Wireless random-access networks and spectra of random graphs

Sfragara, M.

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.

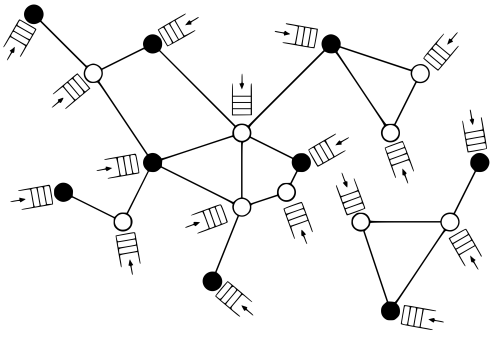

Figure 1.1: The interference graph of a random-access network. Each node represents a communication link between a transmitter and a receiver (active nodes are in black, while inactive nodes are in white). Packets arrive to the transmitter and form a queue. The wireless signals interfere with each other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

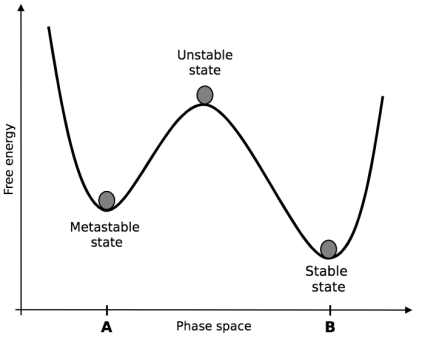

Figure 1.2: The evolution of the system from a metastable state A (local minimum), through an unstable state, to a stable state B (global minimum).

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]).

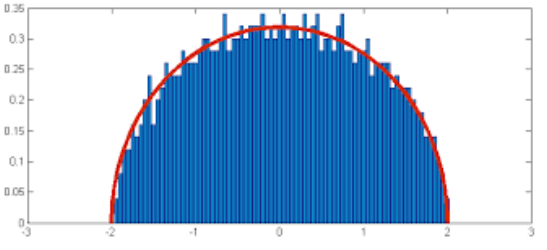

Figure 1.3: Wigner semicircle law.

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.

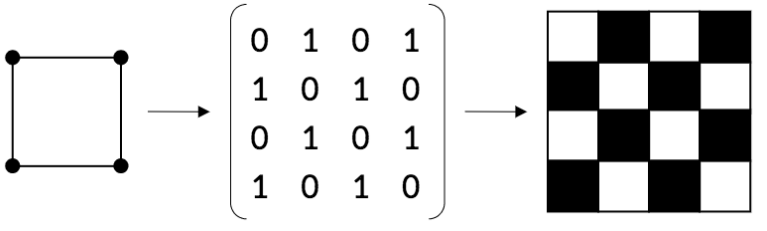

Figure 1.4: Graphon representation of a graph.

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].