

Consequently,

$$\chi(a) \ge 2(a-b)^2, \qquad a \in [0,1], \tag{6.52}$$

and hence

$$I_r(r+\Delta_\epsilon) \ge 2\int_{[0,1]^2} \Delta_\epsilon^2 = 2\|\Delta_\epsilon\|_2^2. \tag{6.53}$$

Next observe that

$$C_r + \epsilon = \|T_{r+\Delta_\epsilon}\| = \|T_r + T_{\Delta_\epsilon}\| \le \|T_r\| + \|T_{\Delta_\epsilon}\| \le C_r + \|\Delta_\epsilon\|_2, \tag{6.54}$$

which gives $\|\Delta_\epsilon\|_2 \geq \epsilon$. Inserting this lower bound into (6.53), we get (6.48). To get (6.49), we need a higher-order expansion of χ, namely,

$$\chi(x) = \tfrac{1}{2}\chi''(b)(x-b)^2 + O((x-b)^3), \qquad x \to b. \tag{6.55}$$

Since r is bounded away from 0 and 1, and the constraint $r + \Delta_\epsilon \in \mathcal{W}$ implies that $\Delta_\epsilon(x,y) \in [-1,1]$, we see that the third-order term is smaller than the second-order term when $\Delta_\epsilon = \epsilon\Delta$. Hence (6.49) follows. □

Next we consider different types of small perturbations in a given region and compute their total cost.

Lemma 6.3.2 (Cost of small perturbations).
Let $B \subseteq [0,1]^2$ be a measurable region with area $|B|$. Suppose that $\Delta_\epsilon = \epsilon^\alpha \Delta$ on B, with $\epsilon > 0$, $\alpha > 0$ and $\Delta\colon [0,1]^2 \to \mathbb{R}$. Then the contribution of B to the cost $I_r(h)$ is

$$\int_B \mathcal{R}(h \mid r) = \epsilon^{2\alpha} \int_B \frac{\Delta^2}{2r(1-r)}[1 + o(1)], \qquad \epsilon \to 0. \tag{6.56}$$

If the integral diverges, then the contribution decays slower than $\epsilon^{2\alpha}$.

Proof. The proof is similar to that of Lemma 6.3.1. □

Approximation by block graphons. We next introduce block graphons, which will be useful for our perturbation analysis. It follows from Lemma 6.3.1 that optimal perturbations with Δ_ϵ must satisfy $\|\Delta_\epsilon\|_2 \asymp \epsilon$, and hence it is desirable to have $\Delta_\epsilon = \epsilon\Delta$. We argue through block graphon approximations that this is indeed the case.

Definition 6.3.3 (Block graphons).
Let $\mathcal{W}_N \subset \mathcal{W}$ be the space of graphons with N blocks having a constant value on each of the blocks, i.e., $f \in \mathcal{W}_N$ is of the form

$$f(x,y) = \begin{cases} f_{i,j}, & \text{if } (x,y) \in B_i \times B_j, \\ 0, & \text{otherwise,} \end{cases} \tag{6.57}$$

where $B_i = [\frac{i-1}{N}, \frac{i}{N})$, $1 \leq i \leq N-1$ and $B_N = [\frac{N-1}{N}, 1]$ and $f_{i,j} \in [0,1]$. Write $B_{i,j} = B_i \times B_j$. With each $f \in \mathcal{W}$ associate the block graphon $f_N \in \mathcal{W}_N$ given by

$$f_N(x,y) = N^2 \int_{B_{i,j}} f(x',y')\,dx'\,dy' = \bar{f}_{N,ij}, \qquad (x,y) \in B_{i,j}. \tag{6.58}$$

Observe that if f_N is the block graphon associated with a graphon f, then

$$\|T_{f_N} - T_f\| = \|T_{f_N - f}\| \leq \|f_N - f\|_2. \tag{6.59}$$

We know from [137, Proposition 2.6] that, for any $f \in \mathcal{W}$ and its associated sequence of block graphons $(f_N)_{N\in\mathbb{N}}$, $\|f_N - f\|_2 \to 0$ and hence $\lim_{N\to\infty} \|T_{f_N}\| = \|T_f\|$. The following lemma shows that the cost function associated with the graphons r and f is well approximated by the cost function associated with the block graphons r_N and f_N.

Lemma 6.3.4 (Convergence of the cost function).
For any $f \in \mathcal{W}$

$$\lim_{N\to\infty} I_{r_N}(f_N) = I_r(f). \tag{6.60}$$

Proof. Since $f \in L^2([0,1]^2)$, f_N is bounded. The assumption in (6.1) implies that $\eta \le r_N \le 1-\eta$ for all $N \in \mathbb{N}$. We know from [145, Lemma 2.3] that there exists a constant $c > 0$ independent of f such that

$$|I_{r_N}(f) - I_r(f)| \le c\,\|r_N - r\|_1 \le c\,\|r_N - r\|_2. \tag{6.61}$$

Hence

$$\begin{aligned} |I_{r_N}(f_N) - I_r(f)| &\le |I_{r_N}(f_N) - I_r(f_N)| + |I_r(f_N) - I_r(f)| \\ &\le c\,\|r_N - r\|_2 + |I_r(f_N) - I_r(f)|. \end{aligned} \tag{6.62}$$

Since $\lim_{N\to\infty} \|r_N - r\|_2 = 0$, the first term tends to zero. Since $\lim_{N\to\infty} \|f_N - f\|_2 = 0$ and I_r is continuous in the L^2-topology on $\mathcal{W}$ (see [180, Lemma 3.4]), also the second term tends to zero and the claim follows. □

Block graphon perturbations. In what follows we fix $N \in \mathbb{N}$, analyze different types of perturbation and identify which one is optimal. For each $N \in \mathbb{N}$, we associate with the perturbed graphon $h = r + \Delta_\epsilon$ the block graphon $h_N \in \mathcal{W}_N$ given by

$$\overline{h}_{N,ij}(x,y) = \overline{r}_{N,ij}(x,y) + \overline{\Delta_\epsilon}_{N,ij}(x,y), \qquad (x,y) \in B_{i,j}, \tag{6.63}$$

with

$$\overline{r}_{N,ij} = N^2 \int_{B_{i,j}} r(x',y')\,dx'\,dy', \quad \overline{\Delta_\epsilon}_{N,ij} = N^2 \int_{B_{i,j}} \Delta_\epsilon(x',y')\,dx'\,dy'. \tag{6.64}$$

Observe that optimal perturbations must have $\|\Delta_\epsilon\|_2 = O(\epsilon)$, and hence the constraint in (6.46) becomes

$$\sum_{i,j=1}^{N} \int_{B_{i,j}} r(x,y)\Delta_\epsilon(x,y)\,dx\,dy = \sum_{i,j=1}^{N} \frac{1}{N^2}\,\overline{r\Delta_\epsilon}_{N,ij} = C_r\epsilon\,[1+o(1)], \qquad \epsilon \to 0. \tag{6.65}$$

The block constraint in (6.65) implies that the sum over each block must be of order ϵ. We therefore must have that

$$\overline{r\Delta_\epsilon}_{N,ij} = O(\epsilon), \qquad \epsilon \to 0, \qquad \forall (i,j), \tag{6.66}$$

which means that

$$\overline{\Delta_\epsilon}_{N,ij} = O(\epsilon), \qquad \epsilon \to 0, \qquad \forall (i,j), \tag{6.67}$$

since (6.1) implies that $\overline{r\Delta_\epsilon}_{N,ij} \asymp \overline{\Delta_\epsilon}_{N,ij}$. There are the following two possible cases.

(I) All blocks contribute to the constraint with a term of order ϵ (*balanced perturbation*).

(II) Some blocks contribute to the constraint with a term of order ϵ and some with $o(\epsilon)$ (*unbalanced perturbation*).

Perturbations of type (I) consist of a small perturbation on each block, i.e., $\overline{\Delta_\epsilon}_{N,ij} \asymp \epsilon$ for each block $B_{i,j}$. By Lemma 6.3.2, this contributes a term of order ϵ^2 to the total cost. Since all blocks have the same type of perturbation, they all contribute in the same way, and so we get $I_{r_N}(h_N) \asymp \epsilon^2$. We will see in Corollary 6.3.6 that perturbations of type (II) are worse than perturbations of type (I). Let $1 \le k \le N^2-1$ be the number of blocks that contribute a term of order $o(\epsilon)$ to the constraint, i.e., $\overline{\Delta_\epsilon}_{N,ij} = o(\epsilon)$. By Lemma 6.3.2, these blocks contribute order $o(\epsilon^2)$ to the total cost. The remaining blocks must fall in the class of blocks of type (I), with a perturbation of order ϵ on each of them. Corollary 6.3.6 below shows that the cost function attains its infimum when the small perturbation of order ϵ is uniform on $[0,1]^2$.

Optimal perturbation. We have shown that perturbations of type (I) lead to the minimal total cost. They consist of perturbations of order ϵ on all blocks, and hence on $[0,1]^2$. A sequence of such perturbations $(\Delta_{\epsilon,N})_{N\in\mathbb{N}}$ converges to a perturbation Δ_ϵ as $N \to \infty$. We can identify the cost of $\Delta_\epsilon = \epsilon\Delta$ with $\Delta\colon [0,1]^2 \to \mathbb{R}$, which we refer to as balanced perturbation.

Lemma 6.3.5 (Balanced perturbations).
Suppose that $\Delta_\epsilon = \epsilon\Delta$ with $\Delta\colon [0,1]^2 \to \mathbb{R}$. Let $\mathcal{M}$ be the set of Lebesgue measure-preserving bijective maps. Then

$$\delta_r(\epsilon) = K_r\epsilon^2\,[1+o(1)], \qquad \epsilon \to 0, \tag{6.68}$$

with

$$K_r = \frac{1}{2}C_r^2 \inf_{\phi\in\mathcal{M}} \frac{D_r^\phi}{(B_r^\phi)^2}, \tag{6.69}$$

where $B_r^\phi = \int_{[0,1]^2} r^\phi r^2(1-r)$ and $D_r^\phi = \int_{[0,1]^2} (r^\phi)^2 r(1-r)$.

Proof. The constraint in (6.46) becomes

$$\int_{[0,1]^2} r\Delta = C_r\,[1+o(1)], \qquad \epsilon \to 0, \tag{6.70}$$

and we get

$$\begin{aligned}\delta_r(\epsilon) &= \inf_{\substack{\Delta\colon [0,1]^2\to\mathbb{R} \\ r+\epsilon\Delta\in\mathcal{W} \\ \int_{[0,1]^2} r\Delta = C_r\,[1+o(1)]}} I_r(r+\epsilon\Delta) \\ &= \inf_{\substack{\Delta\colon [0,1]^2\to\mathbb{R} \\ r+\epsilon\Delta\in\mathcal{W} \\ \int_{[0,1]^2} r\Delta = C_r\,[1+o(1)]}} \int_{[0,1]^2} \mathcal{R}\big((r+\epsilon\Delta)(x,y) \mid r(x,y)\big)\,dx\,dy.\end{aligned} \tag{6.71}$$

By Lemma 6.3.2 (with $\alpha = 1$), we have

$$\delta_r(\epsilon) = K_r\epsilon^2\,[1+o(1)], \qquad \epsilon \to 0, \tag{6.72}$$

with

$$K_r = \inf_{\substack{\Delta\colon [0,1]^2\to\mathbb{R} \\ r+\epsilon\Delta\in\mathcal{W} \\ \int_{[0,1]^2} r\Delta = C_r\,[1+o(1)]}} \int_{[0,1]^2} \frac{\Delta(x,y)^2}{2r(x,y)(1-r(x,y))}\,dx\,dy. \tag{6.73}$$

The term $1+o(1)$ in (6.72) arises after we scale Δ by $1+o(1)$ in order to force $\int_{[0,1]^2} r\Delta = C_r$. Note that the optimization problem in (6.73) no longer depends on ϵ.

We can apply the method of Lagrange multipliers to solve this constrained optimization problem. To that end we define the Lagrangian

$$\mathcal{L}_{A_r}(\Delta) = \int_{[0,1]^2} \frac{\Delta^2}{2\,r(1-r)} + A_r \int_{[0,1]^2} r\Delta, \tag{6.74}$$

where A_r is a Langrange multiplier. Since $\int_{[0,1]^2} r = \int_{[0,1]^2} r^\phi$ for any Lebesgue measure-preserving bijective map $\phi \in \mathcal{M}$, we get that the minimizer (in the space of functions from $[0,1]^2 \to \mathbb{R}$) is of the form

$$\Delta^\phi(x,y) = -A_r\, r^\phi(x,y)r(x,y)(1-r(x,y)), \qquad (x,y)\in[0,1]^2, \qquad \phi\in\mathcal{M}. \tag{6.75}$$

We pick A_r such that the constraint is satisfied, i.e.,

$$-A_r B_r^\phi = C_r\,[1+o(1)] \tag{6.76}$$

with

$$B_r^\phi = \int_{[0,1]^2} r(x,y)^\phi r(x,y)^2(1-r(x,y))\,dx\,dy. \tag{6.77}$$

We get

$$\Delta^\phi(x,y) = \frac{C_r}{B_r^\phi}\, r^\phi(x,y)r(x,y)(1-r(x,y)), \qquad (x,y)\in[0,1]^2, \qquad \phi\in\mathcal{M}, \tag{6.78}$$

and

$$K_r = \inf_{\phi\in\mathcal{M}} \int_{[0,1]^2} \frac{(\Delta^\phi)^2}{2r(1-r)} = \frac{1}{2}C_r^2 \inf_{\phi\in\mathcal{M}} \frac{D_r^\phi}{(B_r^\phi)^2} \tag{6.79}$$

with

$$D_r^\phi = \int_{[0,1]^2} (r^\phi)^2 r(1-r). \tag{6.80}$$

This completes the proof. □

We next show that the infimum in (6.79) is uniquely attained when ϕ is the identity. For this we show that $D_r^\phi/(B_r^\phi)^2 \geq 1/B_r$ with equality if and only if $\phi = \mathrm{Id}$.

Indeed, write

$$\begin{aligned}
&B_r D_r^\phi - (B_r^\phi)^2 \\
&\quad = \int_{[0,1]^2} r(x,y)(1-r(x,y))\,dx\,dy \int_{[0,1]^2} r(\bar{x},\bar{y})(1-r(\bar{x},\bar{y}))\,d\bar{x}\,d\bar{y} \\
&\qquad \times \Big(r(x,y)^2 r^\phi(\bar{x},\bar{y})^2 - r(x,y) r^\phi(x,y) r(\bar{x},\bar{y}) r^\phi(\bar{x},\bar{y}) \Big) \\
&\quad = \int_{[0,1]^2} r(x,y)(1-r(x,y))\,dx\,dy \int_{[0,1]^2} r(\bar{x},\bar{y})(1-r(\bar{x},\bar{y}))\,d\bar{x}\,d\bar{y} \\
&\qquad \times \frac{1}{2} \Big(r(x,y)^2 r^\phi(\bar{x},\bar{y})^2 + r^\phi(x,y)^2 r(\bar{x},\bar{y})^2 - 2 r(x,y) r^\phi(x,y) r(\bar{x},\bar{y}) r^\phi(\bar{x},\bar{y}) \Big) \\
&\quad = \int_{[0,1]^2} r(x,y)(1-r(x,y))\,dx\,dy \int_{[0,1]^2} r(\bar{x},\bar{y})(1-r(\bar{x},\bar{y}))\,d\bar{x}\,d\bar{y} \\
&\qquad \times \frac{1}{2} \Big(r(x,y) r^\phi(\bar{x},\bar{y}) - r^\phi(x,y) r(\bar{x},\bar{y}) \Big)^2 ,
\end{aligned} \tag{6.81}$$

where the second equality uses the symmetry between the integrals. Hence we obtain $B_r D_r^\phi - (B_r^\phi)^2 \geq 0$, with equality if and only if $r(x,y)/r^\phi(x,y) = C$ for almost every $(x,y) \in [0,1]^2$. Clearly, for non-constant r this can hold only for $C = 1$, which amounts to $\phi = \mathrm{Id}$.

We conclude that the infimum in (6.79) equals $1/B_r$, and so we find that

$$K_r = \frac{C_r^2}{2B_r}. \tag{6.82}$$

Finally, note that $C_r = m_2$ by (6.35), and that $B_r = m_3^2 - m_4^2$ by (6.15). This settles the expression for K_r in (6.19).

Corollary 6.3.6 (Unbalanced perturbations).
Perturbations of order ϵ that are not balanced, i.e., that do not cover the entire unit square $[0,1]^2$, are worse than the balanced perturbation in Lemma 6.3.5.

Proof. The argument of the variational formula can be reduced to an integral that considers only those regions that contribute order ϵ^2, which constitute a subset of $[0,1]^2$. Applying the method of Lagrange multipliers as in Lemma 6.3.5, we obtain that the solution is given by

$$\delta_r(\epsilon) = [1 + o(1)]\, K_r' \epsilon^2, \qquad \epsilon \to 0, \tag{6.83}$$

with $K_r' > K_r$. The strict inequality comes from the fact that the optimal balanced perturbation Δ^{Id} found in (6.75) is non-zero everywhere. □

Proof of Theorems 6.1.6(i)–6.1.7(i). We have shown that a balanced perturbation is optimal and we have identified in (6.78) the form of the optimal balanced perturbation. The claim in Theorem 6.1.6(i) is settled by Lemma 6.3.5 and (6.82), while (6.78) settles the claim in Theorem 6.1.7(i). □

§6.3.3 Proof: perturbation near the right end

For $\beta = 1-\epsilon$ consider a graphon of the form $h = 1-\Delta_\epsilon$, where $\Delta_\epsilon\colon [0,1]^2 \to [0,\infty)$ represents a symmetric and bounded perturbation of the constant graphon $h \equiv 1$. We compare

$$\psi_r(1-\epsilon) = \inf_{\substack{\Delta_\epsilon\colon [0,1]^2\to[0,\infty)\\ 1-\Delta_\epsilon\in\mathcal{W}\\ \|T_{1-\Delta_\epsilon}\|=1-\epsilon}} I_r(1-\Delta_\epsilon) \tag{6.84}$$

with

$$C_r^1 = I_r(1) \tag{6.85}$$

by computing the difference

$$\delta_r(\epsilon) = \psi_r(1) - \psi_r(1-\epsilon) \tag{6.86}$$

and studying its behavior as $\epsilon \to 0$. Since $I_r(1)$ is a constant, we can write

$$\delta_r(\epsilon) = \sup_{\substack{\Delta_\epsilon\colon [0,1]^2\to[0,\infty)\\ 1-\Delta_\epsilon\in\mathcal{W}\\ \|T_{1-\Delta_\epsilon}\|=1-\epsilon}} \big(I_r(1) - I_r(1-\Delta_\epsilon)\big). \tag{6.87}$$

We again use the expansion in Lemma 6.2.1. Pick $\bar{h} = 1$ and $h = 1-\Delta_\epsilon$ in (6.28), to get

$$\|T_{1-\Delta_\epsilon}\| = 1 + \sum_{n\in\mathbb{N}} \frac{1}{\|T_{1-\Delta_\epsilon}\|^n}\, \mathcal{F}_n(1-\Delta_\epsilon, 1). \tag{6.88}$$

Since $\|T_{1-\Delta_\epsilon}\| = 1-\epsilon$, this gives

$$1-\epsilon = 1 + \frac{\langle 1, (-\Delta_\epsilon)1\rangle}{1-\epsilon} + \frac{\langle 1, (-\Delta_\epsilon)^2 1\rangle}{(1-\epsilon)^2} + \sum_{n\in\mathbb{N}\setminus\{1,2\}} \frac{\langle 1, (-\Delta_\epsilon)^n 1\rangle}{(1-\epsilon)^n}. \tag{6.89}$$

For $\epsilon \to 0$ we have $\|\Delta_\epsilon\|_2 \to 0$ and $|\langle 1, (-\Delta_\epsilon)^n 1\rangle| = O(\|\Delta_\epsilon\|_2^n)$. Therefore

$$\epsilon(1-\epsilon) = \int_{[0,1]^2} \Delta_\epsilon - \frac{\langle 1, \Delta_\epsilon^2 1\rangle}{(1-\epsilon)} + O\big(\|\Delta_\epsilon\|_2^3\big). \tag{6.90}$$

The restriction $1-\Delta_\epsilon \in \mathcal{W}$ implies that $\Delta_\epsilon \in [0,1]$. Hence $\|\Delta_\epsilon\|_2^2 \le \|\Delta_\epsilon\|_1$. Moreover,

$$1-\epsilon = \|T_{1-\Delta_\epsilon}\| \le \|1-\Delta_\epsilon\|_2 \le \sqrt{\|1-\Delta_\epsilon\|_1}. \tag{6.91}$$

Since $\|1-\Delta_\epsilon\|_1 = 1 - \|\Delta_\epsilon\|_1$, we have

$$\|\Delta_\epsilon\|_1 \le 1 - (1-\epsilon)^2 = \epsilon(2-\epsilon). \tag{6.92}$$

Since $\|\Delta_\epsilon\|_2^3 = O(\epsilon^{3/2})$, (6.90) reads

$$\frac{1}{1-\epsilon}\int_{[0,1]^3} \Delta_\epsilon(x,y)(1-\Delta_\epsilon(y,z))\,dx\,dy\,dz - \frac{\epsilon}{1-\epsilon}\,\|\Delta_\epsilon\|_1 = \epsilon(1-\epsilon) + O(\epsilon^{3/2}), \tag{6.93}$$

which, because $\|\Delta_\epsilon\|_1 = O(\epsilon)$, further reduces to

$$\int_{[0,1]^3} \Delta_\epsilon(x,y)(1-\Delta_\epsilon(y,z))\,dx\,dy\,dz = \epsilon\,[1+O(\epsilon^{1/2})]. \tag{6.94}$$

Note that when $\Delta_\epsilon = \epsilon\Delta$, the constraint reads

$$\int_{[0,1]^2} \Delta = 1 + O(\epsilon^{1/2}), \qquad \epsilon \to 0. \tag{6.95}$$

The following lemma gives an upper bound for $I_r(1) - I_r(1-\Delta_\epsilon)$.

Lemma 6.3.7 (Order of minimal cost).
Let $\Delta_\epsilon : [0,1]^2 \to [0,1]$ be such that $1-\Delta_\epsilon \in \mathcal{W}$ and $\|T_{1-\Delta_\epsilon}\| = 1-\epsilon$. Then, for ϵ small enough,

$$I_r(1) - I_r(1-\Delta_\epsilon) \le \|\Delta_\epsilon\|_1 \log \frac{1}{\|\Delta_\epsilon\|_1} + O(\|\Delta_\epsilon\|_1). \tag{6.96}$$

Moreover, $\delta_r(\epsilon) \le \epsilon \log \frac{1}{\epsilon} + O(\epsilon)$.

Proof. Abbreviate (recall (6.6))

$$\chi(a) = \mathcal{R}(a \mid r) = a \log \frac{a}{r} + (1-a)\log \frac{1-a}{1-r}, \qquad a \in [0,1]. \tag{6.97}$$

Then

$$\chi(1) - \chi(1-\Delta_\epsilon(x,y)) = \Delta_\epsilon(x,y) \log\left(\frac{1-\Delta_\epsilon(x,y)}{\Delta_\epsilon(x,y)} \frac{1-r(x,y)}{r(x,y)}\right) - \log(1-\Delta_\epsilon(x,y)), \tag{6.98}$$

and so

$$I_r(1) - I_r(1-\Delta_\epsilon) = \int_{[0,1]^2} \left(\Delta_\epsilon \log\left(\frac{1-\Delta_\epsilon}{\Delta_\epsilon}\frac{1-r}{r}\right) - \log(1-\Delta_\epsilon)\right). \tag{6.99}$$

Let μ_ϵ be the probability measure on $[0,1]^2$ whose density with respect to the Lebesgue measure is $Z_\epsilon^{-1}(1-\Delta_\epsilon(x,y))$, where $Z_\epsilon = \int_{[0,1]^2}(1-\Delta_\epsilon) = 1 - O(\epsilon)$. Since $u \mapsto \bar{s}(u) = u\log(1/u)$ is strictly concave, by Jensen's inequality we have

$$\begin{aligned} \int_{[0,1]^2} \Delta_\epsilon \log\left(\frac{1-\Delta_\epsilon}{\Delta_\epsilon}\right) &= Z_\epsilon \int_{[0,1]^2} \mu_\epsilon \bar{s}\left(\frac{\Delta_\epsilon}{1-\Delta_\epsilon}\right) \\ &\le Z_\epsilon \bar{s}\big(Z_\epsilon^{-1}\|\Delta_\epsilon\|_1\big) \\ &= \|\Delta_\epsilon\|_1 \log\left(\frac{Z_\epsilon}{\|\Delta_\epsilon\|_1}\right). \end{aligned} \tag{6.100}$$

Moreover,

$$\int_{[0,1]^2} \Delta_\epsilon \log\left(\frac{1-r}{r}\right) = O(\|\Delta_\epsilon\|_1), \quad -\int_{[0,1]^2} \log(1-\Delta_\epsilon) = O(\|\Delta_\epsilon\|_1). \tag{6.101}$$

Hence

$$I_r(1) - I_r(1-\Delta_\epsilon) \le \|\Delta_\epsilon\|_1 \log \frac{1}{\|\Delta_\epsilon\|_1} + O(\|\Delta_\epsilon\|_1), \qquad \epsilon \to 0, \tag{6.102}$$

and since $\|\Delta_\epsilon\|_1 = O(\epsilon)$ also $\delta_r(\epsilon) \le \epsilon \log \frac{1}{\epsilon} + O(\epsilon)$. □

The following is the analogue of Lemma 6.3.2 for perturbations near the right end.

Lemma 6.3.8 (Cost of small perturbations).
Let $B \subseteq [0,1]^2$ be a measurable region of area $|B|$. Suppose that $\Delta_\epsilon = \epsilon^\alpha \Delta$ on B with $\epsilon > 0$, $\alpha > 0$ and $\Delta\colon\ [0,1]^2 \to [0,\infty)$. Then the contribution of B to the cost $I_r(h)$ is

$$\int_B \mathcal{R}(1 \mid r) - \mathcal{R}(h \mid r) = \int_B \epsilon^\alpha \Delta \log\left(\frac{1-r}{\epsilon^\alpha \Delta r}\right)[1+o(1)], \qquad \epsilon \to 0. \tag{6.103}$$

Proof. Observe that

$$\mathcal{R}(1 \mid r) - \mathcal{R}(1-\epsilon^\alpha\Delta \mid r) = \epsilon^\alpha \Delta \log\left(\frac{1-r}{\epsilon^\alpha \Delta r}\right)[1+o(1)], \qquad \epsilon \to 0. \tag{6.104}$$

The proof is analogous to that of Lemma 6.3.2. □

Following the argument in Section 6.3.2, we can approximate the cost function by using block graphons. The constraint becomes

$$\sum_{i,j=1}^N \int_{B_{i,j}} \Delta_{\epsilon,N}(x,y)\,dx\,dy = \sum_{i,j=1}^N \frac{1}{N^2}\overline{\Delta_\epsilon}_{N,ij} = \epsilon\,[1+o(1)], \qquad \epsilon \to 0. \tag{6.105}$$

The block constraint in (6.105) implies that the sum over each block must be of order ϵ. Hence

$$\overline{\Delta_\epsilon}_{N,ij} = O(\epsilon), \qquad \epsilon \to 0, \qquad \forall\,(i,j). \tag{6.106}$$

There are two cases to distinguish: all blocks contribute to the constraint with a term of order ϵ (balanced perturbation), or some of the blocks contribute to the constraint with a term of order ϵ and some with $o(\epsilon)$. Analogously to the analysis in Section 6.3.2, using Lemma 6.3.8, we can compute the total cost that different types of block perturbations produce. This again shows that the optimal perturbations are the balanced perturbations, consisting of small perturbations of order ϵ on every block. As $N \to \infty$, a sequence of such perturbations converges to a perturbation $\Delta_\epsilon = \epsilon\Delta$ with $\Delta\colon\ [0,1]^2 \to [0,\infty)$, which we analyze next.

Lemma 6.3.9 (Balanced perturbations).
Suppose that $\Delta_\epsilon = \epsilon\Delta$ with $\Delta\colon\ [0,1]^2 \to [0,\infty)$. Then

$$\delta_r(\epsilon) = \left(\epsilon + \epsilon \log\left(\frac{N_r^1}{\epsilon}\right)\right)[1+O(\epsilon^{1/2})] + O(\epsilon^2), \qquad \epsilon \to 0. \tag{6.107}$$

Proof. By (6.95) and (6.99),

$$\begin{aligned}\delta_r(\epsilon) &= \sup_{\substack{\Delta\colon\ [0,1]^2\to[0,\infty)\\ 1-\epsilon\Delta\in\mathcal{W}\\ \int_{[0,1]^2}\Delta = 1+O(\epsilon^{1/2})}} (I_r(1) - I_r(1-\epsilon\Delta)) \\ &= \sup_{\substack{\Delta\colon\ [0,1]^2\to[0,\infty)\\ 1-\epsilon\Delta\in\mathcal{W}\\ \int_{[0,1]^2}\Delta = 1+O(\epsilon^{1/2})}} \int_{[0,1]^2} \left(\epsilon\Delta \log\left(\frac{1-\epsilon\Delta}{\epsilon\Delta}\frac{1-r}{r}\right) - \log(1-\epsilon\Delta)\right).\end{aligned} \tag{6.108}$$

The integral in (6.108) equals

$$\int_{[0,1]^2}\left(\epsilon\Delta\log\left(\frac{1-r}{\epsilon\Delta r}\right)-(1-\epsilon\Delta)\log(1-\epsilon\Delta)\right)=\int_{[0,1]^2}\epsilon\Delta\log\left(\frac{1-r}{\epsilon\Delta r}\right)+\epsilon\int_{[0,1]^2}\Delta+O(\epsilon^2). \tag{6.109}$$

Hence

$$\delta_r(\epsilon)=\left(\epsilon+\sup_{\substack{\Delta\colon [0,1]^2\to[0,\infty)\\ \int_{[0,1]^2}\Delta=1}}\int_{[0,1]^2}\epsilon\Delta\log\left(\frac{1-r}{\epsilon\Delta r}\right)\right)[1+\mathrm{O}(\epsilon^{1/2})]+O(\epsilon^2), \tag{6.110}$$

where we scale Δ by $1+O(\epsilon^{1/2})$ in order to force $\int_{[0,1]^2}\Delta=1$. Note that the constraint under the supremum no longer depends on ϵ.

We can solve the optimization problem by applying the method of Lagrange multipliers. To that end we define the Lagrangian

$$\mathcal{L}_{A_r}(\Delta)=\int_{[0,1]^2}\epsilon\Delta\log\left(\frac{1-r}{\epsilon\Delta r}\right)+A_r\int_{[0,1]^2}\Delta, \tag{6.111}$$

where A_r is a Langrange multiplier. Since $\int_{[0,1]^2}\log\frac{1-r}{r}=\int_{[0,1]^2}\log\frac{1-r^\phi}{r^\phi}$ for any Lebesgue measure-preserving bijective map $\phi\in\mathcal{M}$, we get that the minimizer (in the space of functions from $[0,1]^2\to\mathbb{R}$) is of the form

$$\Delta^\phi(x,y)=e^{-\frac{\epsilon-A_r}{\epsilon}}\,\frac{1}{\epsilon}\,\frac{1-r^\phi(x,y)}{r^\phi(x,y)},\qquad (x,y)\in[0,1]^2,\qquad \phi\in\mathcal{M}. \tag{6.112}$$

We pick A_r such that the constraint $\int_{[0,1]^2}\Delta=1$ is satisfied. This gives

$$\Delta^\phi(x,y)=\frac{1}{N_r^1}\,\frac{1-r^\phi(x,y)}{r^\phi(x,y)},\qquad (x,y)\in[0,1]^2,\qquad \phi\in\mathcal{M}, \tag{6.113}$$

with $N_r^1=\int_{[0,1]^2}\frac{(1-r)}{r}$. Hence the supremum in (6.110) becomes

$$\sup_{\phi\in\mathcal{M}}\int_{[0,1]^2}\epsilon\Delta^\phi\log\left(\frac{1-r}{\epsilon\Delta^\phi r}\right). \tag{6.114}$$

We have

$$\int_{[0,1]^2}\epsilon\Delta^\phi\log\left(\frac{1-r}{\epsilon\Delta^\phi r}\right)=\epsilon\log\left(\frac{N_r^1}{\epsilon}\right)-\epsilon\int_{[0,1]^2}\Delta^\phi\log\left(\frac{\Delta^\phi}{\Delta}\right), \tag{6.115}$$

where we use that $\int_{[0,1]^2}\Delta^\phi=1$. Since the function $u\mapsto s(u)=u\log u$ is strictly convex on $[0,\infty)$, Jensen's inequality gives

$$\int_{[0,1]^2}\Delta^\phi\log\left(\frac{\Delta^\phi}{\Delta}\right)=\int_{[0,1]^2}\Delta\,s\left(\frac{\Delta^\phi}{\Delta}\right)\geq s\left(\int_{[0,1]^2}\Delta\,\frac{\Delta^\phi}{\Delta}\right)=s\left(\int_{[0,1]^2}\Delta^\phi\right)=s(1)=0, \tag{6.116}$$

where we use that $\int_{[0,1]^2} \Delta = 1$. Equality holds if and only if $\Delta = \Delta^\phi$ almost everywhere on $[0,1]^2$, which amounts to $\phi = \mathrm{Id}$. Hence the supremum in (6.114) is uniquely attained at $\phi = \mathrm{Id}$ and equals

$$\int_{[0,1]^2} \epsilon \, \frac{1}{N_r^1} \frac{(1-r)}{r} \log\left(\frac{N_r^1}{\epsilon}\right) = \epsilon \log\left(\frac{N_r^1}{\epsilon}\right). \tag{6.117}$$

Consequently, (6.110) gives (6.107), and this completes the proof. □

Proof of Theorems 6.1.6(ii)–6.1.7(ii). The claim in Theorem 6.1.6(ii) is settled by Lemma 6.3.9. Since we have shown that a balanced perturbation is optimal, (6.113) settles the claim in Theorem 6.1.7(ii). □

§6.3.4 Proof: perturbation near the left end

For $\beta = \epsilon$ consider a graphon of the form $h = \Delta_\epsilon$, where $\Delta_\epsilon \colon [0,1]^2 \to [0,\infty)$ represents a symmetric and bounded perturbation of the constant graphon $h \equiv 0$. We compare

$$\psi_r(\epsilon) = \inf_{\substack{\Delta_\epsilon \colon [0,1]^2 \to [0,\infty) \\ \Delta_\epsilon \in \mathcal{W} \\ \|T_{\Delta_\epsilon}\| = \epsilon}} I_r(\Delta_\epsilon) \tag{6.118}$$

with

$$\psi_r(0) = I_r(0) \tag{6.119}$$

by computing the difference

$$\delta_r(\epsilon) = \psi_r(\epsilon) - \psi_r(0) \tag{6.120}$$

and studying its behavior as $\epsilon \to 0$.

We claim that analyzing (6.120) is equivalent to analyzing

$$\delta_{\hat{r}}(\epsilon) = \phi_{\hat{r}}(1) - \phi_{\hat{r}}(1-\epsilon), \tag{6.121}$$

where $\hat{r}$ is the *reflection* of r defined as

$$\hat{r}(x,y) = 1 - r(x,y), \qquad (x,y) \in [0,1]^2. \tag{6.122}$$

Indeed,

$$I_r(0) = \int_{[0,1]^2} \mathcal{R}(0 \mid r) = \int_{[0,1]^2} \log\left(\frac{1}{1-r}\right) = \int_{[0,1]^2} \mathcal{R}(1 \mid \hat{r}) = I_{\hat{r}}(1) \tag{6.123}$$

and

$$\begin{aligned} I_r(\Delta_\epsilon) &= \int_{[0,1]^2} \mathcal{R}(\Delta_\epsilon \mid r) = \int_{[0,1]^2} \left(\Delta_\epsilon \log\left(\frac{\Delta_\epsilon}{r}\right) + (1-\Delta_\epsilon)\log\left(\frac{1-\Delta_\epsilon}{1-r}\right)\right) \\ &= \int_{[0,1]^2} \mathcal{R}(1-\Delta_\epsilon \mid \hat{r}) = I_{\hat{r}}(1-\Delta_\epsilon). \end{aligned} \tag{6.124}$$

We can therefore use the results in Section 6.3.3. From Lemma 6.3.9 we know that

$$\delta_{\hat{r}}(\epsilon) = \left(\epsilon + \epsilon \log\left(\frac{N^1_{\hat{r}}}{\epsilon}\right)\right)[1 + O(\epsilon^{1/2})] + O(\epsilon^2), \qquad \epsilon \to 0, \tag{6.125}$$

and hence we obtain

$$\delta_r(\epsilon) = \left(\epsilon + \epsilon \log\left(\frac{N^0_r}{\epsilon}\right)\right)[1 + O(\epsilon^{1/2})] + O(\epsilon^2), \qquad \epsilon \to 0. \tag{6.126}$$

The optimal perturbation is then given by the balanced perturbation $\Delta_\epsilon = \epsilon\Delta$ with

$$\Delta(x,y) = \frac{1}{N^0_r}\,\frac{r(x,y)}{1-r(x,y)}, \qquad (x,y) \in [0,1]^2, \tag{6.127}$$

with $N^0_r = \int_{[0,1]^2} \frac{r}{1-r}$.

Proof of Theorems 6.1.6(iii)–6.1.7(iii). The claim in Theorem 6.1.6(iii) is settled by the scaling in (6.126). Since we have shown that a balanced perturbation is optimal, (6.127) settles the claim in Theorem 6.1.7(iii). □

§F Appendix: finite-rank expansion

The following lemma shows how the expansion around reference graphons of rank 1 can be extended to finite rank.

Lemma F.1 (Finite-rank expansion).
Consider a graphon $\bar{h} \in \mathcal{W}$ such that

$$\bar{h}(x,y) = \sum_{i=1}^{k} \theta_i \bar{\nu}_i(x)\bar{\nu}_i(y), \qquad (x,y) \in [0,1]^2, \tag{6.128}$$

for some $k \in \mathbb{N}$, where $\theta_1 > \theta_2 \geq \ldots \geq \theta_k \geq 0$ and $\{\bar{\nu}_1, \bar{\nu}_2, \ldots, \bar{\nu}_k\}$ is an orthonormal set in $L^2([0,1])$. Then there exists an $\epsilon > 0$ such that, for any $h \in \mathcal{W}$ satisfying $\|T_{h-\bar{h}}\| < \min(\epsilon, \|T_h\|)$, the operator norm $\|T_h\|$ solves the equation

$$\|T_h\| = \lambda_k\left(\sum_{n\in\mathbb{N}_0} \frac{1}{\|T_h\|^n}\mathcal{F}_n(h,\bar{h})\right), \tag{6.129}$$

where $\lambda_k(M)$ is the largest eigenvalue of a $k \times k$ Hermitian matrix M, and $\mathcal{F}_n(h,\bar{h})$ is a $k \times k$ matrix whose (i,j)-th entry is

$$\sqrt{\theta_i\theta_j}\int_{[0,1]^2} \bar{\nu}_i(x)(h-\bar{h})^n(x,y)\bar{\nu}_j(y)\,dx\,dy \tag{6.130}$$

for $1 \leq i,j \leq k$ and $n \in \mathbb{N}_0$.

Proof. Put $\mu = \|T_h\|$, and let u be the eigenfunction corresponding to μ, i.e.,

$$T_h u = \mu u. \tag{6.131}$$

Put $g = h - \bar{h}$ and rewrite the above as

$$(\mu - T_g)u = T_{\bar{h}}u. \tag{6.132}$$

The assumption $\|T_{h-\bar{h}}\| < \|T_h\|$ implies that $\mu - T_g$ is invertible, which allows us to write

$$u = (\mu - T_g)^{-1}T_{\bar{h}}u = \sum_{j=1}^{k} \theta_j \langle \bar{\nu}_j, u\rangle (\mu - T_g)^{-1}\bar{\nu}_j. \tag{6.133}$$

For fixed $1 \leq i \leq k$, it follows that

$$\langle \bar{\nu}_i, u\rangle = \sum_{j=1}^{k} \theta_j \langle \bar{\nu}_j, u\rangle \langle \bar{\nu}_i, (\mu - T_g)^{-1}\bar{\nu}_j\rangle. \tag{6.134}$$

Multiplying both sides by $\mu\sqrt{\theta}_i$, we get

$$Mv = \mu v, \tag{6.135}$$

where $M = (M_{ij})_{1\leq i,j\leq k}$ is the $k \times k$ real symmetric matrix with elements

$$M_{ij} = \sqrt{\theta_i\theta_j}\left\langle \bar{\nu}_i, \left(1 - \frac{T_g}{\mu}\right)^{-1}\bar{\nu}_j\right\rangle, \qquad 1 \leq i,j \leq k, \tag{6.136}$$

and

$$v = \left(\sqrt{\theta}_1\langle \bar{\nu}_1, u\rangle, \ldots, \sqrt{\theta}_k\langle \bar{\nu}_k, u\rangle\right)'. \tag{6.137}$$

The first entry of v is non-zero for ϵ small with $\|T_g\| < \epsilon$. Thus, (6.135) means that μ is an eigenvalue of M. By studying the diagonal entries of M, we can shown with the help of the Gershgorin circle theorem that, for small $\|T_g\|$,

$$\mu = \lambda_k(M). \tag{6.138}$$

With the help of the observation

$$M_{ij} = \sqrt{\theta_i\theta_j}\sum_{n\in\mathbb{N}_0}\frac{1}{\mu^n}\langle \bar{\nu}_i, g^n\bar{\nu}_j\rangle, \qquad 1 \leq i,j \leq k, \tag{6.139}$$

i.e.,

$$M = \sum_{n\in\mathbb{N}_0}\frac{1}{\mu^n}\mathcal{F}_n(h,\bar{h}), \tag{6.140}$$

this completes the proof. □

Bibliography

References of Part I

[1] N. Abramson. *The ALOHA System - Another alternative for computer communication*. In: Proceedings of the Fall 1970 AFIPS Computer Conference 37, 281–285, 1970.

[2] F. Baccelli, B. Błaszczyszyn. *Stochastic Geometry and Wireless Networks: Volume I Theory*, Foundations and Trends in Networking 3(3/4), 249–449, 2009.

[3] F. Baccelli, B. Błaszczyszyn. *Stochastic Geometry and Wireless Networks: Volume II Applications*, Foundations and Trends in Networking 4(1/2), 1–312, 2009.

[4] J. Beltrán, C. Landim. *Tunneling and metastability of continuous time Markov chains*, Journal of Statistical Physics 140(6), 1065–1114, 2010.

[5] G. Ben Arous, R. Cerf. *Metastability of the three-dimensional Ising model on a torus at verylow temperatures*, Electronic Journal of Probability 1(10), 1–55, 1996.

[6] A. Bianchi, A. Gaudilliere. *Metastable states, quasi-stationary and soft measures*, Stochastic Processes and their Applications 126(6), 1622–1680, 2016.

[7] T. Bonald, S.C. Borst, N. Hegde, M. Jonckheere, A. Proutiere. *Flow-level performance and capacity of wireless networks with user mobility*, Queueing Systems 63(1-4), 131–164, 2009.

[8] T. Bonald, S.C. Borst, A. Proutiere. *How mobility impacts the flow-level performance of wireless data systems*. In: Proceedings of the IEEE INFOCOM 2004 Conference, volume 3, 1872–1881, 2004.

[9] R.R. Boorstyn, A. Kershenbaum. *Throughput analysis of multihop packet radio networks*. In: Proceedings of the ICC 1980 Conference, 1361–1366, 1980.

[10] R.R. Boorstyn, A. Kershenbaum, B. Maglaris, V. Sahin. *Throughput analysis in multihop CSMA packet radio networks*, IEEE Transactions on Communications 25(3), 267–274, 1987.

[11] S.C. Borst, N. Hegde, A. Proutiere. *Mobility-driven scheduling in wireless networks*. In:Proceedings of the IEEE INFOCOM 2009 Conference, 1260–1268, 2009.

[12] S.C. Borst, F. den Hollander, F.R. Nardi, M. Sfragara. *Transition time asymptotics of queue-based activation protocols for random-access neworks*. Stochastic Processes and Their Applications, 2020.

[13] S.C. Borst, F. den Hollander, F.R. Nardi, M. Sfragara. *Wireless random-access networks with bipartite interference graphs*. [arXiv:2001.02841], 2020.

[14] S.C. Borst, F. den Hollander, F.R. Nardi, S. Taati. *Crossover times in bipartite networks with activity constraints and time-varying switching rates*. [arXiv:1912.13011], 2019.

[15] S.C. Borst, A. Proutiere, N. Hegde. *Capacity of wireless data networks with intra- and inter-cell mobility*. In: Proceedings of the IEEE INFOCOM 2006 Conference, 1058–1069, 2006.

[16] S.C. Borst, F. Simatos. *A stochastic network with mobile users in heavy traffic*, Queueing Systems 74(1), 1–40, 2013.

[17] N. Bouman. Queue-based random access in wireless networks. Ph.D. thesis, Eindhoven University of Technology, 2013.

[18] N. Bouman, S.C. Borst, J.S.H. van Leeuwaarden. *Delay performance in random-access networks*, Queueing Systems 77, 211–242, 2014.

[19] A. Bovier, M. Eckhoff, V. Gayrard, M. Klein. *Metastability and small eigenvalues in Markov chains*, Journal of Physics A: Mathematical and General 33(46), L447–L451, 2000.

[20] A. Bovier, M. Eckhoff, V. Gayrard, M. Klein. *Metastability in stochastic dynamics of disordered mean-field models*, Probability Theory and Related Fields 119(1), 99–161, 2001.

[21] A. Bovier, M. Eckhoff, V. Gayrard, M. Klein. *Metastability and low lying spectra in reversible Markov chains*, Communications in Mathematical Physics 228(2), 219–255, 2002.

[22] A. Bovier, F. den Hollander. Metastability. Springer, 2005.

[23] A. Bovier, F. den Hollander. Metastability: A Potential-Theoretic Approach. Springer, 2015.

[24] A. Bovier, F. den Hollander, F.R. Nardi. *Sharp asymptotics for Kawasaki dynamics on a finite box with open boundary*, Probability Theory and Related Fields 135(2), 265–310, 2006.

[25] A. Bovier, F. den Hollander, C. Spitoni. *Homogeneous nucleation for Glauber and Kawasaki dynamics in large volumes at low temperatures*, Annals of Probability 38(2), 661–713, 2010.

[26] A. Bovier, F. Manzo. *Metastability in Glauber dynamics in the low temperature limit: beyond exponential asymptotics*, Journal of Statistical Physics 107(3-4), 757–779, 2002.

[27] M. Cassandro, A. Galves, E. Olivieri, M.E. Vares. *Metastable behavior of stochastic dynamics: A pathwise approach*, Journal of Statistical Physics 35(5-6), 603–634, 1984.

[28] E. Castiel, S.C. Borst, L. Miclo, F. Simatos, P.A. Whiting. *Induced idleness leads to deterministic heavy traffic limits for queue-based random-access algorithms*. [arXiv:1904.03980], 2019.

[29] F. Cecchi, S.C. Borst, J.S.H. van Leeuwaarden. *Throughput of CSMA networks with buffer dynamics*, Performance Evaluation 79, 216–234, 2014.

[30] R. Cerf, F. Manzo. *Nucleation and growth for the Ising model in d dimensions at very low temperatures*, Annals of Probability 41(6), 3697–3785, 2013.

[31] R. Cerf, F. Manzo. *A d dimensional nucleation and growth model*, Probability Theory and Related Fields 155(1/2), 427–449, 2013.

[32] E.N.M. Cirillo, F.R. Nardi. *Metastability for a stochastic dynamics with a parallel heat bath updating rule*, Journal of Statistical Physics 110(1-2), 183–217, 2003.

[33] E.N.M. Cirillo, F.R. Nardi. *Relaxation height in energy landscapes: An application to multiple metastable states*, Journal of Statistical Physics 150, 1080–1114, 2013.

[34] E.N.M. Cirillo, F.R. Nardi, J. Sohier. *Metastability for general dynamics with rare transitions: Escape time and critical configurations*, Journal of Statistical Physics 161, 365–403, 2015.

[35] E.N.M. Cirillo, F.R. Nardi, C. Spitoni. *Metastability for reversible probabilistic cellular automata with self-interaction*, Journal of Statistical Physics 132(3), 431–471, 2008.

[36] E.N.M. Cirillo, F.R. Nardi, C. Spitoni. *Sum of exit times in series of metastable states in Probabilistic Cellular Automata*, Cellular Automata and Discrete Complex Systems, Lecture Notes in Computer Science 9664, 105–119, 2016.

[37] E.N.M. Cirillo, F.R. Nardi, C. Spitoni. *Sum of exit times in a series of two metastable states*, European Physical Journal Special Topics 226(10), 2421–2438, 2017.

[38] P. Dehghanpour, R.H. Schonmann. *A nucleation and growth model*, Probability Theory and Related Fields 107(1), 123–135, 1997.

[39] P. Dehghanpour, R.H. Schonmann. *Metropolis dynamics relaxation via nucleation and growth*, Communications in Mathematical Physics 188(1), 89–119, 1997.

[40] A. Dembo, O. Zeitouni. Large Deviations Techniques and Applications, second edition. Springer, 1998.

[41] I. Dimitriou, N. Pappas. *Stable throughput and delay analysis of a random access network with queue-aware transmission*, IEEE Transactions on Wireless Communications 17(5), 3170–3184, 2018.

[42] I. Dimitriou, N. Pappas. *A queue-based random access scheme in network-level cooperative wireless networks*. In: Prooceedings of the ICC 2019 Conference, 1–6, 2019.

[43] S. Dommers. *Metastability of the Ising model on random regular graphs at zero temperature*, Probability Theory and Related Fields 167(1), 305–324, 2017.

[44] S. Dommers, F. den Hollander, O. Jovanovski, F.R. Nardi. *Metastability for Glauber dynamics on random graphs*, Annals of Applied Probability 27(4), 2130–2158, 2017.

[45] M. Durvy, O. Dousse, P. Thiran. *Border effects, fairness, and phase transitions in large wireless networks*. In: Proceedings of the IEEE INFOCOM 2008 Conference, 601–609, 2018.

[46] M. Durvy, O. Dousse, P. Thiran. *Modeling the 802.11 protocol under different capture and sensing capabilities*. In: Proceedings of the IEEE INFOCOM 2007 Conference, 2356–2560, 2017.

[47] M. Durvy, O. Dousse, P. Thiran. *On the fairness of large CSMA networks*, IEEE Journal on Selected Areas in Communications 27, 1093–1104, 2009.

[48] M. Durvy, P. Thiran. *A packing approach to compare slotted and non-slotted medium access control*. In: Proceedings of the IEEE INFOCOM 2006 Conference, 1–12, 2006.

[49] R. Fernandez, F. Manzo, F.R. Nardi, E. Scoppola. *Asymptotically exponential hitting times and metastability: a pathwise approach without reversibility*, Electronic Journal of Probability 20(122), 1–37, 2015.

[50] R. Fernandez, F. Manzo, F.R. Nardi, E. Scoppola, J. Sohier. *Conditioned, quasi-stationary, restricted measures and escape from metastable states*, Annals of Applied Probability 26(2), 760–793, 2016.

[51] M. Garetto, T. Salonidis, E.W. Knightly. *Modeling per-flow throughput and capturing starvation in CSMA multi-hop wireless networks*, IEEE/ACM Transactions on Networking 16(4), 864–877, 2008.

[52] A. Gaudillière, P. Milanesi, M.E. Vares. *Asymptotic exponential law for the transition time to equilibrium of the metastable kinetic Ising model with vanishing magnetic field*, Journal of Statistical Physics 179(2), 263–308, 2020.

[53] A. Gaudillière, E. Olivieri, E. Scoppola. *Nucleation pattern at low temperature for local Kawasaki dynamics in two dimensions*, Markov Processes and Related Fields 11(4), 553–628, 2005.

[54] J. Ghaderi, S.C. Borst, P.A. Whiting. *Queue-based random-access algorithms: fluid limits and stability issues*, Stochastic Systems 4, 81–156, 2014.

[55] J. Ghaderi, R. Srikant. *On the design of efficient CSMA algorithms for wireless networks*. In: Proceedings of the CDC 2010 Conference, 954–959, 2010.

[56] M. Grossglauser, D.N.C. Tse. *Mobility increases the capacity of ad hoc wireless networks*, IEEE/ACM Transactions on Networking 10(4), 477–486, 2002.

[57] F. den Hollander, O. Jovanovski. *Metastability on the hierarchical lattice*, Journal of Physics A: Theoretical and Mathematical 50(30), 305001, 2017.

[58] F. den Hollander, F.R. Nardi, E. Olivieri, E. Scoppola. *Droplet growth for three-dimensional Kawasaki dynamics*, Probability Theory and Related Fields 125(2), 153–194, 2003.

[59] F. den Hollander, F.R. Nardi, S. Taati. *Metastability of hard-core dynamics on bipartite graphs*. Electronic Journal of Probability 23(97), 1–65, 2018.

[60] F. den Hollander, F.R. Nardi, A. Troiani. *Kawasaki dynamics with two types of particles: stable/metastable configurations and communication heights*, Journal of Statistical Physics 145(6), 1423–1457, 2011.

[61] F. den Hollander, F.R. Nardi, A. Troiani. *Kawasaki dynamics with two types of particles: critical droplets*, Journal of Statistical Physics 149(6), 1013–1057, 2012.

[62] F. den Hollander, F.R. Nardi, A. Troiani. *Metastability for Kawasaki dynamics at low temperature with two types of particles*, Electronic Journal of Probability 17(2), 1–26, 2012.

[63] F. den Hollander, E. Olivieri, E. Scoppola. *Metastability and nucleation for conservative dynamics*, Journal of Mathematical Physics 41(3), 1424–1498, 2000.

[64] L. Jiang, D. Shah, J. Shin, J. Walrand. *Distributed random access algorithm: scheduling and congestion control*, IEEE Transactions on Information Theory 56, 6182–6207, 2010.

[65] L. Jiang, J. Walrand. *A distributed CSMA algorithm for throughput and utility maximization in wireless networks*, IEEE/ACM Transactions on Networking 18(3), 960–972, 2010.

[66] O. Jovanovski. *Metastability for the Ising model on the hypercube*, Journal of Statistical Physics 167(1), 135–159, 2017.

[67] J. Keilson. Markov Chain Models - Rarity and Exponentiality. Applied Mathematical Sciences 28, Springer, 1979.

[68] F.P. Kelly. *Stochastic models of computer-communication systems*, Journal of the Royal Statistical Society Series B 47(3), 379–395, 1985.

[69] A. Kershenbaum, R.R. Boorstyn, M.-S. Chen. *An algorithm for evaluating throughput in multihop packet radio networks with complex topologies*, IEEE Journal on Selected Areas in Communications 5(6), 1003–1012, 1987.

[70] H. Kesten, R.H. Schonmann. *On some growth models with a small parameter*, Probability Theory and Related Fields 101, 435–468, 1995.

[71] L. Kleinrock, F.A. Tobagi. *Packet switching in radio channels: Part I - Carrier Sense Multiple-Access modes and their throughput-delay characteristics*. IEEE Transactions on Communications 23(12), 1400–1416, 1975.

[72] R. Kotecký, E. Olivieri. *Shapes of growing droplets — A model of escape from a metastable phase*, Journal of Statistical Physics 75(3-4), 409–506, 1994.

[73] C. Landim, P. Lemire. *Metastability of the two-dimensional Blume-Capel model with zero chemical potential and small magnetic field*, Journal of Statistical Physics 164(2), 346–376, 2016.

[74] R. Laufer, L. Kleinrock. *On the capacity of wireless CSMA/CA multihop networks*. In: Proceedings of the IEEE INFOCOM 2013 Conference, 1312–1320, 2013.

[75] S.C. Liew, C.H. Kai, J. Leung, B. Wong. *Back-of-the-envelope computation of throughput distributions in CSMA wireless networks*. In: Proceedings of the ICC 2009 Conference, 1–19, 2009.

[76] M.R.H. Mandjes. Large deviations for Gaussian queues. Wiley, 2007.

[77] F. Manzo, F.R. Nardi, E. Olivieri, E. Scoppola. *On the essential features of metastability: Tunnelling time and critical configurations*, Journal of Statistical Physics 115(1/2), 591–642, 2004.

[78] F. Manzo, E. Olivieri. *Dynamical Blume–Capel model: competing metastable states at infinite volume*, Journal of Statistical Physics 104(5), 1029–1090, 2001.

[79] F. Manzo, E. Olivieri. *Relaxation patterns for competing metastable states: a nucleation and growth model*, Markov Processes and Related Fields 4(4), 549–570, 1998.

[80] L. Miclo. *Sur les problèmes de sortie discrets inhomogènes*, Annals of Applied Probability 6(4), 1112–1156, 1996.

[81] F.R. Nardi, C. Spitoni. *Sharp asymptotics for stochastic dynamics with parallel updating rule*, Journal of Statistical Physics 146(4), 701–718, 2012.

[82] F.R. Nardi, A. Zocca. *Tunneling behavior of Ising and Potts models on grid graphs*, Stochastic Processes and their Applications 129(11), 4556–4575, 2019.

[83] F.R. Nardi, A. Zocca, S.C. Borst. *Hitting time asymptotics for hard-core interactions on grids*, Journal of Statistical Physics 162, 522–576, 2016.

[84] E.J. Neves, R.H. Schonmann. *Critical droplets and metastability for a Glauber dynamics at very low temperatures*, Communications in Mathematical Physics 137(2), 209–230, 1991.

[85] J. Ni, B. Tan, R. Srikant. *Q-CSMA: queue-length based CSMA/CA algorithms for achieving maximum throughput and low delay in wireless networks*. In: Proceedings of the IEEE INFOCOM 2010 Mini-Conference, 1–5, 2010.

[86] E. Olivieri, M.E. Vares. Large Deviations and Metastability. Cambridge University Press, 2005.

[87] E. Pinsky, Y. Yemini. *The asymptotic analysis of some packet radio networks*, IEEE Journal on Selected Areas in Communications 4(6), 938–945, 1986.

[88] S. Rajagopalan, D. Shah, J. Shin. *Network adiabatic theorem: An efficient randomized protocol for contention resolution*, ACM SIGMETRICS Performance Evaluation Review 37, 133–144, 2009.

[89] T.S. Rappaport. Wireless Communications: Principles and Practice, second edition. Prentice Hall, 2002.

[90] L.G. Roberts. *ALOHA packet system with and without slots and capture*, ACM SIGCOMM Computer Communication Review 5(2), 28–42, 1975.

[91] R.H. Schonmann, S. Shlosman. *Wulff droplets and the metastable relaxation of kinetic Ising models*, Communications in Mathematical Physics 194(2), 389–462, 1998.

[92] M. Sfragara. *Introducing an edge dynamic in wireless random-access networks*. Preprint, 2020.

[93] D. Shah, J. Shin. *Delay-optimal queue-based CSMA*, ACM SIGMETRICS Performance Evaluation Review 38(1), 373–374, 2010.

[94] D. Shah, J. Shin. *Randomized scheduling algorithms for queueing networks*, Annals of Applied Probability 22, 128–171, 2012.

[95] A. Shwartz, A. Weiss. Large Deviations for Performance Analysis: Queues, Communications, and Computing. Chapman & Hall, 1995.

[96] F. Simatos, A. Simonian. *Mobility can drastically improve the heavy-traffic performance from* $1/(1-rho)$ *to* $-log(1-rho)$. [arXiv:1909.04383], 2019.

[97] V.G. Subramanian, M. Alanyali. *Delay performance of CSMA in networks with bounded degree conflict graphs*. In: Proceedings of the IEEE ISIT 2011 Conference, 2373–2377, 2011.

[98] D.N.C. Tse, P. Viswanath. Fundamentals of Wireless Communications. Cambridge University Press, 2005.

[99] N. van der Velden. A least degree randomized algorithm on a bipartite graph. B.Sc. thesis, Leiden University, 2019.

[100] P.M. van de Ven, S.C. Borst, J.S.H. van Leeuwaarden, A. Proutière. *Insensitivity and stability of random-access networks*, Performance Evaluation 67(11), 1230–1242, 2010.

[101] P. Viswanath, D.N.C. Tse, R. Laroia. *Opportunistic beamforming using dumb antennas*, IEEE Transactions on Information Theory 48(6), 1277–1294, 2002.

[102] X. Wang, K. Kar. *Throughput modeling and fairness issues in CSMA/CA based ad-hoc networks*, In: Proceedings of the IEEE INFOCOM 2005 Conference, 23–34, 2005.

[103] Y. Yemini. *A statistical mechanics of distributed resource sharing mechanisms.* In: Proceedings of the IEEE INFOCOM 1983 Conference, 531–539, 1983.

[104] A. Zocca. Spatio-temporal dynamics of random-access networks: An interacting-particle approach. Ph.D. thesis, Eindhoven University of Technology, 2015.

[105] A. Zocca. *Tunneling of the hard-core model on finite triangular lattices*, Random Structures and Algorithms 55; 215–246, 2017.

[106] A. Zocca. *Low-temperature behavior of the multicomponent Widom–Rowlison model on finite square lattices*, Journal of Statistical Physics 171, 1–37, 2018.

References of Part II

[107] O.H. Ajanki, L. Erdős, T. Kruger. *Universality for general Wigner-type matrices*, Probability Theory and Related Fields 169, 667–727, 2017.

[108] G.W. Anderson, A. Guionnet, O. Zeitouni. An Introduction to Random Matrices. Cambridge University Press, 2010.

[109] G.W. Anderson, O. Zeitouni. *A CLT for a band matrix model*, Probability Theory and Related Fields 134(2), 283–338, 2006.

[110] G.B. Arous, A. Dembo, A. Guionnet. *Aging of spherical spin glasses*, Probability Theory and Related Fields 120(1), 1–67, 2001.

[111] G.B. Arous, A. Guionnet. *Large deviations for Wigner's law and Voiculescu's non-commutative entropy*, Probability Theory and Related Fields 108(4), 517–542, 1997.

[112] F. Augeri. *Large deviations principle for the largest eigenvalue of Wigner matrices without Gaussian tails*, Electronic Journal of Probability 21(32), 1–49, 2016.

[113] F. Augeri, A. Guionnet, J. Husson. *Large deviations for the largest eigenvalue of sub-Gaussian matrices*. [arXiv:1911.10591], 2019.

[114] Z. Bai, J.W. Silverstein. Spectral analysis of large dimensional random matrices, second edition. Springer Series in Statistics, 2010.

[115] M. Bauer, O. Golinelli. *Random incidence matrices: moments of the spectral density*, Journal of Statistical Physics 103(1-2), 301–337, 2001.

[116] F. Benaych-Georges, C. Bordenave, A. Knowles. *Spectral radii of sparse random matrices*. [arXiv:1704.02945], 2017.

[117] F. Benaych-Georges, C. Bordenave, A. Knowles. *Largest eigenvalues of sparse inhomogeneous Erdős-Rënyi graphs*. Annals of Probability 47(3), 1653–1676, 2019.

[118] H. Bercovici, D. Voiculescu. *Free convolution of measures with unbounded support*. Indiana University Mathematics Journal 42, 733–773, 1993.

[119] S. Bhamidi, S.N. Evans, A. Sen. *Spectra of large random trees*, Journal of Theoretical Probability 25(3), 613–654, 2012.

[120] P. Biane. *On the free convolution with a semi-circular distribution*, Indiana University Mathematics Journal 46(3), 705–718, 1997.

[121] C. Bordenave, P. Caputo. *A large deviation principle for Wigner matrices without Gaussian tails*, Annals of Probability 42(6), 2454–2496, 2014.

[122] C. Bordenave, M. Lelarge. *Resolvent of large random graphs*, Random Structures & Algorithms 37(3), 332–352, 2010.

[123] C. Borgs, J.T. Chayes, H. Cohn, Y. Zhao. *An L^p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions*, Transactions of the American Mathematical Society 372, 3019–3062, 2019.

[124] C. Borgs, J.T. Chayes, H. Cohn, Y. Zhao. *An L^p theory of sparse graph convergence II: LD convergence, quotients and right convergence*, Annals of Probability 46(1), 337–396, 01 2018.

[125] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, K. Vesztergombi. *Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing*, Advances in Mathematics 219(6), 1801–1851, 2008.

[126] C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, K. Vesztergombi. *Convergent sequences of dense graphs II. Multiway cuts and statistical physics*, Annals of Mathematics 176(1), 151–219, 2012.

[127] P. Bourgade, L. Erdős, H.-T. Yau, J. Yin. *Fixed energy universality for generalized Wigner matrices* Communications on Pure and Applied Mathematics 69(10), 1815–1881, 2016.

[128] W. Bryc, A. Dembo, T. Jiang. *Spectral measure of large random Hankel, Markov and Toeplitz matrices*, Annals of Probability 34(1), 1–38, 2006.

[129] A. Chakrabarty. *The Hadamard product and the free convolutions*, Statistics and Probability Letters 127, 150–157, 2017.

[130] A. Chakrabarty, S. Chakraborty, R.S. Hazra. *Eigenvalues outside the bulk of inhomogeneous Erdős-Rényi random graphs.* [arXiv:1911.08244], 2019.

[131] A. Chakrabarty, R.S. Hazra. *Remarks on absolute continuity in the context of free probability and random matrices*, Proceedings of the American Mathematical Society 144, 1335–1441, 2016.

[132] A. Chakrabarty, R.S. Hazra, F. den Hollander, M. Sfragara. *Spectra of adjacency and Laplacian matrices of Inhomogeneous Erdős-Rényi random graphs.* Random Matrices: Theory and Applications, 2020.

[133] A. Chakrabarty, R.S. Hazra, F. den Hollander, M. Sfragara. *Large deviation principle for the maximal eigenvalue of inhomogeneous Erdős-Rényi random graphs.* [arXiv:2008.08367], 2020.

[134] A. Chakrabarty, R.S. Hazra, D. Sarkar. *Limiting spectral distribution for Wigner matrices with dependent entries*, Acta Physica Polonica B 46(9), 1637–1652, 2015.

[135] A. Chakrabarty, R.S. Hazra, D. Sarkar. *From random matrices to long range dependence*, Random Matrices: Theory and Applications 5(2), 1650008, 2016.

[136] S. Chatterjee. *A simple invariance theorem.* [arXiv:math/0508213v1], 2005.

[137] S. Chatterjee. Large deviations for random graphs. Lecture Notes in Mathematics 2197, École d'Été de Probabilités de Saint-Flour XLV - 2015, Springer, 2017.

[138] S. Chatterjee, R.S. Varadhan. *The large deviation principle for the Erdős-Rényi random graph*, European Journal of Combinatorics 32, 1000–1017, 2011.

[139] F.R.K. Chung. Spectral graph theory. CBMS Regional Conference Series in Mathematics 92, Conference Board of the Mathematical Sciences, 1997.

[140] F. Chung, 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, 2002.

[141] F. Chung, L. Lu, V. Vu. *Spectra of random graphs with given expected degrees*, Proceedings of the National Academy of Sciences 100(11), 6313–6318, 2003.

[142] D.M. Cvetković, M. Doob, H. Sachs. Spectra of Graphs, Theory and Applications. Academic Press, 1980.

[143] K. Deimling. Nonlinear Functional Analysis. Dover publications, 1985.

[144] A. Dembo, E. Lubetzky. *Empirical spectral distributions of sparse random graphs.* [arXiv:1610.05186], 2016.

[145] S. Dhara, S. Sen. *Large deviation for uniform graphs with given degrees.* [arXiv:1904.07666v4], 2020.

[146] X. Ding, T. Jiang. *Spectral distributions of adjacency and Laplacian matrices of random graphs*, Annals of Applied Probability 20(6), 2086–2117, 2010.

[147] M. Disertori, F. Merkl, S.W.W. Rolles. *Localization for a nonlinear sigma model in a strip related to vertex reinforced jump processes*, Communications in Mathematical Physics 332, 78–825, 2014.

[148] F.J. Dyson. *A Brownian-motion model for the eigenvalues of a random matrix*, Journal of Mathematical Physics 3, 1191–1198, 1962.

[149] F.J. Dyson. *Statistical theory of the energy levels of complex systems*, I, II, and III, Journal of Mathematical Physics 3, 140–156, 157–165, 166–175, 1962.

[150] I. Dumitriu, S. Pal. *Sparse regular random graphs: spectral density and eigenvectors*, Annals of Probability 40(5), 2197–2235, 2012.

[151] L. Erdős, A. Knowles, H.-T. Yau, J. Yin. *Spectral statistics of Erdős-Rényi graphs I: Local semicircle law*, Annals of Probability 41(3B), 2279–2375, 2013.

[152] L. Erdős, A. Knowles, H.-T. Yau, J. Yin. *Spectral statistics of Erdős-Rényi Graphs II: Eigenvalue spacing and the extreme eigenvalues*, Communications in Mathematical Physics 314(3), 587–640, 2012.

[153] L. Erdős, S. Péché, J.A. Ramírez, B. Schlein, H.-T. Yau. *Bulk universality for Wigner matrices*, Communications on Pure and Applied Mathematics 63(7), 895–925, 2010.

[154] L. Erdős, J.A. Ramírez, B. Schlein, T. Tao, V. Vu, H.-T. Yau. *Bulk universality for Wigner Hermitian matrices with subexponential decay*, Mathematical Research Letters 17(4), 667–674, 2010.

[155] L. Erdős, B. Schlein, H.-T. Yau. *Universality of random matrices and local relaxation flow*, Inventiones mathematicae 185(1), 75–119, 2011.

[156] L. Erdős, H.-T. Yau. *Gap universality of generalized Wigner and β-ensembles*, Journal of the European Mathematical Society 17(8), 1927–2036, 2015.

[157] L. Erdős, H.-T. Yau, J. Yin. *Bulk universality for generalized Wigner matrices*, Probability Theory and Related Fields 154(1-2), 1–67, 2012.

[158] P. Erdős, A. Rényi. *On random graphs I*, Publicationes Mathematicae 6, 290–297, 1959.

[159] I.J. Farkas, I. Derényi, A.L. Barabási, T. Vicsek. *Spectra of "real-world" graphs: Beyond the semicircle law*, Physical Review E 64(2), 026704, 2001.

[160] Z. Füredi, J. Komlós. *The eigenvalues of random symmetric matrices*, Combinatorica 1, 233–241, 1981.

[161] P. Frenkel. *Convergence of graphs with intermediate density*, Transactions of the American Mathematical Society 370(5), 3363–3404, 2018.

[162] M. Gaudin. *Sur la loi limite de l'espacement des valeurs propres d'une matrice aléatoire*, Nuclear Physics 25, 447–458, 1961.

[163] V. Girko, W. Kirsch, A. Kutzelnigg. *A necessary and sufficient conditions for the semicircle law*, Random Operators and Stochastic Equations 2(2), 195–202, 1994.

[164] W. Hachem, P. Loubaton, J. Najim. *Deterministic equivalents for certain functionals of large random matrices*, Annals of Applied Probability 17(3), 875–930, 2007.

[165] R. van der Hofstad. Random Graphs and Complex Networks, volume 1. Cambridge University Press, 2017.

[166] P.W. Holland, K.B. Laskey, S. Leinhardt. *Stochastic blockmodels: First steps*, Social networks 5(2), 109–137, 1983.

[167] F. den Hollander. Large Deviations. Fields Institute Monograph 14, American Mathematical Society, 2000.

[168] J. Huang, B. Landon, H.-T. Yau. *Bulk universality of sparse random matrices*, Journal of Mathematical Physics 56(12), 123301, 2015.

[169] E.T. Jaynes. *Information theory and statistical mechanics*, Physical Review, 106(4), 620–630, 1957.

[170] T. Jiang. *Empirical distributions of Laplacian matrices of large dilute random graphs*, Random Matrices: Theory and Applications 1(3), 1250004, 2012.

[171] T. Jiang. *Low eigenvalues of Laplacian matrices of large random graphs*, Probability Theory and Related Fields 153(3-4), 671–690, 2012.

[172] O. Khorunzhy, M. Shcherbina, V. Vengerovsky. *Eigenvalue distribution of large weighted random graphs*, Journal of Mathematical Physics 45(4), 1648–1672, 2004.

[173] M. Krivelevich, B. Sudakov. *The largest eigenvalue of sparse random graphs*, Combinatorics, Probability and Computing 12, 61–72, 2003.

[174] D. Kunszenti-Kovács, L. Lovász, B. Szegedy. *Measures on the square as sparse graph limits*, Journal of Combinatorial Theory Series B 138, 1–40, 2019.

[175] J.O. Lee, K. Schnelli. *Local law and Tracy-Widom limit for sparse random matrices*. Probability Theory and Related Fields 171, 543–616, 2018.

[176] L. Lovász. Large networks and graph limits. American Mathematical Society Colloquium Publications 60, American Mathematical Society, 2012.

[177] L. Lovász and B. Szegedy. *Limits of dense graph sequences*, Journal of Combinatorial Theory Series B 96(6), 933–957, 2006.

[178] L. Lovász and B. Szegedy. *Szemerédi's lemma for the analyst*, Geometric and Functional Analysis 17(1), 252–270, 2007.

[179] E. Lubetzky, Y. Zhao. *On replica symmetry of large deviations in random graphs*, Random Structures & Algorithms 47(1), 109–146, 2015.

[180] M.J.R. Markering. The large deviation principle for inhomogeneous Erdős-Rényi random graphs. B.sc. thesis, Leiden University, 2020.

[181] M.L. Mehta, M. Gaudin. *On the density of eigenvalues of a random matrix*, Nuclear Physics B 18, 420–427, 1960.

[182] J.A. Mingo, R. Speicher. Free Probability and Random Matrices. Springer, 2017.

[183] A. Nica, R. Speicher. Lectures on the Combinatorics of Free Probability. Cambridge University Press, 2006.

[184] F. Sauvigny. Partial Differential Equations 2: Functional Analytic Methods. Springer Science and Business Media, 2012.

[185] D. Shlyakhtenko. *Random Gaussian band matrices and freeness with amalgamation*, International Mathematics Research Notices 1996(20), 1013–1025, 1996.

[186] R. Speicher. *Free probability theory*. In: The Oxford handbook of random matrix theory, 452–470, Oxford University Press, 2011.

[187] T. Squartini, J. de Mol, F. den Hollander, D. Garlaschelli. *Breaking of ensemble equivalence in networks*, Physical Review Letters 115, 268701, 2015.

[188] E. Szemerédi. *Regular partitions of graphs*. In: Problèmes combinatoires et théorie des graphes (Colloq. Internat. CNRS, Univ. Orsay, Orsay, 1976), Colloques Internationaux CNRS 260, 399–401, 1978.

[189] T. Tao. Topics in random matrix theory. Graduate Studies in Mathematics 132, American Mathematical Society, 2012.

[190] T. Tao, V. Vu. *Random matrices: universality of local eigenvalue statistics*, Acta mathematica 206(1), 127–204, 2011.

[191] L.V. Tran, V.H. Vu, K. Wang. *Sparse random graphs: Eigenvalues and eigenvectors*, Random Structures & Algorithms 42(1), 110–134, 2013.

[192] D.V. Voiculescu. *Symmetries of some reduced free product C^*-algebras*. In: Operator Algebras and their Connections with Topology and Ergodic Theory, Proceedings of the OATE 1983 Conference, Lecture Notes in Mathematics 1132, 556–588, Springer, 1985.

[193] D.V. Voiculescu. *Limit laws for random matrices and free products* Inventiones mathematicae 104(1), 201–220, 1991.

[194] E.P. Wigner. *Characteristic vectors of bordered matrices with infinite dimensions*, Annals of Mathematics 62(3), 548–564, 1955.

[195] J. Wishart. *The generalised product moment distribution in samples from a normal multivariate population*, Biometrika 20A(1/2), 32–52, 1928.

[196] Y. Zhu. *Graphon approach to limiting spectral distributions of Wigner-type matrices*. [arXiv:1806.11246], 2018.

Summary

This thesis is divided into two parts. In Part I we study metastability properties of queue-based random-access protocols for wireless networks. The network is modeled as a bipartite graph whose edges represent interference constraints. In Part II we study spectra of inhomogeneous Erdős-Rényi random graphs. We focus in particular on the limiting spectral distribution of the adjacency and Laplacian matrices and on the largest eigenvalue of the adjacency matrix.

Part I

Random-access protocols have been introduced in the context of wireless networks with the aim to avoid collisions between ongoing transmissions of wireless signals. The main idea behind these protocols is to associate with each device a random clock that determines when the device attempts to transmit. Hence, devices decide autonomously when to start a transmission using only local information. We consider random-access networks where each node represents a server with a queue. The nodes can be either active or inactive: a node deactivates at unit rate, while it activates at a rate that depends on its queue length, provided none of its neighbors is active. Wireless random-access networks are known to exhibit metastability effects. In a regime where the activation rates become large, the stationary distribution of the joint activity process concentrates on states where the maximum number of nodes is active, with extremely slow transitions between them. Individual nodes may experience prolonged periods of starvation, resulting in severe build-up of queues and long delays. A deeper understanding of metastability properties is important for designing mechanisms to improve the network performance. We model the network as a bipartite graph and, in the limit as the queues become large, we study the transition time between the two states where one half of the network is active and the other half is inactive.

In Chapter 2 we consider complete bipartite graphs. We compare the transition time of an internal model in which the activation rates depend on the current queue lengths with that of an external model in which the activation rates depend on the current mean queue lengths. We define two perturbed models with externally driven activation rates that sandwich the queue lengths of the internal model and its transition time. We show with the help of coupling that with high probability the mean transition time and its distribution for the internal model are asymptotically the same as for the external model. The law of the transition time divided by its mean is exponential, truncated polynomial or deterministic, depending on the activation rate functions.

In Chapter 3 we consider arbitrary bipartite graphs. We decompose the transition into a succession of transitions on complete bipartite subgraphs. This succession

depends in a delicate manner on the full structure of the graph. We formulate a greedy algorithm to analyze the most likely transition paths between dominant states. By combining our results for complete bipartite graphs with a detailed analysis of the algorithm, we determine the mean transition time and its distribution along each path. Depending on the activation rate functions, we again identify three regimes of behavior.

In Chapter 4 we consider dynamic bipartite graphs. In order to try to capture the effects of user mobility in wireless networks, we analyze dynamic interference graphs where the edges are allowed to appear and disappear over time. A node can activate either when its neighbors are simultaneously inactive or when the edges connecting it with its neighbors disappear. Interpolation between these two situations gives rise to different scenarios and interesting behavior. We identify how the order of the mean transition time depends on the speed of the dynamics.

Part II

Eigenvalues play a central role in our understanding of graphs. Spectral graph theory studies the properties of eigenvalues and eigenvectors of the associated adjacency and Laplacian matrices. The eigenvalues of the adjacency matrix carry information about topological features of the graph, such as connectivity and subgraph counts. The eigenvalues of the Laplacian matrix carry information about random walks on the graph and allow us to analyze approximation algorithms. The standard Erdős-Rényi random graph model, which is the most basic model in random graph theory, is formed by connecting each pair of vertices with a certain fixed probability, independently of each other. The spectra of both the adjacency and Laplacian matrices are well studied and well understood. We consider inhomogeneous Erdős-Rényi random graphs, where each pair of vertices is connected with a certain probability that is not necessarily the same for all pairs, but still independently of each other.

In Chapter 5 we consider inhomogenerous Erdős-Rényi random graphs in the non-dense non-sparse regime, where the degrees of the vertices diverge sublinearly with the size of the graph. We study the limiting behavior of the emprical spectral distributions of the adjacency and Laplacian matrices. When the connection probabilities have a multiplicative structure, we give an explicit description of the scaling limits using tools from free probability theory. Furthermore, we apply our results to constrained random graphs, Chung-Lu random graphs and social networks.

In Chapter 6 we consider inhomogenerous Erdős-Rényi random graphs in the dense regime, where the degrees of the vertices are proportional to the size of the graph. Using the theory of graphons, we derive a large deviation principle for the largest eigenvalue and analyze the associated rate function in detail. Indeed, the structure of the graph conditional on a large deviation can be expressed in terms of a variational problem involving graphons. When the connection probabilities have a multiplicative structure, we analyze the variational formula in order to identify the scaling properties of the rate function.

Samenvatting

Dit proefschrift is opgedeeld in twee delen. In deel I bestuderen we de metastabiliteitseigenschappen van zogeheten "op wachtrij gebaseerde toegangsprotocollen" voor draadloze netwerken. Het netwerk wordt gemodelleerd door een bipartiete graaf waarvan de zijden van de graaf de storingsbeperkingen representeren. In deel II bestuderen we de spectra van inhomogene Erdős-Rényi random grafen. In het bijzonder bekijken we de verdeling van de spectra van de nabuurmatrix en de Laplace matrix. Daarnaast bestuderen we de grootste eigenwaarde van de nabuurmatrix.

Deel I

Toegangsprotocollen voor draadloze netwerken zijn ingevoerd om botsingen tussen draadloze signalen te voorkomen. Het belangrijkste idee achter deze protocollen is dat elk apparaat in het netwerk een willekeurige klok heeft die bepaalt wanneer het een poging doet om een signaal te zenden. Elk apparaat beslist dus zelf wanneer het start met het zenden van signalen en gebruikt daarbij alleen lokale informatie. We beschouwen netwerken met toevallige toegangsprotocollen, waarbij elke knoop een server met een wachtrij representeert. De knopen kunnen actief of inactief zijn: een actieve knoop wordt inactief met snelheidsparameter één, terwijl een inactieve knoop geactiveerd wordt met een snelheidsparameter die afhangt van de lengte van de wachtrij, indien geen van de naburige knopen actief is. Van netwerken met draadloze toegangsprotocollen is bekend dat zij metastabiliteit vertonen. In een regiem waarbij de activatieparameters groot zijn, concentreert de evenwichtsverdeling van het gezamenlijke activiteitsproces zich op toestanden waarin een maximaal aantal knopen actief is. Na zeer lange tijden kan een overgang tussen twee van zulke toestanden plaatsvinden. Individuele knopen kunnen lang gedeactiveerd zijn, wat resulteert in lange wachtrijen en grote vertragingen. Om mechanismen te ontwerpen die het functioneren van draadloze netwerken verbeteren, is een goed begrip van de metastabiliteitseigenschappen belangrijk. We modelleren het netwerk als een bipartiete graaf. In de limiet dat de wachtrijen lang worden, bestuderen we de transitietijden tussen de toestanden waarbij enkel de helft van het netwerk actief is en de andere helft inactief is.

In hoofdstuk 2 beschouwen we complete bipartiete grafen. We vergelijken de transitietijden van een "intern" model, waarbij de activatieparameters afhangen van de lengte van de wachtrijen op dat moment, met de transitietijden van een "extern" model, waarbij de activatieparameters afhangen van de gemiddelde lengte van de wachtrijen op dat moment. We definiëren twee gepertubeerde externe modellen met activatieparameters die de activatieparameters van het interne model insluiten. Met behulp van koppelingstechnieken laten we zien dat de kans groot is dat de gemiddelde transitietijden en de verdeling van het interne model asymptotisch gelijk zijn aan die

van het externe model. Afhankelijk van de activatieparameters is de verdeling van de transitietijden, gedeeld door de gemiddelde transitie tijd, exponentieel, afgekapt polynomiaal of deterministisch.

In hoofdstuk 3 beschouwen we willekeurige bipartiete grafen. We beschrijven de transitie van de graaf als een opeenvolging van transities van complete bipartiete deelgrafen. Deze opeenvolging hangt op een subtiele manier af van de volledige structuur van de graaf. We formuleren een "greedy algorithmöm de meest waarschijnlijke transitiepaden tussen dominante toestanden te analyseren. Door de resultaten voor complete bipartiete grafen te combineren met een gedetailleerde analyse van het algoritme, bepalen we voor elk pad de gemiddelde transitietijd en de verdeling van de transitietijd. Afhankelijk van de activatieparameters kunnen we opnieuw het gedrag van de transitietijden in drie regimes indelen.

In hoofdstuk 4 beschouwen we dynamische bipartiete grafen. Om het effect van bewegende gebruikers in een draadloos netwerk te begrijpen, analyseren we dynamische storingsgrafen waarbij de randen kunnen verschijnen en verdwijnen over de tijd. Een knoop kan activeren als de naburige knopen allemaal inactief zijn of als de verbindingszijde met naburige knopen verdwijnt. Wanneer we interpoleren tussen deze twee situaties komen er verschillende scenario's en interessant gedrag naar voren. We identificeren hoe de orde van grootte van de gemiddelde transitietijd afhangt van de snelheid van de dynamica.

Deel II

Eigenwaarden spelen een belangrijke rol in de analyse van grafen. Spectrale grafentheorie bestudeert de eigenschappen van eigenwaarden en eigenvectoren van de bijbehorende nabuurmatrices en Laplace matrices. De eigenwaarden van de nabuurmatrix bevatten informatie over de topologische eigenschappen van de graaf, zoals de verbondenheid en het aantal deelgrafen met een bepaalde eigenschap. De eigenwaarden van de Laplace matrix bevatten informatie over toevalswandelingen op de graaf en stellen ons in staat om benaderingsalgoritmen te analyseren. In het standaard Erdős-Rényi model, dat het meest elementaire model in de theorie van toevallige grafen is, zijn elk paar knopen met een vaste kans met elkaar te verbonden. De spectra van zowel de nabuurmatrices als van de Laplace matrices zijn reeds in detail bestudeerd en worden goed begrepen. We beschouwen inhomogene Erdős-Rényi random grafen, waar elk paar knopen verbonden is met een zekere kans die niet voor elk paar knopen hetzelfde is, maar waarbij de kansen nog steeds onafhankelijk van elkaar zijn.

In hoofdstuk 5 beschouwen we inhomogene Erdős-Rényi random grafen in het niet-dichte en niet-dunne regime, waarbij de graad van de knopen sublineair groeit met de grootte van de graaf. We bestuderen het limietgedrag van de empirische spectrale verdeling van de nabuurmatrix en Laplace matrix. Wanneer de verbindingskansen een multiplicatieve structuur hebben, dan geven we een expliciete beschrijving van de schalingslimieten met behulp van technieken uit de theorie van vrije kansrekening. Daarnaast passen we onze resultaten toe op grafen met restricties, Chung-Lu grafen en sociale netwerken.

In hoofdstuk 6 beschouwen we inhomogene Erdős-Rényi random grafen in het dichte regime, waar het aantal verbindingen van de knopen lineair groeit met de grootte van de graaf. Met behulp van grafons leiden we een grote-afwijkingen-principe af voor de grootste eigenwaarde, en analyseren we de bijbehorende entropiefunctie in detail. De structuur van de graaf geconditioneerd op een grote afwijking kan uitgedrukt worden in termen van een variationeel probleem voor grafons. Wanneer de verbindingskansen een multiplicatieve structuur hebben, dan analyseren we dit variationeel probleem om de schalingseigenschappen van de entropiefunctie te bepalen.

Acknowledgements

This thesis is the outcome of four years of PhD research. It could not have been completed without the help and support of many people to whom I would like to express my gratitude.

To begin with, I would like to thank my supervisors Frank, Sem and Francesca. They have guided me through my research, allowing me to grow as a mathematician and a scientist. Frank has been of incredible support with his positivity in every situation, his excitement for new ideas and his brilliant mind. I enjoyed very much working with him and sharing experiences together, also outside the work enviroment. Sem has been a source of many fruitful discussions and stimulating questions. I admire his critical thinking and his ability to detect the key issue where no one could see it. Francesca has always been available, even if based in another country. I enjoyed a lot the various weeks spent in Florence working with her and learning about academic life in Italy. I really appreciated the scientific freedom they all gave me, as well as the precious advices on pursuing my career.

I would also like to thank Rajat and Arijit, with whom I collaborated during my internship at the Indian Statistical Institute in Kolkata, and whom I visited three times in India. Together with them and Frank, we carried on a very interesting research project, which materialized into the second part of this thesis. I am very glad I had the opportunity to meet such great people and mathematicians, who helped me a lot integrating into the Indian culture and learning about new topics in mathematics.

I would like to thank the NETWORKS program, which funded my research, for giving me the possibility to travel to many places in the Netherlands and abroad. I really enjoyed the open-minded atmosphere, the multicultural collaborative environment of the training weeks and the many other meetings. A sincere thanks goes to all the current and former members of the Probability Theory group at the Mathematical Institute of Leiden University for the nice atmosphere they created and for all the interesting discussions we had in these years.

Besides mathematics, in the past four years I have been practising many different activities and hobbies. I would like to thank all the wonderful people I met through playing football, music, events and trips. Unfortunately, I cannot name them all, but I would like to mention some of them. A huge thanks goes to my “Italian family" in the Netherlands, among which Alessandro, Davide, Francesco, Gianluca, Giovanni, Matteo, Niccolò, Nicolò and Pierfrancesco. Their friendship and support made me miss our country less and continuously feel like home. I would also like to thank Miguel, Laura and Pedro, for all the beautiful moments shared together. A sincere thanks goes in general to all my friends who have always been there for me during these years. Their constant presence in my life is what gives me the strength to pursue

my goals and follow my passion.

The last two years have been in particular characterized by the indispensable affection of Fabiana. I woud like to thank her for always being such an inspiration to me through her determination, ambition, honesty and ethics. She helped me discovering my true self and finding stability in life. I am glad to share this adventure with her and I hope many more will follow.

Last but not least, I want to deeply thank my brother Lorenzo and my parents Ignazio and Patrizia. They are my family and they will always be by my side. I thank them for always helping me taking the right decisions, for always pushing me to be a better person and for their never-ending support.

Curriculum Vitae

Matteo Sfragara was born on February 20th 1991 in Pescara, Italy. He grew up in Verona, where he attended the Scientific High School Galileo Galilei and obtained the final diploma in 2010.

He then moved to Padova to study at the University of Padova. There he obtained a Bachelor of Science in Mathematics in 2013 and a Master of Science in Mathematics in 2015. During the first year of his master studies he joined the Erasmus Program and lived in Stockholm (Sweden), where he attended various courses both at Stockholm University and at KTH Royal Institute of Technology. During the last semester of his master studies he lived in Sydney (Australia), where he joined the Practicum Program at the University of New South Wales. There he wrote his master thesis entitled "The switch Markov chain for sampling irregular directed graphs", under the supervision of prof. dr. Catherine Greenhill. This research experience motivated him to pursue a PhD in Mathematics.

In September 2016 he moved to Leiden to start his PhD as a member of the probability group at the Mathematical Institute of Leiden University. His research was part of the Gravitation Program NETWORKS funded by the Dutch Ministry of Education, Culture and Science. He worked under the supervision of prof. dr. W.T.F. den Hollander (Leiden University), prof. dr. S.C. Borst (Eindhoven University of Technology) and dr. F.R. Nardi (University of Florence). As part of his PhD, he did an internship at the Indian Statistical Institute in Kolkata (India), where he collaborated with dr. R.S. Hazra and dr. A. Chakrabarty. He visited India three times.

During his PhD, he cultivated his passion for Mathematics in different ways. He supervised two pairs of high-school students attending Leiden University's Pre-University College. The first pair was awarded the Jan Kijne-onderzoeksprijs for best research. He also supervised a bachelor thesis project at the Mathematical Institute. He served as teaching assistant for various courses in Leiden and The Hague. He attended workshops, conferences and summer schools, and presented his research on several occasions in France, India, Italy, the Netherlands and Sweden.