Topics in random graphs, combinatorial optimization, and statistical inference

Marc Lelarge

Topics in random graphs, combinatorial optimization, and statistical inference

par

Marc Lelarge

Mémoire pour l'obtention de

l'Habilitation à Diriger des Recherches

de l'École normale supérieure

(Spécialité Informatique)

Soutenue le 23 février 2015 devant le jury composé de:

M. Venkatachalam Anantharam, *Professeur*
M. François Baccelli, *Directeur de Recherche*
M. Djalil Chafai, *Professeur*
M. Francis Comets, *Professeur*
M. Bruce Hajek, *Professeur*
M. Remco Van Der Hofstad, *Professeur*
M. Philippe Robert, *Directeur de Recherche*

au vu des rapports de:

M. Francis Comets, *Professeur*
M. Bruce Hajek, *Professeur*
M. Remco Van Der Hofstad, *Professeur*

Contents

Introduction

"The accreditation to conduct research acknowledges the high scientific level of the candidate, the originality of his approach in a scientific field, his ability to master a strategy in a sufficiently wide academic field and his capability to supervise younger researchers. [...]"

Extracted from the ministerial decree of November 23d, 1998, about the accreditation to conduct research (habilitation à diriger des recherches).
source: `http://www.ens.fr/recherche/hdr/?lang=en`

This Habilitation Thesis manuscript does NOT summarize my main research activities since the end of my Ph.D. Instead, I decided to use this introduction to briefly describe the research contributions of my work (articles are numbered according to my publication list at the end of the introduction). After this short self-promotion, the rest of the manuscript is made of three chapters presenting three different topics on which I worked with Ph.D. students. Each chapter can be read independently of the others and should be relatively self-contained. Chapter 1 is a gentle introduction to the theory of random graphs with an emphasis on contagions on such networks. In Chapter 2, I explain the main ideas of the objective method developed by Aldous and Steele [6] applied to the spectral measure of random graphs and the monomer-dimer problem. This topic is dear to me and I hope that this chapter will convince the reader that it is an exciting field of research. Chapter 3 deals with problems in high-dimensional statistics which now occupy a large proportion of my time. Unlike Chapters 1 and 2 which could be easily extended in lecture notes, I felt that the material in Chapter 3 was not ready for such a treatment. This field of research is currently very active and I decided to present two of my recent contributions.

My research interests are in network science and the theory of communications with emphasis on statistics, mathematical modeling and algorithm design. My main theoretical contributions are grounded in applications. Understanding tools developed by various communities (computer science, theoretical physics or machine learning) allowed me to attack pure mathematical problems with original ideas and to solve mathematical conjectures out of reach with the standard tools developed in the random graphs literature. In the other direction, a mathematical understanding of heuristic methods such as the cavity method in statistical physics or belief propagation in machine learning allowed me to design new algorithms with optimal performances for load balancing or statistical inference in networks. I strongly believe in the dialog between applied and theoretical scientific disciplines.

Main Scientific Contributions

(a) Large deviations with applications to communication networks (2001-2006) Motivated by statistical evidences for the presence of files with heavy-tailed distributions on the Internet, I developed with my Ph.D. advisor François Baccelli (INRIA, now UT Austin) and Serguei Foss (Heriot-Watt University) a new theory of large deviations for heavy-tailed (subexponential) distributions. Combined with an algebraic representation of networks, we obtained the first large deviations results for non-markovian networks in their stationary regime for heavy-tailed distributions in [J23],[J21]. I then developed a technique to deal with light-tailed distributions in [J19], [J17]. I extended these results [J15] to design control algorithms like in [J18] where an optimal load balancing rule is designed in order to minimize resequencing delays.

(b) Streaming algorithms with security constraints (2004-2006) In applications such as network monitoring, telecommunications data management or clickstream monitoring, data takes the form of continuous data streams rather than finite stored data sets, and clients require long-running continuous queries as opposed to one-time queries. During my postdoc at IBM, I developed with Zhen Liu (now Microsoft) and Anton Riabov (IBM) algorithms to schedule tasks on streams with security constraints [C31], which are now part of a product of IBM called System S.
(c) Economics of security in networks (2005-2009) I started this applied program in a collaboration with Jean Bolot (SPRINT now Technicolor). We developed a very general model to analyze the economics of epidemic processes, by overlaying an economic framework on a general epidemic propagation model. Using a local mean-field approach, we solved the model, and derived insight in a specific problem, namely the economic aspects of security investments in networks subject to virus and botnet attacks [C26], [C24]. We then considered the problem of designing incentives to entities in the Internet so they invest in security at a socially efficient level [C20], [J10], and find that insurance is a powerful such incentive [C27], [C21].
(d) Combinatorial optimization over random data and its scaling exponents (2006-2009) I started this theoretical program under the guidance of David Aldous (UC Berkeley), working with Charles Bordenave (postdoc, now with CNRS) on combinatorial optimization over random data and its scaling exponents. We considered combinatorial optimization problems defined over random ensembles, and studied how solution cost increases when the optimal solution undergoes a small perturbation. Note that Freshman calculus tells us that for a smooth cost function f with x^* being its minimum, we have $\epsilon(\delta) = \min\{f(x) - f(x^*),\ |x - x^*| \geq \delta\}$ that scales like δ^2. For the minimum spanning tree and for a simple example of an algorithmic problem solvable by dynamic programming, we were able to show in [J16], [J14], that the increase in cost scales also as δ^2, corresponding to a scaling exponent of 2. For a combinatorial optimization problem, a larger exponent means that there are more near-optimal solutions, suggesting that the algorithmic problem of finding the optimal solution is intrinsically harder. So scaling exponents may serve to separate combinatorial optimization problems of an appropriate type into a small set of classes of increasing difficulty. For instance, the minimum matching problem is expected to have scaling exponent 3, and thus be in the same class as TSP in a quantitative way, as distinct from their qualitative similarity as NP-complete problems under worst-case inputs. This parallels the notion of universality class in statistical physics. A rigorous proof of a scaling exponent of 3 is still an open problem.
(e) Contagion in social networks (2008-2014) In [J11], I analyzed the spread of new behaviors in social networks modeled by random graphs. Building on the work of S. Janson, I showed a new phase transition: when the network is sparse, the contagion is limited by its low global connectivity; when it is dense, the contagion is limited by the stability of the high-degree nodes. With Emilie Coupechoux (Ph.D. student), we extended these results to analyze the impact of clustering (i.e. presence of triangles) [J4] and [J1].
(f) Distances in weighted random graphs (2010-2014) With Moez Draief (Imperial college) and Hamed Amini (Ph.D. student), we studied the impact of random exponential edge weights on the distances in a random graph [J8] and, in particular, on its diameter. Our main result consists of a precise asymptotic expression for the maximal weight of the shortest weight paths between all vertices (the weighted diameter) of sparse random graphs [J2].
(g) Matching, load balancing and caching (2008-) Leveraging my expertise in the techniques developed by D. Aldous and M. Steele, called objective method or local weak convergence of graphs, I was able to solve an open problem in the random graphs literature with C. Bordenave and J. Salez (disproving a conjecture of N. Wormald): we showed in [J7] that the size of a maximum matching depends only upon local statistics of the graph and we were able to compute explicitly its value for large random graphs [C14]. With Laurent Massoulié (Technicolor now INRIA-MSR) and Mathieu Leconte (Ph.D student), we extended these results [C8] and showed how they can be used for caching in Content Distribution Networks [C12]. In another direction, with Florian Bourse (ENS student) and Milan Vojnovic (MSR), we looked at the balanced edge partition problem for the purpose of scaling out parallel computations and we demonstrate efficiency of natural greedy online assignments in [C2].
(h) Spectra of random graphs (2007-) With Charles Bordenave, we discovered that the local weak convergence gives new results for the convergence of the spectrum of large random graphs [J13] (i.e. spectrum of the associated symmetric adjacency matrix). Our work combined with ideas from random matrix theory and analysis of algorithms, allowed us with Justin Salez (Ph.D. student) to compute in

[J12] the rank of such graphs (answering an open question of K. Costello and V. Vu). This is one of the very few quantitative results for diluted random matrices with non i.i.d. entries. Wit Laurent Massoulié, we are now leveraging our expertise in random matrix theory to study weak Ramanujan property of random graphs with applications to community detection.

(i) High-dimensional statistical inference (2011-) I started this new line of research while visiting Andrea Montanari (Stanford). Using approximate message passing algorithms, we proved the universality of a certain phase transition arising in polytope geometry and compressed sensing [J3]. This solved a conjecture by David Donoho and Jared Tanner. More recently, with Laurent Massoulié and Jiaming Xu (Ph.D. UIUC), I obtained new results for community detection using spectral methods, first for a labeled stochastic bloc model [C6] and then for a more general latent variable model [C3]. With Se-Young Yun (postdoc) and Alexandre Proutière (KTH), we designed a streaming and memory limited algorithm for community detection in [C1] with guarantees that outperform all existing schemes.

PUBLICATIONS

Refereed Journal Publications

[J1] Emilie Coupechoux, Marc Lelarge, Contagions in Random Networks with Overlapping Communities, *Applied Probability Trust*, under revision.

[J2] Hamed Amini, Marc Lelarge, The Diameter of Weighted Random Graphs, *Annals of Applied Probability*, to appear.

[J3] Mohsen Bayati, Marc Lelarge, Andrea Montanari, Universality in Polytope Phase Transitions and Message Passing Algorithms, *Annals of Applied Probability* to appear.

[J4] Emilie Coupechoux, Marc Lelarge, How Clustering Affects Epidemics in Random Networks, *Advances in Applied Probability*, 46(4), 2014.

[J5] Marc Lelarge, Hang Zhou, Sublinear-time algorithms for monomer-dimer systems on bounded degree graphs, *Theoretical Computer Science*, 548:68-78, 2014.

[J6] Marc Lelarge, Antonella Villani. Exact tail asymptotics of a queue with LRD input traffic, *Journal of Mathematical Sciences*, 196(1):57-69, 2014.

[J7] Charles Bordenave, Marc Lelarge, Justin Salez, Matchings on infinite graphs, *Probability Theory and Related Fields*, 157(1-2):183-208, 2013.

[J8] Hamed Amini, Moez Draief, Marc Lelarge, Distances in Weighted Sparse Random Graphs, *SIAM Journal on Discrete Mathematics*, 27(1):1-26, 2013.

[J9] Hamed Amini, Marc Lelarge, Upper deviations for split times of branching processes, *Journal of Applied Probability*, 49(4):1134-1143, 2012.

[J10] Marc Lelarge, Coordination in Network Security Games: a Monotone Comparative Static Approach, *IEEE Journal on Selected Areas in Communications*, 30(11):2210-2219, 2012.

[J11] Marc Lelarge, Diffusion and Cascading Behavior in Random Networks, *Games and Economic Behavior*, 75(2):752-775, 2012.

[J12] Charles Bordenave, Marc Lelarge, Justin Salez, The rank of diluted random graphs, *Annals of Probability*, 39(3):1097-1121, 2011.

[J13] Charles Bordenave, Marc Lelarge, Resolvent of Large Random Graphs, *Random Structures and Algorithms*, 37(3):322-352, 2010.

[J14] David J. Aldous, Charles Bordenave, Marc Lelarge, Dynamic Programming Optimization over Random Data: the Scaling Exponent for Near-optimal Solutions, *SIAM Journal on Computing*, 38(6):2382-2410, 2009.

[J15] Marc Lelarge, Asymptotic behavior of Generalized Processor Sharing queues under subexponential assumptions, *Queueing Systems. Theory and Applications* 62(1):51-73, 2009.

[J16] David J. Aldous, Charles Bordenave, Marc Lelarge, Near-Minimal Spanning Trees: a Scaling Exponent in Probability Models, *Annales de l'Institut Henri Poincaré (B)* 44(5):962-976, 2008.

[J17] Marc Lelarge, Tail asymptotics for Discrete Event Systems, *Discrete Event Dynamic Systems* 18(4): 563-584, 2008.

[J18] Marc Lelarge, Packet Reordering in Networks with Heavy-Tailed Delays, *Mathematical Methods of Operations Research* 67(2):341-371, 2008.

[J19] Marc Lelarge, Tail asymptotics for monotone-separable networks, *Journal of Applied Probability* 44(2):306-320, 2007.

[J20] Ton Dieker, Marc Lelarge, Tails for (max,plus) recursions under subexponentiality, *Queueing Systems. Theory and Applications* 53(4):213-230, 2006.

[J21] François Baccelli, Serguei Foss, Marc Lelarge, Tails in Generalized Jackson Networks with Subexponential Service Time Distributions, *Journal of Applied Probability* 42(2):513-530, 2005.

[J22] Marc Lelarge, Fluid limit of generalized Jackson queueing networks with stationary and ergodic arrivals and service times, *Journal of Applied Probability* 42(2):491-512, 2005.

[J23] François Baccelli, Marc Lelarge, Serguei Foss, Asymptotics of Subexponential Max Plus Networks the Stochastic Event Graph Case, *Queueing Systems. Theory and Applications* 46(1-2):75-96, 2004

Refereed Conference and Workshop Publications

[C1] Se-Young Yun, Marc Lelarge, Alexandre Proutière, Streaming, Memory Limited Algorithms for Community Detection, *NIPS* 2014.

[C2] Florian Bourse, Marc Lelarge, Milan Vojnovic, Balanced Graph Edge Partition, *KDD* 2014: 1456-1465.

[C3] Jiaming Xu, Marc Lelarge, Laurent Massoulié, Edge Label Inference in Generalized Stochastic Block Models: from Spectral Theory to Impossibility Results, *COLT* 2014: 903-920.

[C4] Marc Lelarge, Hang Zhou, Sublinear-Time Algorithms for Monomer-Dimer Systems on Bounded Degree Graphs, *ISAAC* 2013: 141-151.

[C5] Marc Lelarge, Alexandre Proutière, Sadegh Talebi, Spectrum Bandit Optimization, *ITW* 2013.

[C6] Marc Lelarge, Laurent Massoulié, Jiaming Xu, Reconstruction in the Labeled Stochastic Block Model, *ITW* 2013.

[C7] Marc Lelarge, Bypassing Correlation Decay for Matchings with an Application to XORSAT, *ITW* 2013.

[C8] Mathieu Leconte, Marc Lelarge, Laurent Massoulié, Convergence of multivariate belief propagation, with applications to cuckoo hashing and load balancing, *ACM-SIAM SODA*, 2013: 35-46.

[C9] Simon Heimlicher, Marc Lelarge, Laurent Massoulié, Community Detection in the Labelled Stochastic Block Model, *NIPS 2012 Workshop: Algorithmic and Statistical Approaches for Large Social Networks.*

[C10] Stéphane Caron, Branislav Kveton, Marc Lelarge, Smriti Bhagat, Leveraging Side Observations in Stochastic Bandits, *UAI*, 2012: 142-151.

[C11] Mohsen Bayati, Marc Lelarge, Andrea Montanari, Universality in Polytope Phase Transitions and Iterative Algorithms, *ISIT*, 2012: 1643-1647.

[C12] Mathieu Leconte, Marc Lelarge, Laurent Massoulié, Bipartite graph structures for efficient balancing of heterogeneous loads, *ACM SIGMETRICS*, 2012: 41-52.

[C13] Marc Lelarge, Coordination in Network Security Games, *IEEE INFOCOM*, 2012: 2856-2860.

[C14] Marc Lelarge, A New Approach to the Orientation of Random Hypergraphs, *ACM-SIAM SODA*, 2012: 251-264.

[C15] Emilie Coupechoux, Marc Lelarge, Impact of Clustering on Diffusions and Contagions in Random Networks, *NetGCooP*, 2011: 1-7.

[C16] Marc Lelarge, Diffusion and Cascading Behavior in Random Networks, *MAMA workshop at Sigmetrics 2011* and *Interdisciplinary Workshop on Information and Decision in Social Networks*, MIT, 2011.

[C17] Hamed Amini, Moez Draief, Marc Lelarge, Flooding in Weighted Random Graphs. *SIAM ANALCO*, 2011: 1-15.

[C18] Charles Bordenave, Marc Lelarge, The Rank of Diluted Graphs, *ACM-SIAM SODA*, 2010: 1389-1402.

[C19] Marc Lelarge, Efficient Control of Epidemics over Random Networks, *ACM SIGMETRICS*, 2009: 1-12.

[C20] Marc Lelarge, Economics of Malware: Epidemic Risks Model, Network Externalities and Incentives, *Allerton Conference* 1353-1360, and *WEIS*, 2009.

[C21] Marc Lelarge, Jean Bolot, Economic Incentives to Increase Security in the Internet: The Case for Insurance, *IEEE INFOCOM*, 2009: 1494-1502.

[C22] Hamed Amini, Moez Draief, Marc Lelarge, Marketing in Random Networks, *NetCoop08*, LNCS 5425, 2009: 17-25.

[C23] Marc Lelarge, Diffusion of Innovations on Random Networks: Understanding the Chasm, *WINE*, LNCS 5385, 2008: 178-185.

[C24] Marc Lelarge, Jean Bolot, A Local Mean Field Analysis of Security Investments in Networks, ACM SIGCOMM *NetEcon*, 2008: 25-30.

[C25] Marc Lelarge, Jean Bolot, Cyber Insurance as an Incentive for Internet Security, *WEIS*, 2008.

[C26] Marc Lelarge, Jean Bolot, Network Externalities and the Deployment of Security Features and Protocols in the Internet, *ACM SIGMETRICS*, 2008: 37-48.

[C27] Jean Bolot, Marc Lelarge, A New Perspective on Internet Security using Insurance, *IEEE INFOCOM*, 2008: 1948-1956.

[C28] Marc Lelarge, Online Bandwidth packing with symmetric distributions, *Conference on Analysis of Algorithms*, DMTCS 2007: 471-482.

[C29] Cathy H. Xia, Zhen Liu, Don Towsley, Marc Lelarge, Scalability of Fork/Join queueing networks with blocking, *ACM SIGMETRICS*, 2007: 133-144.

[C30] Marc Lelarge, Tail asymptotics for Discrete Event Systems, *ValueTools*, 2006.

[C31] Marc Lelarge, Zhen Liu, Anton Riabov, Automatic Composition of Secure Workflow, *Automatic and Trusted Computing*, LNCS 4158, 2006: 322-331.

[C32] François Baccelli, Marc Lelarge, David McDonald, Metastable Regimes for multiplexed TCP flows, *Allerton conference*, 2004.

[C33] Marc Lelarge, Zhen Liu and Cathy H. Xia, Asymptotic Tail Distribution of End-to-End Delay in Networks of Queues with Self-Similar Cross Traffic, *IEEE INFOCOM*, 2004: 2352-2363.

Chapter 1

Percolation in random graphs

In this first chapter, we start with an introduction to the theory of random graphs. We first define the configuration model denoted by $G(n, \boldsymbol{d})$ in Section 1.1, where n is the number of nodes, while $\boldsymbol{d} = (d_1, \ldots, d_n)$ is the sequence of degrees. This model is a very natural probabilistic model as $G(n, \boldsymbol{d})$ will have the uniform distribution among all graphs with degree sequence given by $\boldsymbol{d}$. The sequence $\boldsymbol{d}$ is called graphic iff there is a simple graph (without loops and multiple edges) that has $\boldsymbol{d}$ as degree sequence. Erdős and Gallai [49] gave a necessary and sufficient condition for a sequence to be graphic (see also [92]): $\sum_{i=1}^{n} d_i$ is even and

$$\sum_{i=1}^{k} d_i \leq k(k-1) + \sum_{j=k+1}^{n} \min(k, d_j), \text{ for each } k \leq n.$$

However it is always possible to construct a multigraph, i.e. a graph having possibly self-loops and multiple edges between pairs of vertices, provided $\sum_{i=1}^{n} d_i$ is even. Hence for $\boldsymbol{d}$ such that $\sum_{i=1}^{n} d_i$ is even, we construct the multigraph $G^*(n, \boldsymbol{d})$ by pairing the half-edges uniformly at random (see the formal definition below). Then if $\boldsymbol{d}$ is graphic, the law of the multigraph $G^*(n, \boldsymbol{d})$ conditioned on being simple is exactly $G(n, \boldsymbol{d})$. In the sequel, we always assume that $\sum_{i=1}^{n} d_i$ is even and for each statement about $G(n, \boldsymbol{d})$, we also tacitly assume that $\boldsymbol{d}$ is graphic.

In this chapter, we are interested in questions concerning the size and structure of the largest component in a random graph when the number of vertices n tends to infinity. We consider a sequence of degree sequences $\boldsymbol{d}^{(n)}$ and the associated sequence of random graphs $G^*(n, \boldsymbol{d}^{(n)})$ and $G(n, \boldsymbol{d}^{(n)})$. These questions have attracted a lot of attention and there are by now quite a number of textbooks studying random graphs [24], [61]. For a modern treatment, we strongly recommend the lecture notes [95] (note that my opinion might be biased by the fact that Remco van der Hofstad kindly agreed to review my thesis!). Under some regularity conditions on the sequence of degrees $\boldsymbol{d}^{(n)}$, there is at most one giant component (i.e. containing a positive fraction of the vertices) and we give conditions for the existence of such a giant component. These results go back to the work of Molloy and Reed [79] and a simple proof is provided by Janson and Luczak in [60] based on a very nice coupling argument and Glivenko-Cantelli theorem (see Proposition 4.24 in [65]). The proof is also explained in [95]. Indeed, the technique of proof was first developed for the k-core problem in [59], i.e. to compute the size of the largest induced subgraph with minimal degree at least k. In the case of the giant component problem, the proof is based on the analysis of an exploration process and in the case of the k-core problem, the proof is based on the analysis of a greedy algorithm removing low degree vertices. As low degree vertices are removed, it will possibly produce other low degree vertices among its neighbors and so on. The exploration process or this greedy algorithm can be interpreted as an epidemic process on the random graph activating (all or some) neighbors of already active vertices. In the case of the giant component problem, activated vertices are the ones that have been explored while in the greedy algorithm constructing the k-core, the activated vertices are those removed by the algorithm. Note that at the end of the first process, activated nodes constitute the explored component while at the end of the second process, the k-core is the subgraph induced by all vertices which are not active. In this chapter, we choose to follow the epidemics interpretation and show that the proof technique extends to much more general type of epidemics. As an example, we will obtain new results on cascading behavior in random networks. The proofs for the results in Section 1.2 can be found in [69].

Results in Section 1.2 concern static properties of the random graph. Although a dynamic process is introduced and analyzed, its main purpose is to allow for an explicit computation of its final state. In particular, the dynamic of the process has no natural meaning in term of epidemics. Note however that a recent work [58] extends the techniques of [60] to study the SIR epidemics on random graphs. In Section 1.3, we follow a different approach and consider first-passage percolation on random graphs. Imagine the random graph is connected and a piece of information has to be broadcast to all vertices from one initial vertex. We compute the time needed for all vertices to get informed if each edge requires an exponential time to transmit the information from one end to the other. This time depends on the initial node starting the broadcast and we compute the time associated to a typical vertex and also to a worst vertex. In other words, we compute weighted distances in random graphs with exponential weights. The proofs for the results in Section 1.3 can be found in [10].

1.1 Configuration model

One popular and important type of random graph is given by the uniformly distributed random graph with a given degree sequence, defined as follows. Let $n \in \mathbb{N}$ and let $\boldsymbol{d} = (d_i)_1^n$ be a sequence of non-negative integers. We let $G(n, \boldsymbol{d})$ be a random graph with degree sequence $\boldsymbol{d}$, uniformly chosen among all possibilities (assuming there is any such graph at all; in particular, $\sum_i d_i$ has to be even).

It is well-known that it is often simpler to study the corresponding random multigraph $G^*(n, \boldsymbol{d})$ with given degree sequence $\boldsymbol{d} = (d_i)_1^n$, defined for every sequence $\boldsymbol{d}$ with $\sum_i d_i$ even by the configuration model (introduced in [23], [15]): take a set of d_i half-edges for each vertex i, and combine the half-edges into pairs by an uniform random matching of the set of all half-edges (this pairing is called a configuration); each pair of half-edges is then joined to form an edge of $G^*(n, \boldsymbol{d})$.

Probability asymptotics: we consider asymptotics as the number of vertices n tends to infinity, and thus we assume throughout the paper that we are given, for each n, a sequence $\boldsymbol{d}^{(n)} = (d_i)_1^n$ with $\sum_i d_i^{(n)}$ even. For notational simplicity we will usually not show the dependency on n explicitly; we thus write $\boldsymbol{d}$ and d_i, and similarly for other (deterministic or random) quantities. All unspecified limits and other asymptotics statements are for $n \to \infty$. For example, w.h.p. (with high probability) means with probability tending to 1 as $n \to \infty$ and $\xrightarrow{p}$ means convergence in probability as $n \to \infty$. Similarly, we use o_p , Ω_p and O_p in a standard way, see [61]. For example, if $X^{(n)} = X$ is a parameter of the random graph, $X = o_p(n)$ means that $\mathbb{P}(|X| > \epsilon n) \to 0$ as $n \to \infty$ for every $\epsilon > 0$, equivalently $|X|/n \xrightarrow{p} 0$, or for every $\epsilon > 0$, $|X| < \epsilon n$ w.h.p.

We assume that we are given $\boldsymbol{d} = (d_i)_1^n$ satisfying the following regularity conditions:

Condition 1. *For each n, $\boldsymbol{d} = (d_i^{(n)})_1^n$ is a sequence of non-negative integers such that $\sum_{i=1}^n d_i$ is even and, for some probability distribution $\boldsymbol{p} = (p_r)_{r=0}^\infty$ independent of n,*

(i) $|\{i : d_i = r\}|/n \to p_r$ *for every* $r \geq 0$ *as* $n \to \infty$*;*

(ii) $\lambda := \sum_{r \geq 0} r p_r \in (0, \infty)$*;*

(iii) $\sum_{i \in [n]} d_i^2 = O(n)$*.*

In words, we assume that the empirical distribution of the degree sequence converges to a fixed probability distribution $\boldsymbol{p}$ with a finite mean λ. Condition 1 (iii) ensures that $\boldsymbol{p}$ has a finite second moment and implies that the empirical mean of the degrees converges to λ (uniform integrability of the empirical distribution of the degrees).

We will study both the random multigraph $G^*(n, \boldsymbol{d})$ and the (simple) random graph $G(n, \boldsymbol{d})$. Note that $G^*(n, \boldsymbol{d})$ does not have the uniform distribution over all multigraphs with the given degree sequence (there is a weight with a factor $1/j!$ for every edge of multiplicity j, and a factor $1/2$ for every loop). However, conditioned on the multigraph being a simple graph, we obtain $G(n, \boldsymbol{d})$, the uniformly distributed random graph with given degree sequence. By [57], we know that the Condition 1(iii) implies $\liminf \mathbb{P}(G^*(n, \boldsymbol{d}) \text{ is simple}) > 0$. In this case, many results transfer immediately from $G^*(n, \boldsymbol{d})$ to $G(n, \boldsymbol{d})$, for example, every result of the type $\mathbb{P}(\mathcal{E}_n) \to 0$ for some events $\mathcal{E}_n$, and thus every result saying that some parameter converges in probability to some non-random value. This includes every results in this chapter. The proofs of the results given in this chapter first deal with the random multigraph $G^*(n, \boldsymbol{d})$ and in a last step, transfer the results to $G(n, \boldsymbol{d})$ by conditioning.

Remark 1.1. The results of this chapter, can be applied to some other random graphs models too by conditioning on the degree sequence. In particular, our results will apply whenever the random graph conditioned on the degree sequence has a uniform distribution over all possibilities. Notable examples of such graphs are $G(n,p)$, the Bernoulli random graph with n vertices and edge probability p and $G(n,m)$, the uniformly random graph with n vertices and m edges. For example, for $G(n,p)$ with $np \to \lambda \in (0,\infty)$ or $G(n,m)$ with $2m/n \to \lambda$, the Condition 1 holds in probability with (p_k) a Poisson distribution with parameter λ, $p_k = e^{-\lambda}\frac{\lambda^k}{k!}$ (see Appendix B in [10]).

1.2 An unified framework for contagions

In this section, we describe the contagion process for any given finite graph G with vertex set $[n]$. We still denote by d_i the degree of node $i \in [n]$. From now on, a vertex i is either active or inactive. In our model, the initial set of active nodes S (the seed) will remain active during the whole process of the contagion. We will consider **single activation** where the seed contains only one (or $o(n)$) vertex and **partial activation** where a positive fraction $\alpha > 0$ of the vertices is in the seed. In this last case, we consider a simple model where each node i draws independently of each other a Bernoulli random variable σ_i with parameter α and is considered as initially active (i.e. in the seed) if $\sigma_i = 1$ and not initially active otherwise. Clearly, then a fraction α chosen uniformly at random among the population constitutes the initial seed of the contagion.

Symmetric threshold model: we first present the symmetric threshold model which generalizes the bootstrap percolation presented above. Given a map $k : \mathbb{N} \mapsto \mathbb{N}$, the local rule is the following: a vertex of degree d becomes active if and only if its number of active neighbors exceeds $k(d)$.

Now the progressive dynamic of the diffusion on a finite graph $G = (V,E)$ operates as follows: some set of nodes S starts out being active; all other nodes are inactive. Time operates in discrete steps $t = 1, 2, 3, \ldots$. At a given time t, any inactive node i becomes active if its number of active neighbors is at least $k(d_i) + 1$. This in turn may cause other nodes to become active. It is easy to see that the final set of active nodes (after n time steps if $|V| = n$) only depends on the initial set S (and not on the order of the activations) and can be obtained as follows: set $X_i = \mathbb{1}(i \in S)$ for all i. Then as long as there exists i such that $\sum_{j\sim i} X_j > k(d_i)$, set $X_i = 1$, where $j \sim i$ means that i and j share an edge in G. When this algorithm finishes, the final state of node i is represented by X_i: $X_i = 1$ if node i is active and $X_i = 0$ otherwise.

Percolated threshold model: this model builds on previous model and depends on an additional parameter $\pi \in [0,1]$. Given any graph G and initial set S, we now proceed in two phases.

- bond percolation: randomly delete each edge with probability $1-\pi$ independently of all other edges. We denote the resulting random graph by G_π;
- apply the symmetric threshold model with thresholds $k(d)$: set $X_i = \mathbb{1}(i \in S)$ and then as long as there is i such that $\sum_{j\sim_\pi i} X_j > k(d_i)$, set $X_i = 1$, where $j \sim_\pi i$ means that i and j share an edge in G_π and d_i is the degree of node i in the original graph G.

Clearly if $\pi = 1$, this is exactly the symmetric threshold model. If in addition $k(d) = k$, then this model is the bootstrap percolation. On the other hand if $\pi \in (0,1)$ and $k(d) = 0$ for any d, then this process explores the connected components of the vertices in S in the graph G_π. Note that the percolated threshold model is not equivalent to the symmetric threshold model on the (bond) percolated graph since threshold depends on the degree in the original graph (and not in the percolated graph).

1.2.1 Contagion with partial activation

For integers $s \geq 0$ and $0 \leq r \leq s$, we denote by b_{sr} the binomial probabilities $b_{sr}(p) := \mathbb{P}(\mathrm{Bi}(s,p) = r) = \binom{s}{r}p^r(1-p)^{s-r}$.

For a graph G, let $v(G)$ and $e(G)$ denote the number of vertices and edges in G respectively. In this section, we assume that (in addition of the asymptotic degree distribution $\boldsymbol{p}$) the followings are given: a positive fraction α for the initial seed, a threshold function $k(d)$ and a bond percolation parameter π. These parameters are independent of n. The subgraph of $G(n,\boldsymbol{d})$ induced by the active (resp. inactive) nodes at the end of the diffusion is denoted by H (resp. I). For $r \leq s$, we denote by $v_{sr}(I)$ the number of vertices in I with degree s in G and r in I, i.e. the number of vertices with degree s in G which are

not activated and with r neighbors which are not activated either. We denote by $v_s(H)$ the number of activated vertices of degree s in G (and with possibly lower degree in H).

We define the functions:

$$\begin{aligned} h(z) &:= (1-\alpha)\sum_s p_s \sum_{r\geq s-k(s)} r b_{sr}(1-\pi+\pi z), \\ g(z) &:= \lambda z(1-\pi+\pi z) - h(z), \\ h_1(z) &:= (1-\alpha)\sum_s p_s \sum_{r\geq s-k(s)} b_{sr}(1-\pi+\pi z), \end{aligned}$$

where $\lambda = \sum_s s p_s$. We define (a justification of the use of the max follows from the detailed proof)

$$\hat{z} := \max\{z \in [0,1] :\ g(z) = 0\}. \tag{1.1}$$

Theorem 1.2. *Consider the random graph $G(n,\boldsymbol{d})$ (or $G^*(n,\boldsymbol{d})$) for a sequence $\boldsymbol{d} = (d_i)_1^n$ satisfying Condition 1. We consider the percolated threshold diffusion on this graph with a positive fraction α for the initial seed, a threshold function $k(d)$ and a bond percolation parameter π. We assume that either $\hat{z} = 0$, or if $\hat{z} \in (0,1]$ then there exists $\epsilon > 0$ such that $g(z) < 0$ for any z in the interval $(\hat{z}-\epsilon, \hat{z})$. We have:*

$$\begin{aligned} v(H)/n &\xrightarrow{p} 1 - h_1(\hat{z}), \\ v_s(H)/n &\xrightarrow{p} p_s - (1-\alpha)\sum_{k\geq s-k(s)} p_s b_{sk}(1-\pi+\pi\hat{z}), \\ v_{sr}(I)/n &\xrightarrow{p} (1-\alpha)\sum_{i+k(s)\geq s-r} p_s b_{sr}(\hat{z}) b_{s-r,i}(1-\pi), \\ e(I)/n &\xrightarrow{p} \left(\mathbb{1}(\pi\neq 1)\frac{\hat{z}}{2(1-\pi+\pi\hat{z})} + \mathbb{1}(\pi=1)\frac{1}{2}\right) h(\hat{z}). \end{aligned}$$

If we condition the induced graph I^* of inactive nodes in $G^*(n,\boldsymbol{d})$ on its degree sequence $\boldsymbol{d}^{I^*}$ and let n^{I^*} be the number of its vertices, then I^* has the distribution of $G^*(n^{I^*}, \boldsymbol{d}^{I^*})$.

Note that the bootstrap percolation corresponds to the particular case of the percolated threshold model with $\pi = 1$ and $K(d) = \theta \geq 0$ and our Theorem 1.2 allows to recover the results in [13] for random regular graphs and in [7] for random graphs with given vertex degrees.

1.2.2 Contagion with a single activation

In this section, we look at the contagion with one (or a small number $o(n)$ of) initial active node(s) in $G(n,\boldsymbol{d})$.

In this case, we need to define vertices playing a particular role. We call them pivotal vertices and denote by P the set of pivotal vertices defined as follows: start from the graph G_π, i.e. the initial graph in which each edge has been removed with probability $1-\pi$ and then remove all vertices with $k(d_i) \geq 1$. The set of pivotal vertices is the set of vertices spanned by the largest connected component of the remaining graph. Since we kept only connected vertices in G_π with $k(d_i) = 0$, we see that activating any of the pivotal vertex will activate its pivotal neighbors in the next step and at the end of the diffusion, all pivotal vertices (and perhaps others) will be active.

For $u \in [1,n]$, let $\mathcal{C}(u)$ (resp. $\mathcal{C}(1,\ldots,j)$) be the subgraph induced by the final active nodes with initial active node u (resp. initial active nodes $1,\ldots,k$). We also define $\mathcal{I}(u)$ as the subgraph induced by the inactive nodes with initial active node u. The set of vertices of $\mathcal{C}(u)$ and $\mathcal{I}(u)$ is a partition of the vertices of the original graph. From previous discussion, we see that if $u \in P$, then we have $P \subset \mathcal{C}(u)$.

In the case of random graphs $G(n,\boldsymbol{d})$, we will show that if the set of pivotal vertices is small (i.e. $o_p(n)$), then the final size of the diffusion is also small w.h.p. Of course, if the set of pivotal vertices is $\Omega_p(n)$, then the final set of active vertices is also $\Omega_p(n)$ as it contains all the pivotal vertices and our next theorem gives its asymptotic size. Before that, we need to define two sets of functions corresponding to activated vertices and pivotal vertices.

For activated nodes, the functions are exactly the same as in previous section with $\alpha = 0$:

$$\begin{aligned} h^a(z) &:= \sum_s p_s \sum_{r \geq s-k(s)} r b_{sr}(1-\pi+\pi z), \\ g^a(z) &:= \lambda z(1-\pi+\pi z) - h^a(z), \\ h_1^a(z) &:= \sum_s p_s \sum_{r \geq s-k(s)} b_{sr}(1-\pi+\pi z). \end{aligned}$$

We define

$$z^a := \sup\{z \in [0,1) : \ g^a(z) = 0\}. \tag{1.2}$$

We now define the functions associated to the pivotal vertices:

$$\begin{aligned} g^p(z) &= (1-\pi+\pi z)\left(\lambda z - \sum_{s:\ k(s)\geq 1} s p_s - \sum_{s:\ k(s)=0} s p_s (1-\pi+\pi z)^{s-1}\right), \\ h_1^p(z) &= \sum_{s:\ k(s)=0} p_s (1-\pi+\pi z)^s + \sum_{s:\ k(s)\geq 1} p_s, \end{aligned}$$

and $z^p = \sup\{z \in [0,1) : \ g^p(z) = 0\}$.

We call the following condition the cascade condition:

$$\pi \sum_{r:\ k(r)=0} r(r-1)p_r > \sum_r r p_r, \tag{1.3}$$

which can be rewritten as $\pi \mathbb{E}[D(D-1)\mathbb{1}(k(D) = 0)] > \mathbb{E}[D]$ where D is a random variable with distribution $\boldsymbol{p}$.

Theorem 1.3. *Consider the random graph $G(n,\boldsymbol{d})$ (or $G^*(n,\boldsymbol{d})$) for a sequence $\boldsymbol{d} = (d_i)_1^n$ satisfying Condition 1 and assume that in addition, we have for each n, $\sum_{i\in[n]} d_i^3 = O(n)$ (which in particular implies Condition 1 (iii)). We consider the percolated threshold diffusion on this graph with a threshold function $k(d)$ and a bond percolation parameter $\pi \in [0,1]$.*

(i) If the cascade condition (1.3) is satisfied, then

$$\lim_n \frac{v(P)}{n} = 1 - h_1^p(z^p) > 0.$$

Moreover, for any $u \in P$, we have w.h.p.

$$\liminf_n \frac{v(\mathcal{C}(u))}{n} = \liminf \frac{v(\cap_{u\in P}\mathcal{C}(u))}{n} \geq 1 - h_1^a(z^a) > 0,$$

where z^a is defined by (1.2). Moreover if $z^a = 0$ or z^a is such that there exists $\epsilon > 0$ with $g^a(z) < 0$ for $z \in (z^a - \epsilon, z^a)$, then we have for any $u \in P$:

$$\begin{aligned} v(\mathcal{C}(u))/n &\xrightarrow{p} 1 - h_1^a(z^a), \\ v_{sr}(\mathcal{I}(u))/n &\xrightarrow{p} \sum_{i+k(s)\geq s-r} p_s b_{sr}(z^a) b_{s-r,i}(1-\pi), \\ v_s(\mathcal{C}(u))/n &\xrightarrow{p} p_s - \sum_{j\geq s-k(s)} p_s b_{sj}(1-\pi+\pi z^a), \\ e(\mathcal{I}(u))/n &\xrightarrow{p} \left(\mathbb{1}(\pi \neq 1)\frac{z^a}{2(1-\pi+\pi z^a)} + \mathbb{1}(\pi = 1)\frac{1}{2}\right) h^a(z^a). \end{aligned}$$

If we condition the induced graph $\mathcal{I}^*(u)$ of inactive nodes in $G^*(n,\boldsymbol{d})$ on its degree sequence $\boldsymbol{d}^{\mathcal{I}^*(u)}$ and let $n^{\mathcal{I}^*(u)}$ be the number of its vertices, then $\mathcal{I}^*(u)$ has the distribution of $G^*(n^{\mathcal{I}^*(u)}, \boldsymbol{d}^{\mathcal{I}^*(u)})$.

(ii) If $\pi \sum_{r,\, k(r)=0} r(r-1)r_r < \sum_r r p_r$, *then for any* $j = o(n)$, $v(\mathcal{C}(1,\dots,j)) = o_p(n)$.

Remark 1.4. The technical condition on the sign of g^a is required to avoid some (pathological) cases. To illustrate it, consider the particular case $k(d) = 0$ and $\pi = 1$, so that the diffusion is a simple exploration process of the connected component of the initial active vertex in G. Then the cascade condition becomes with D a random variable with distribution $\boldsymbol{p}$:

$$\mathbb{E}[D(D-1)] > \mathbb{E}[D].$$

A simple computation shows that

$$h^a(z) = z\phi'_D(z),\ g^a(z) = z\left(\lambda z - \phi'_D(z)\right),\ \text{and},\ h^a_1(z) = \phi_D(z),$$

where $\phi_D(x) = \mathbb{E}[x^D]$ is the generating function of the asymptotic degree distribution. By an easy concavity argument, we see that the condition on the sign of g^a around z^a is satisfied unless $p_2 = 1$. Hence if $p_2 < 1$, we recover the condition for the existence of a giant component [79], [60]. Note that the case $p_2 = 1$ is quite exceptional and rather different behaviors can occur (see Remark 2.7 in [60]).

For the same reason, in (ii) we require the inequality to be strict which is in some cases a (slightly) conservative condition.

1.2.3 Application: cascading behavior in random networks

To illustrate our last result, we consider a simple case with $k(d) = qd$ and $\pi = 1$. Then, every inactive vertex of degree d that has $> qd$ active neighbors becomes active. This model may be a reasonable model for the spread of rumors or beliefs. To keep the analysis simple, we will only consider the Erdős-Rényi graph $G(n, \lambda/n)$ with $\lambda \in (0, \infty)$. Hence we have only two parameters: λ the mean degree in the graph and q which represents the required fraction of active nodes among its neighbors in order to become active.

Clearly, the lower q is, the easier the diffusion spreads. Indeed, for each λ, there exists a value $q_c(\lambda)$ such that: when $q < q_c(\lambda)$, the cascade condition 1.3 holds, whereas for $q > q_c(\lambda)$, it does not and we can apply Theorem 1.3 (ii) so that cascades are small. As shown in Figure 1.1 (Left), we see that q_c is a non-decreasing function of λ for $\lambda \leq 2$ and a non-increasing function of λ, for $\lambda \geq 4$. The second curve in Figure 1.1 corresponds to the contagion threshold for a scale-free random network whose degree distribution $p_r = \frac{r^{-\gamma}}{\zeta(\gamma)}$ (with $\zeta(\gamma) = \sum r^{-\gamma}$) is parametrized by the decay parameter $\gamma > 1$.

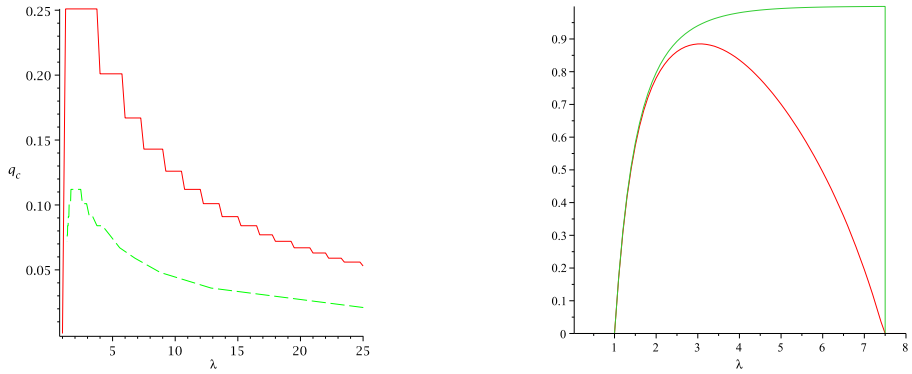

Figure 1.1: Left: $q_c(\lambda)$ for Erdős-Rényi random graphs and for power law graphs (dashed curve) as a function of the average degree λ. Right: Size of the cascade (in percent of the total population) for Erdős-Rényi random graphs as a function of λ the average degree for a fixed $q = 0.15$. The lower curve gives the asymptotic fraction of pivotal vertices.

We also observe that in both cases, for q sufficiently low, there are two critical values for the parameter λ, $1 < \lambda_i(q) < \lambda_s(q)$ such that a global cascade for a fixed q is only possible for $\lambda \in (\lambda_i(q); \lambda_s(q))$. The heuristic reason for these two thresholds is that a cascade can be prematurely stopped at high-degree nodes. For Erdős-Rényi random graphs, when $1 \leq \lambda < \lambda_i(q)$, there exists a giant component, i.e.

a connected component containing a positive fraction of the nodes. The high-degree nodes are quite infrequent so that the diffusion should spread easily. However, for λ close to one, the diffusion does not branch much and progresses along a very thin tree, "almost a line", so that its progression is stopped as soon as it encounters a high-degree node. Due to the variability of the Poisson distribution, this happens before the diffusion becomes large for $\lambda < \lambda_i(q)$. Nevertheless the condition $\lambda > \lambda_i(q)$ is not sufficient for a global cascade. Global diffusion also requires that the network be not too highly connected. This is reflected by the existence of the second threshold $\lambda_s(q)$ where a further transition occurs, now in the opposite direction. For $\lambda > \lambda_s(q)$, the diffusion will not reach a positive fraction of the population. The intuition here is clear: the frequency of high-degree nodes is so large that diffusion cannot avoid them and typically stops there since it is unlikely that a high enough fraction of their many neighbors eventually adopts. Following [98], we say that these nodes are locally stable.

The lower curve in Figure 1.1 (Right) represents the number of pivotal nodes in an Erdős-Rényi random graph as a function of λ the average connectivity for $q^{-1} = 6.666...$: hence we keep only the largest connected component of an Erdős-Rényi random graph where we removed all vertices of degree greater than 6. By the same heuristic argument as above, we expect two phase transitions for the size of the set of pivotal players. In the proof of Theorem 1.3, we show that it is indeed the case and moreover that the phase transitions occur at the same values $\lambda_i(q)$ and $\lambda_s(q)$ as can be seen on Figure 1.1 (Right) where the normalized size (i.e. fraction) $\gamma(q, \lambda)$ of the set of pivotal players is positive only for $\lambda \in (\lambda_i(q), \lambda_s(q))$. Hence a cascade is possible if and only if there is a 'giant' component of pivotal players. Note also that both phase transitions for the pivotal players are continuous, in the sense that the function $\lambda \mapsto \gamma(q, \lambda)$ is continuous. This is not the case for the second phase transition for the normalized size of the cascade $s(q, \lambda)$ which can be computed from Theorem 1.3 (see [69]). The function $\lambda \mapsto s(q, \lambda)$ is continuous in $\lambda_i(q)$ but not in $\lambda_s(q)$ as depicted on Figure 1.1 (Right). This has important consequences: around $\lambda_i(q)$ the propagation of cascades is limited by the connectivity of the network as in standard epidemic models. But around $\lambda_s(q)$, the propagation of cascades is not limited by the connectivity but by the high-degree nodes which are locally stable.

1.3 First Passage Percolation

In previous section, we ignored the dynamic of the diffusion and concentrated on the final set of active/inactive vertices. In this section, we will concentrate on a careful analysis of the dynamic for a very simple model of diffusion also called first passage percolation. First-passage percolation (FPP) describes the dynamics of a fluid spreading within a random medium. In this section, we study the impact of the introduction of edge weights on the typical distances in a random graph and, in particular, on its diameter.

A weighted graph (G, w) is the data of a graph $G = (V, E)$ and a collection of weights $w = \{w_e\}_{e \in E}$ associated to each edge $e \in E$. We suppose that all the edge weights are non-negative. For two vertices a and $b \in V$, a path between a and b is a sequence $\pi = (e_1, e_2, \dots e_k)$ where $e_i = \{v_{i-1}, v_i\} \in E$ and $v_i \in V$ for $i \in \{1, \dots, k\} = [1, k]$, with $v_0 = a$ and $v_k = b$. We write $e \in \pi$ if the edge $e \in E$ belongs to the path π, i.e., if $e = e_i$ for an $i \in [1, k]$. For $a, b \in V$, the weighted distance between a and b is given by

$$\mathrm{dist}_w(a, b) = \mathrm{dist}_w(a, b; G) = \min_{\pi \in \Pi(a,b)} \sum_{e \in \pi} w_e \ ,$$

where the minimum is taken over all the paths between a and b in the graph G. If there is no such path, we set the distance to be infinite. The *weighted diameter* is then given by

$$\mathrm{diam}_w(G) = \max\{\mathrm{dist}_w(a, b),\ a, b \in V,\ \mathrm{dist}_w(a, b) < \infty\} \ .$$

and the *weighted flooding time* is

$$\mathrm{flood}_w(a, G) = \max\{\mathrm{dist}_w(a, b),\ b \in V,\ \mathrm{dist}_w(a, b) < \infty\}.$$

1.3.1 Main results

We will consider the random graph $G(n, \boldsymbol{d})$ for a sequence $\boldsymbol{d}$ satisfying Condition 1 with the additional condition: for some $\epsilon > 0$,

$$\sum_{i=1}^{n} d_i^{2+\epsilon} = O(n). \tag{1.4}$$

We define $\boldsymbol{q} = \{q_k\}_{k=0}^{\infty}$ the size-biased probability mass function corresponding to $\boldsymbol{p}$, by

$$\forall k \geq 0, \quad q_k := \frac{(k+1)p_{k+1}}{\lambda}, \tag{1.5}$$

and let ν denote its mean:

$$\nu := \sum_{k=0}^{\infty} k q_k \in (0, \infty) \qquad \text{(which is finite by condition (1.4))}. \tag{1.6}$$

Let $\phi_p(z)$ be the probability generating function of $\{p_k\}_{k=0}^{\infty}$: $\phi_p(z) = \sum_{k=0}^{\infty} p_k z^k$, and let $\phi_q(z)$ be the probability generating function of $\{q_k\}_{k=0}^{\infty}$: $\phi_q(z) = \sum_{k=0}^{\infty} q_k z^k = \phi_p'(z)/\lambda$. In this section, we will consider only the case where $\nu > 1$. In particular, there exists a unique μ in $(0, 1)$ such that $\mu = \phi_q(\mu)$ and if $\mathcal{C}$ is the size (in number of vertices) of the largest component of $G(n, (d_i)_1^n)$, then we have by [Molloy, Reed [79] - Janson, Luczak [60]]: $\mathcal{C}/n \xrightarrow{p} 1 - \phi_p(\mu) > 0$. In addition, we introduce

$$\mu_* = \phi_q'(\mu) = \sum_{k=1}^{\infty} k q_k \mu^{k-1} \in [0, 1)\,. \tag{1.7}$$

We can now state our main theorem.

Theorem 1.5. *Let $(G(n, (d_i)_1^n), w)$ be a sequence of random weighted graphs where $w = \{w_e\}_{e \in E}$ are i.i.d. rate one exponential random variables.*
Assume Condition 1 *and (1.4) hold. Assume that ν defined in (1.6) is such that $\nu > 1$.*
Assume that all the graphs have the same minimum degree denoted by $d_{\min} = \min_{i \in [1,n]} d_i \geq 1$ and moreover that $p_{d_{\min}} > 0$. Let $\Gamma : \mathbb{N}^* \to \mathbb{R}$ be defined by:

$$\Gamma(d) := d\, \mathbb{1}[d \geq 3] + 2(1 - q_1)\, \mathbb{1}[d = 2] + (1 - \mu_*)\, \mathbb{1}[d = 1]. \tag{1.8}$$

If we condition the vertices 1 and 2 to be connected, we have

$$\frac{\mathrm{dist}_w(1, 2; G(n, (d_i)_1^n))}{\log n} \quad \xrightarrow{p} \quad \frac{1}{\nu - 1}. \tag{1.9}$$

If we condition the vertex 1 to be in the largest component, we have

$$\frac{\mathrm{flood}_w(1, G(n, (d_i)_1^n))}{\log n} \quad \xrightarrow{p} \quad \frac{1}{\nu - 1} + \frac{1}{\Gamma(d_{\min})}. \tag{1.10}$$

Finally, we have:

$$\frac{\mathrm{diam}_w(G(n, (d_i)_1^n))}{\log n} \quad \xrightarrow{p} \quad \frac{1}{\nu - 1} + \frac{2}{\Gamma(d_{\min})}. \tag{1.11}$$

Remark 1.6. Note that $\nu > 1$ implies that there is a positive fraction of nodes in $G(n, (d_i)_1^n)$ with degree 3 or larger. In particular, we have $q_1 = 2p_2/\lambda < 1$ and $\mu_* < 1$ so that we have $\Gamma(d) > 0$ for all $d \in \mathbb{N}^* = \{1, 2 \dots\}$.

We now comment our result with respect to related literature. Our main contribution is (1.11) while results (1.9) and (1.10) follow from the analysis required to prove (1.11). Indeed, a much stronger version of (1.9) has been proved for a slightly different model of random graphs by Bhamidi, van der Hofstad, and Hooghiemstra in [19]. Theorem 3.1 in [19] shows that if the sequence $(d_i)_1^n$ is a sequence of i.i.d.

(non-degenerate) random variables with $d_{\min} \geq 2$ and finite variance, then there exists a random variable V such that (conditioning on 1 and 2 being connected)

$$\mathrm{dist}_w(1,2;G^*(n,(d_i)_1^n))) - \frac{\log n}{\nu - 1} \xrightarrow{d} V.$$

We expect this result to be valid for our model of random graphs $G(n,(d_i)_1^n)$ where the degrees d_i satisfy Condition 1 (but we did not try to prove it). [19] and [18] give also results when the degree sequence has no finite second moment and no finite first moment.

Motivated by the analysis of the diameter of the largest component of a critical Erdős-Rényi random graph (without edge weights), Ding, Kim, Lubetzky, and Peres [36] show that if $d_i = r \geq 3$ for all i, then we have with high probability:

$$\mathrm{diam}_w(G^*(n,r)) = \left(\frac{1}{r-2} + \frac{2}{r}\right)\log n + O(\log\log n).$$

In a previous work with M. Draief [8], we extended the analysis of [36] to graphs with minimum degree 3 to get:

$$\frac{\mathrm{diam}_w(G(n,(d_i)_1^n))}{\log n} \xrightarrow{p} \frac{1}{\nu - 1} + \frac{2}{d_{\min}}.$$

The intuition behind this formula is simple: consider a vertex with minimal degree $d_{\min}$, its closest neighbor is at distance given by an exponential random variable with rate $d_{\min}$ (i.e. the minimum of $d_{\min}$ exponential rate one random variables). Hence the probability for this distance to be larger than $\log n/d_{\min}$ is n^{-1}. Since there are order $p_{d_{\min}} n$ vertices with minimal degree, a simple argument shows that we will find two nodes with degree $d_{\min}$ and with closest neighbors at distance $\log n/d_{\min}$. The diameter will be obtained by taking a shortest path between these two nodes. Each such node will first give a contribution of $\log n/d_{\min}$ to reach its closest neighbor and then the path between these neighbors will be typical, of the order $\log n/(\nu - 1)$. This simple heuristic argument shows that our result on the diameter depends crucially on the weights being exponentially distributed or at least have an exponential tail. We refer to [20] for recent results on distances with i.i.d. weights. As we will see, the presence of nodes with degree one and two makes the analysis much more involved than in [36] or [8]. As soon as a fraction of nodes have degree two, there will be long paths constituted by a chain of such nodes and we will see that these paths contribute to the diameter.

We end this section by a simple remark. As explained in Section 1.1, our results can be applied to some other random graphs models too by conditioning on the degree sequence. Notable examples of such graphs are $G(n,p)$, the Bernoulli random graph with n vertices and edge probability p and $G(n,m)$, the uniformly random graph with n vertices and m edges. For example, for $G(n,p)$ with $np \to \lambda \in (0,\infty)$ or $G(n,m)$ with $2m/n \to \lambda$, our result gives in this case (note that $\phi_q(z) = e^{-\lambda(1-z)}$):

Theorem 1.7. *Let $\lambda > 1$ be fixed, and let $\mu_* < 1$ satisfy $\mu_* e^{-\mu_*} = \lambda e^{-\lambda}$. Assume $G_n = G(n,p)$ where $np \to \lambda \in (0,\infty)$ (or $G_n = G(n,m)$ with $2m/n \to \lambda \in (0,\infty)$) with i.i.d. rate 1 exponential weights on its edges. Then*

$$\frac{\mathrm{diam}_w(G_n)}{\log n} \xrightarrow{p} \frac{1}{\lambda - 1} + \frac{2}{1 - \mu_*}. \tag{1.12}$$

This result improves on a lower bound of the weighted diameter given by Bhamidi, van der Hofstad and Hooghiemstra in [17] Theorem 2.6. Note that [17] also deals with the case $np \to \infty$ which is out of the scope of the present paper.

1.3.2 Overview of the proof

Our work is a direct generalization of [36] with significantly more involved calculations. The first key idea of the proof from [36] is to grow balls centered at all vertices of the graph simultaneously. The time when two balls centered at a and b respectively intersect is exactly the half of the weighted distance between a and b. (In what follows, we will sometimes deliberately use the term time instead of the term weighted distance.) Hence the weighted diameter becomes twice the time when the last two balls intersect. A

simple argument shows that any two balls containing slightly more than $\sqrt{n}$ vertices ($2\sqrt{rn\log n}$ vertices for r-regular case) will intersect with high probability; see Proposition 3.1 in [10]. Hence it will be enough to control the time at which all balls have reached this critical size of order $\sqrt{n}$ in order to prove an upper bound for the weighted diameter. For a proof of the upper bound on the diameter, we apply an union bound argument as in [36]. Hence, we need to find the right time such that the probability for a (typical) ball to reach size $\sqrt{n}$ is of order n^{-1}. In order to do so, we use the second main idea of the proof: we couple the exploration process on the weighted graph with a continuous time Markov branching process. This coupling argument is quite standard and we will deal here with the same branching process approximation for the exploration process on the graph as in [19]. However, we are facing here new difficulties as we need to consider events of small probability for this exploration process (of order n^{-1}). In particular, we need to show that the coupling is still valid for such large deviations. When $d_{\min} \geq 3$, the argument of [36] can be extended easily [8]. But as soon as $d_{\min} \leq 2$, several complications happen. First as shown in [9], the asymptotics for the large deviations of the branching process depend on the minimal possible offspring. Second, as soon as $d_{\min} = 1$, the small components of the graph contain now a positive fraction of the nodes. We need to bound the diameter of these small components and to study the diameter of the largest component, we need to condition our exploration process on 'non-extinction'. Similarly, the presence of degree one nodes, significantly complicates the proofs. When $d_{\min} = 1$, the longest shortest path in a random graph will be between a pair of vertices a and b of degree one. Furthermore, this path consists of a path from a to the 2-core, a path through the 2-core, and a path from the 2-core to b. Also to apply the second moment method as in [36] in order to get a lower bound, we need to first remove vertices with degree one iteratively to work with the 2-core of the graph (indeed an augmented version of this 2-core, see Section 4.2 in [10] for details).

In [10], we first prove an upper bound for the weighted diameter. This will consist in defining the two parameters α_n and β_n with the following significance. (i) Two balls of size at least β_n intersect almost surely, (ii) considering the growing balls centered at a vertex in the graph, the time it takes for the balls to go from size α_n to size β_n have all the same asymptotic for all the vertices of the graph, and the asymptotic is half of the typical weighted distance in the graph, and (iii) the time it takes for the growing balls centered at a given vertex to reach size at least α_n is upper bounded by $\frac{1+\epsilon}{\Gamma(d_{\min})}\log n$ for all $\epsilon > 0$ with high probability (w.h.p.). This will show that the diameter is w.h.p. bounded above by $(1+\epsilon)(\frac{1}{\nu-1} + \frac{2}{\Gamma(d_{\min})})\log n$, for all $\epsilon > 0$. The last section provides the corresponding lower bound. To obtain the lower bound, we show that w.h.p. (iv) there are at least two nodes with degree $d_{\min}$ such that the time it takes for the balls centered at these vertices to achieve size at least α_n is worst than the other vertices, and is lower bounded by $\frac{1-\epsilon}{\Gamma(d_{\min})}\log n$, for all $\epsilon > 0$. And using this, we conclude that the diameter is w.h.p. bounded below by $(1-\epsilon)(\frac{1}{\nu-1} + \frac{2}{\Gamma(d_{\min})})\log n$, for all fixed $\epsilon > 0$, finishing the proof of our main theorem.

Chapter 2

Local weak convergence, spectral measures and combinatorial optimization

In this chapter, we give an unified view of three different papers: in [28] and [29], we dealt with spectral properties of random graphs, while in [30], we analyzed the monomer-dimer model on random graphs. It came somehow as a surprise to us that our results on these two different topics were so closely related. I hope that the presentation chosen in this chapter will elucidate this coincidence. At least, writing it helped me to get a better understanding of these problems! The results presented in this section extend partly those written in the original papers. We now have a much better understanding of the spectral properties of random graphs. The lecture notes by Charles Bordenave [27] give a recent overview of this field. Concerning the monomer-dimer model, [2] and [1] are recent works with a similar approach as the one developed in this chapter. Some of their results are restricted to lattices and we extend them to unimodular graphs.

This chapter is completely independent from previous chapter as the techniques used here are different. I would say that the general philosophy to approach the problem is rather original compared to the usual tools developed by the community working on random graphs and combinatorics. The general belief behind our approach is the following: the asymptotic study of algebraic, combinatorial or probabilistic parameters associated to graphs as the number of edges and vertices tends to infinity in a comparable way should depend only upon the limiting local geometry of the graphs. Originating from the study of disordered systems in statistical physics, this general heuristic has given rise to a powerful non-rigorous computational formalism known as the cavity method [77]. If true, this heuristic has important implications:

- practically, the parameter of interest can be efficiently approximated via distributed, local message-passing algorithms;
- theoretically, a unique self-consistent infinite limit can be defined and determined by purely local specifications.

In this chapter, I concentrate on this second point and give a mathematical confirmation of the cavity method for the spectral measure of large graphs [85], [22], [91] and the monomer-dimer problem [99] using the framework of the objective method by Aldous and Steele [6].

I start by defining the empirical spectral measure and the monomer-dimer model for a finite graph in Sections 2.1 and 2.2. In order to extend these notions to infinite graphs, I introduce some basic concepts of functional analysis in Section 2.3 (see [84] for a self-contained exposition of the fundamental principles). In Section 2.4, I define the framework of local weak convergence, the notion of unimodularity and mean spectral measure. The main results are presented in Section 2.5 and in Section 2.6, I get explicit asymptotics for large random graphs. Finally, in Section 2.7, I explain how part of the results can be extended to a much more general framework.

Notations: a graph G is a triple consisting of a vertex set V (possibly infinite but countable), an edge set E and a map that associates to every edge two vertices (not necessarily distinct) called its endpoints. When two vertices u and v are endpoints of an edge, we say they are neighbors and write $u \sim v$ to indicate this. We denote by the same symbol ∂v the set of neighbors of node $v \in V$ and the set of edges incident to v. The degree of v is the cardinal of ∂v. We will work only with locally finite graph, where the degree of every vertex is finite. For a graph $G = (V, E)$, a walk of length ℓ from x to y is a sequence $v_0 = x, v_1, \dots, v_\ell = y$, with $v_i \in V$ and $e_i = \{v_i, v_{i+1}\} \in E$ for $i = 0, \dots \ell - 1$. A simple cycle is a path with $\ell \geq 2$, $x = y$ and all other vertices are pairwise distinct. When such a path exists, x is said to be connected to y. Being connected to is clearly an equivalence relation on V and the associated equivalence classes are called the connected components of G. When there is only one connected component, we say that the graph is connected. A connected graph without simple cycles is called a tree.

$\mathbb{C}$ is the set of complex numbers and $\mathbb{H}$ is the set of complex numbers with positive imaginary part:

$$\mathbb{H} = \{z \in \mathbb{C},\ \Im(z) > 0\}.$$

Any probability measure μ on $\mathbb{R}$ is characterized by its Borel-Stieltjes transform, which is an analytic function from $\mathbb{H}$ to $\mathbb{H}$ (see [56] for details):

$$F_\mu(z) = \int_{\mathbb{R}} \frac{d\mu(\lambda)}{\lambda - z}.$$

We will use the following formula which is true for symmetric measures:

$$\mu(\{0\}) = \lim_{t \to 0} -\sqrt{-1} t F_\mu(\sqrt{-1} t). \tag{2.1}$$

We denote by $\Rightarrow$ the weak convergence of probability measures.

2.1 Empirical spectral measure of finite graphs

A finite graph $G = (V, E)$ on n vertices $V = \{1, \dots, n\}$ can be represented by a $\{0,1\}$-valued $n \times n$ matrix A called its adjacency matrix: $A_{uv} = \mathbb{1}(u \sim v)$. Since A is symmetric, it has n real eigenvalues $\lambda_1 \leq \cdots \leq \lambda_n$. The empirical spectral measure of the graph G is the probability measure on $\mathbb{R}$:

$$\mu_G = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i},$$

where δ_x denotes the Dirac measure concentrated on x.

We now describe a rooted spectral measure. We show that it is a "local functional" and that the empirical spectral distribution is the average of this rooted spectral measure when the root is chosen uniformly at random. Take an orthonormal basis of eigenvectors $(\phi_1, \dots, \phi_n)$ corresponding to the eigenvalues $\lambda_1, \dots, \lambda_n$. For $v \in V$, we denote by χ_v the vector with all entries equal to 0 except the v-th one equal to 1. The spectral measure of G rooted at $v \in V$ is defined as:

$$\mu_{(G,v)} = \sum_{i=1}^{n} \langle \phi_i | \chi_v \rangle^2 \delta_{\lambda_i}, \tag{2.2}$$

where $\langle \cdot | \cdot \rangle$ denotes the inner product in $\mathbb{R}^n$.

Note that we have:

$$\int x^k \mu_{(G,v)}(dx) = \langle \chi_v | A^k \chi_v \rangle,$$

which is the number of returning walks of length k starting at v. This quantity is local in the sense that it depends only on the neighborhood of v at depth k. Moreover, this description of the moment of $\mu_{(G,v)}$ shows that the measure does not depend on the particular choice of the basis. Also, it is easy to check that $\mu_{(G,v)}$ is a probability distribution on $\mathbb{R}$ and we have (see Lemma 2.1 in [87])

$$\mu_G = \frac{1}{n} \sum_{v=1}^{n} \mu_{(G,v)}. \tag{2.3}$$

In words, the empirical spectral measure is the spatial average of the rooted spectral measures.

2.2 The monomer-dimer model and the matching measure

A matching on a finite graph $G = (V, E)$ is a subset of edges with no common vertex. A matching is encoded by a binary vector $\mathbf{B} = (B_e,\ e \in E) \in \{0,1\}^E$ defined by $B_e = 1$ if and only if the edge e belongs to the matching. The constraint $\sum_{e\in\partial v} B_e \leq 1$ for all $v \in V$ ensures that $\mathbf{B}$ is a matching. Vertices v with $\sum_{e\in\partial v} B_e = 0$ are called monomer and edges e with $B_e = 1$ are called dimers. The size of the matching is given by $\sum_{e\in E} B_e$. For a finite graph G, we define the matching number of G as $\nu(G) = \max\{\sum_e B_e\}$ where the maximum is taken over matchings of G.

We introduce the family of probability distributions on the set of matchings parametrized by a parameter $z > 0$:

$$\mu_G^z(\mathbf{B}) = \frac{z^{\sum_e B_e}}{P_G(z)}, \tag{2.4}$$

where

$$\begin{aligned} P_G(z) &= \sum_{\mathbf{B}} z^{\sum_e B_e} \prod_{v\in V} \mathbb{1}\left(\sum_{e\in\partial v} B_e \leq 1\right) \\ &= \sum_{k=0}^{\nu(G)} m_k(G) z^k, \end{aligned}$$

where $m_k(G)$ is the number of matchings of size k ($m_0(G) = 1$).

We denote by $\nu^z(G)$ the mean size of a matching under μ_G^z, that is:

$$\begin{aligned} \nu^z(G) &= \sum_{e\in E} \mu_G^z(B_e = 1) \\ &= \frac{1}{P_G(z)} \sum_{k=1}^{\nu(G)} k m_k(G) z^k = \frac{z P_G'(z)}{P_G(z)}, \end{aligned}$$

$\mu_G^z(B_e = 1)$: sum over all matchings that include the edge e

where $P_G'(z)$ is the derivative of $P_G(z)$ with respect to z. For any finite graph, when z tends to infinity, the distribution μ_G^z converges to the uniform distribution over maximum matchings so that we have

$$\nu(G) = \lim_{z\to\infty} \nu^z(G). \tag{2.5}$$

The matching polynomial is defined as:

$$Q_G(z) = \sum_{k=0}^{\nu(G)} (-1)^k m_k(G) z^{n-2k} = z^n P_G(-z^{-2}).$$

We define the matching measure of G denoted by ρ_G as the uniform distribution over the roots of the matching polynomial of G:

$$\rho_G = \frac{1}{\nu(G)} \sum_{i=1}^{\nu(G)} \delta_{z_i}, \tag{2.6}$$

matching measure (fin)

where the z_i's are the roots of Q_G. Note that $Q_G(-z) = (-1)^n Q_G(z)$ so that ρ_G is symmetric.

The fundamental theorem for the matching polynomial is the following.

Theorem 2.1. *(Heilmann Lieb [54]) The roots of the matching polynomial $Q_G(z)$ are real.*

In particular, the matching measure of G is a probability measure on $\mathbb{R}$. Again, we can write ρ_G as a spatial average of rooted measures thanks to the work of Godsil [53]. A walk in a graph is self-avoiding if it touches every vertex at most once. For a finite graph G and a root vertex v, one can construct $T_v(G)$, the tree of self-avoiding walks at v as follows: its vertices correspond to the finite self-avoiding walks in G starting at v, and we connect two walks if one of them is a one-step extension of the other. Using [53], the following theorem is proved in [2]; it relates the matching measure of G with the rooted spectral measures of the $T_v(G)$'s:

Theorem 2.2. *Let G be a finite graph, then we have*

$$\rho_G = \frac{1}{n}\sum_{v=1}^{n} \mu_{(T_v(G),v)}.$$

In particular, Theorem 2.2 gives one of the several known proofs for the Heilmann-Lieb theorem. Moreover if G is a tree, we have $T_v(G) = G$ for any $v \in V$ so that in this case by (2.3), we have $\rho_G = \mu_G$.

Of course, the polynomials $P_G(z)$ or $Q_G(z)$ contains the same information as the matching measure ρ_G. We can express the quantity of interest in term of ρ_G: for $z > 0$,

$$\begin{aligned}
\frac{1}{n}\ln P_G(z) &= \frac{1}{2}\int \ln(1+z\lambda^2)d\rho_G(\lambda),\\
\frac{1}{n}\nu^z(G) &= \frac{1}{2}\int \frac{z\lambda^2}{1+z\lambda^2}d\rho_G(\lambda),\\
\frac{1}{n}\nu(G) &= \frac{1}{2}\left(1-\rho_G(\{0\})\right).
\end{aligned}$$

2.3 Rooted spectral measure of infinite graphs

In this section, we extend the notion of rooted spectral measure described in Section 2.1 to infinite graphs. For the sake of simplicity, we will restrict ourselves to graphs G with bounded degree, i.e. finite maximum degree. Most of the theory extends to locally finite graphs, see chapter 2.2 in [87].

A graph $G = (V, E)$ with bounded degree can be identified with a linear operator A on the Hilbert space

$$\ell^2(V) = \left\{\psi : V \to \mathbb{C} : \sum_{v\in V} |\psi(v)|^2 < \infty\right\},$$

endowed with its canonical inner product $\langle\psi|\phi\rangle = \sum_{v\in V}\overline{\psi(v)}\phi(v)$. Specifically, the domain of A is the dense subspace of finitely supported functions and the action of A on the canonical orthonormal basis $(\chi_v)_{v\in V}$, where $\chi_v(w) = \mathbb{1}(v = w)$, is given by

$$A\chi_v = \sum_{w\in\partial v}\chi_w.$$

A is called the adjacency operator of G. It is symmetric: $\langle\chi_v|A\chi_w\rangle = \langle A\chi_v|\chi_w\rangle$ and bounded as the graph G has bounded degree. As a consequence, A and G are said to be self-adjoint, the spectrum of A is real. Moreover, the spectral theorem for self-adjoint operators then guarantees that for every $\psi \in \ell^2(V)$ with $\|\psi\| = 1$, there exists a unique probability measure μ_ψ on $\mathbb{R}$ satisfying

$$\langle\psi|(A - zI)^{-1}\psi\rangle = \int_{\mathbb{R}} \frac{d\mu_\psi(\lambda)}{\lambda - z},$$

for all $z \in \mathbb{H}$, where $\mathbb{H}$ is the set of complex numbers with positive imaginary part (see [84]).

In particular, for each vertex $v \in V$, we may consider the spectral measure associated with the vector χ_v. We denote it by $\mu_{(G,v)}$ and call it the rooted spectral measure of G at v. We now check that this definition is compatible with (2.2) in the finite dimensional case, i.e. when A is a finite matrix. In this case, we have with the notations of Section 2.1:

$$\begin{aligned}
(A - zI)\psi &= \sum_{i=1}^{n}(\lambda_i - z)\langle\psi|\phi_i\rangle\phi_i,\\
(A - zI)^{-1}\psi &= \sum_{i=1}^{n}(\lambda_i - z)^{-1}\langle\psi|\phi_i\rangle\phi_i,\\
\langle\psi|(A - zI)^{-1}\psi\rangle &= \sum_{i=1}^{n}(\lambda_i - z)^{-1}\langle\psi|\phi_i\rangle^2.
\end{aligned}$$

Hence taking $\psi = \chi_v$, we see that we recover the expression (2.2) for $\mu_{(G,v)}$. Intuitively, $\mu_{(G,v)}$ may be thought as the local contribution of the vertex v to the spectral measure of A (which is a projection-valued measure on $\mathbb{R}$ in the infinite dimensional case). In particular, we can still interpret the k-th moment of $\mu_{(G,v)}$ as the number of returning walks of length k starting at v.

Now, we want to extend the relation (2.3) in a meaningful way to infinite graphs. To do this, we need to introduce the notion of unimodularity which is the main topic of the next section.

2.4 The framework of local weak convergence

This section gives a brief account of the framework of local weak convergence. For more details, we refer to the seminal paper [16] and to the surveys [6, 5].

Rooted graphs. A *rooted graph* (G, o) is a graph $G = (V, E)$ together with a distinguished vertex $o \in V$, called the *root*. We let $\mathcal{G}_\star$ denote the set of all locally finite connected rooted graphs considered up to *rooted isomorphism*, i.e. $(G, o) \equiv (G', o')$ if there exists a bijection $\gamma \colon V \to V'$ that preserves roots ($\gamma(o) = o'$) and adjacency ($\{i, j\} \in E \iff \{\gamma(i), \gamma(j)\} \in E'$). We write $[G, o]_h$ for the (finite) rooted subgraph induced by the vertices lying at graph-distance at most $h \in \mathbb{N}$ from o. The distance

$$\textsc{dist}\left((G, o), (G', o')\right) := \frac{1}{1 + r} \quad \text{where} \quad r = \sup\left\{h \in \mathbb{N} \colon [G, o]_h \equiv [G', o']_h\right\},$$

turns $\mathcal{G}_\star$ into a complete separable metric space, see [5]. why separable?

With a slight abuse of notation, (G, o) will denote an equivalence class of rooted graph also called unlabeled rooted graph in graph theory terminology. Note that the rooted spectral measures of two isomorphic rooted graphs are equal. Hence $\mu_{(G,o)}$ is the same for any choice of a rooted graph in its rooted-isomorphism class and our abuse of notation does not introduce any ambiguity. Similarly if two rooted graphs are isomorphic, then their rooted trees of self-avoiding walks are also isomorphic. It thus makes sense to define $(T_o(G), o)$ for elements $(G, o) \in \mathcal{G}_\star$.

what does $\mathcal{U}(G)$ look like

Local weak limits. Let $\mathcal{P}(\mathcal{G}_\star)$ denote the set of Borel probability measures on $\mathcal{G}_\star$, equipped with the usual topology of weak convergence (see e.g. [21]). Given a finite deterministic graph $G = (V, E)$, we construct a random element of $\mathcal{G}_\star$ by choosing uniformly at random a vertex $o \in V$ to be the root, and restricting G to the connected component of o. The resulting law is denoted by $\mathcal{U}(G)$. If $\{G_n\}_{n\geq 1}$ is a sequence of finite graphs such that $\{\mathcal{U}(G_n)\}_{n\geq 1}$ admits a weak limit $\mathcal{L} \in \mathcal{P}(\mathcal{G}_\star)$, we call $\mathcal{L}$ the *local weak limit* of $\{G_n\}_{n\geq 1}$. If (G, o) denotes a random element of $\mathcal{G}_\star$ with law $\mathcal{L}$, we shall use the following slightly abusive notation : $G_n \rightsquigarrow (G, o)$ and for $f : \mathcal{G}_\star \to \mathbb{R}$:

$$\mathbb{E}_{(G,o)}\left[f(G, o)\right] = \int_{\mathcal{G}_\star} f(G, o) d\mathcal{L}(G, o).$$

Unimodularity. Let $\mathcal{G}_{\star\star}$ denote the space of locally finite connected graphs with a distinguished oriented edge, taken up to the natural isomorphism relation and equipped with the natural distance, which turns it into a complete separable metric space. With $f : \mathcal{G}_{\star\star} \to \mathbb{R}$, we associate a function $\partial f : \mathcal{G}_\star \to \mathbb{R}$, defined by:

$$\partial f(G, o) = \sum_{i \in \partial o} f(G, o, i),$$

and also the reversal $f^* : \mathcal{G}_{\star\star} \to \mathbb{R}$ of f defined by:

$$f^*(G, o, i) = f(G, i, o).$$

It is shown in [5] that any (G, o) with law $\mathcal{L}$ arising as the local weak limit of some sequence of finite graphs satisfies

$$\mathbb{E}_{(G,o)}\left[\partial f(G, o)\right] = \mathbb{E}_{(G,o)}\left[\partial f^*(G, o)\right] \tag{2.7}$$

for any Borel $f\colon \mathcal{G}_{\star\star} \to [0,\infty)$. A measure $\mathcal{L} \in \mathcal{P}(\mathcal{G}_\star)$ satisfying this invariance is called *unimodular*, and the set of all unimodular probability measures on $\mathcal{G}_\star$ is denoted by $\mathcal{P}_u(\mathcal{G}_\star)$. Note that (2.7) can be expanded to:

$$\int_{\mathcal{G}_\star} \sum_{i\in\partial o} f(G,o,i)\, d\mathcal{L}(G,o) = \int_{\mathcal{G}_\star} \sum_{i\in\partial o} f(G,i,o)\, d\mathcal{L}(G,o).$$

Mean spectral measure. With the notation introduced above, the spectral measure μ_G of a finite graph G (2.3) can be expressed as the mean of $\mu_{(G,o)}$:

$$\mu_G = \mathbb{E}_{(G,o)}\left[\mu_{(G,o)}\right]. \tag{2.8}$$

The definition (2.8) makes sense for (possibly infinite) $(G,o) \in \mathcal{P}_u(\mathcal{G}_\star)$ and we call it the mean spectral measure of (G,o) (following the terminology used for random measures). More explicitly, for any Borel set B on $\mathbb{R}$, we have by definition:

$$\mu_G(B) = \int \mu_{(G,o)}(B) d\mathcal{L}(G,o), \text{ where, } \mu_{(G,o)} \text{ was defined in Section 2.3.}$$

Note that we are slightly abusing notation as μ_G depends on the law $\mathcal{L}$ of (G,o). Indeed $\mu_{\mathcal{L}}$ can be defined by the right-hand side of (2.8) for any $\mathcal{L} \in \mathcal{P}(\mathcal{G}_\star)$. If $\mathcal{L}$ is unimodular, we take the simplified notation $\mu_{\mathcal{L}} = \mu_G$ when no confusion is possible (it follows from Theorem 3.1 in [5] that for fixed underlying graphs, there is at most one unimodular measure).

Sofic measure. A measure $\mathcal{L} \in \mathcal{P}(\mathcal{G}_\star)$ is sofic if there exists a sequence of finite graphs $(G_n)_{n\geq 1}$ whose local weak limit is $\mathcal{L}$. The set of sofic measures is the closure for the topology of weak convergence of the set $\{\mathcal{U}(G)\,, G \text{ finite}\}$. We denote by $\mathcal{P}_s(\mathcal{G}_\star)$ the set of sofic measures. All sofic measures are unimodular: $\mathcal{P}_s(\mathcal{G}_\star) \subset \mathcal{P}_u(\mathcal{G}_\star)$ and the converse is open, see [5] for a discussion.

2.5 Main results

We are now ready to extend the results from finite graphs to infinite graphs.

Theorem 2.3. *Let $(G_n)_{n\geq 1}$ be a sequence of finite graphs with uniformly bounded degrees such that $G_n \rightsquigarrow (G,o)$. In particular $(G,o) \in \mathcal{P}_s(\mathcal{G}_\star)$. Then we have as $n \to \infty$,*

$$\begin{aligned} (i) \quad \mu_{G_n} &\Rightarrow \mathbb{E}_{(G,o)}\left[\mu_{(G,o)}\right] = \mu_G, \\ (i') \quad \rho_{G_n} &\Rightarrow \mathbb{E}_{(G,o)}\left[\mu_{(T_o(G),o)}\right] = \rho_G. \end{aligned}$$

For all $x \in \mathbb{R}$, we have

$$\begin{aligned} (ii) \quad \lim_{n\to\infty} \mu_{G_n}(\{x\}) &= \mu_G(\{x\}), \\ (ii') \quad \lim_{n\to\infty} \rho_{G_n}(\{x\}) &= \rho_G(\{x\}). \end{aligned}$$

Note that if G is a finite graph, wit $G_n = G$ for all n, we recover results in Sections 2.1 and 2.2. Theorem 2.3 deals only with sofic measures and Proposition 2.2 in [27] shows that (i) extends to converging sequences of unimodular measures.

As a consequence of Theorem 2.3, we have for the matching problem (results following this argument can already be found in the paper of Heilmann and Lieb, see Section 8 [54]):

Theorem 2.4. *Let (G_n) be a sequence of finite graphs with uniformly bounded degrees such that $G_n \rightsquigarrow (G,o)$. Then we have as $n \to \infty$, for $z > 0$,*

$$\begin{aligned} \frac{1}{n}\ln P_{G_n}(z) &\to \frac{1}{2}\int_{\mathbb{R}} \ln(1+z\lambda^2) d\rho_G(\lambda), \\ \frac{1}{n}\nu^z(G_n) &\to \frac{1}{2}\int_{\mathbb{R}} \frac{z\lambda^2}{1+z\lambda^2} d\rho_G(\lambda), \\ \frac{1}{n}\nu(G_n) &\to \frac{1}{2}\left(1-\rho_G(\{0\})\right). \end{aligned}$$

Let $T = (V, E)$ be a tree with bounded degree. Thanks to the recursive structure of the tree T, we show that the Borel-Stieltjes transform of $\mu_{(T,o)}$ can be computed recursively. Once a root $o \in V$ is fixed, we use the notation $i \to j$ to mean that j is a child of i in the rooted tree (T, o). Given $z \in \mathbb{H}$, a solution to the cavity recursion on the rooted tree (T, v) is a family $(x_i(z))_{i \in V} \in \mathbb{H}^V$ satisfying the local equations:

$$x_i(z) = \frac{-1}{z + \sum_{j \to i} x_j(z)}. \tag{2.9}$$

We then have the following result:

Theorem 2.5. *Let $T = (V, E)$ be a tree with bounded degree and $o \in V$. For every $z \in \mathbb{H}$, the cavity recursion on the rooted tree (T, o) has a unique solution. Moreover, we have*

$$F_{\mu_{(T,o)}}(z) = x_o(z).$$

Note that we do not need the assumption of unimodularity for this theorem to be true. Note also that the recursion (2.9) involves the whole tree but only the value computed at the root of the tree gives the correct Borel-Stieltjes transform of the rooted spectral measure. Salez proved a much stronger result characterizing the self-adjointness of trees by the existence of a unique solution to the cavity recursion (see Theorem 2.2 in [87]).

We now characterize more precisely $\mu_{(T,o)}(\{0\})$ thanks to (2.1). The idea is to iterate once the recursion (2.9) with $z = \sqrt{-1}t$ and to let t tends to zero. We then obtain what we call the cavity recursion at zero temperature. A solution to the cavity recursion at zero temperature on the rooted tree (T, o) is a family $(m_i)_{i \in V} \in [0, 1]^V$ satisfying the local equations:

$$m_i = \left(1 + \sum_{j \to i} \left(\sum_{k \to j} m_k \right)^{-1} \right)^{-1}, \tag{2.10}$$

with the conventions $1/0 = \infty$, $1/\infty = 0$ and the sum over the empty set is equal to zero. Then we have:

Theorem 2.6. *Let $T = (V, E)$ be a tree with bounded degree and $o \in V$. There exists a largest solution $(m_i)_{i \in V}$ in $[0, 1]^V$ to the cavity recursion at zero temperature on the rooted tree (T, o), i.e. for any solution $(m'_i)_{i \in V}$, we have $m'_i \leq m_i$ for all $i \in V$. Moreover, we have $\mu_{(T,o)}(\{0\}) = m_o$.*

Hence as a direct consequence of Theorem 2.6, we get

Theorem 2.7. *Let $(G_n)_{n \geq 1}$ be a sequence of finite graphs with uniformly bounded degrees such that $G_n \rightsquigarrow (G, o)$. Then we have $F_{\mu_G}(z) = \mathbb{E}_{(G,o)}[x_o(z)]$, where $x_o(z)$ is the value at the root o of the solution of the cavity recursion on $(T_o(G), o)$. Moreover, we have*

$$\lim_{n \to \infty} \frac{\nu(G_n)}{n} = \frac{1}{2} \left(1 - \mathbb{E}_{(G,o)}[m_o] \right),$$

where m_o is the value at the root o of the largest solution of the cavity recursion at zero temperature on $(T_o(G), o)$.

At this stage, the interest of this last theorem might not be apparent to the reader. In the next section, we will use it to compute explicitly the limit of the matching number when the limiting tree has the branching property. Of course, in the case where the limit of $(G_n)_{n \geq 1}$ is concentrated on trees, we have the following result:

Proposition 2.8. *Let $(G_n)_{n \geq 1}$ be a sequence of finite graphs with uniformly bounded degrees such that $G_n \rightsquigarrow (T, o)$ where T is a tree. Then we have $\mu_T = \rho_T$.*

2.6 Unimodular Galton-Watson trees and recursive distributional equations

In this section, we explain how recursions on trees automatically give rise to distributional fixed point equations when specialized to Galton-Watson trees. Such equations are a common ingredient in the objective method, see [6].

As explained in Section 2.4, a measure concentrated on trees arising as the local weak limit of some sequence of finite graphs (i.e. a sofic measure) needs to be unimodular. Hence, we first need to define a unimodular version of family trees of Galton-Watson branching processes.

Unimodular Galton-Watson trees. Let $\boldsymbol{p} = \{p_r\}_{r\geq 0}$ be a probability distribution on $\mathbb{N}$ with non-zero finite mean, and let $\widehat{\boldsymbol{p}} = \{\widehat{p}_r\}_{r\geq 0}$ denote its size-biased version:

$$\widehat{p}_r \quad = \quad \frac{(r+1)p_{r+1}}{\sum_k kp_k} \qquad (r \in \mathbb{N}) \tag{2.11}$$

A *unimodular Galton-Watson tree* with degree distribution $\boldsymbol{p}$ is a random rooted tree (T, o) obtained by a Galton-Watson branching process where the root has offspring distribution $\boldsymbol{p}$ and all its descendants have offspring distribution $\widehat{\boldsymbol{p}}$. The law of (T, o) is unimodular, and is denoted by $\textsc{ugwt}(\boldsymbol{p})$. Such trees play a distinguished role as they are the local weak limits of many natural sequences of random graphs, including those produced by the configuration model described in previous chapter.

Note that there are other unimodular probability measures supported on rooted trees for which the computations described below are still tractable, for example the skeleton tree [6] or bipartite versions of unimodular Galton-Watson trees [70]. We also note that all unimodular probability measures supported on rooted trees with maximum degree less than $d \geq 2$ are sofic, see [48].

Owing to the Markovian nature of the branching process, the set of equations (2.10) takes the much simpler form of a recursive distributional equations (RDE), which we now make explicit. We denote $\mathcal{P}(\mathbb{N})$ (resp., $\mathcal{P}([0,1])$) the space of probability distributions on $\mathbb{N}$ ($[0,1]$, resp.). Given $\boldsymbol{p}, \boldsymbol{p}' \in \mathcal{P}(\mathbb{N})$ and $\nu \in \mathcal{P}([0,1])$, we denote by $\Theta_{\boldsymbol{p},\boldsymbol{p}'}(\nu)$ the distribution of the $[0,1]$-valued r.v.

$$Y = \frac{1}{1 + \sum_{i=1}^{N} (\sum_{j=1}^{N_i'} X_{ij})^{-1}}, \tag{2.12}$$

where $N \sim \boldsymbol{p}$, $N_i' \sim \boldsymbol{p}'$ and $X_{ij} \sim \nu$, all of them being independent. With this notation in hands, Theorem 2.6 clearly implies the following.

Corollary 2.9. *Let $d \geq 2$. Assume $(T, o) \sim \textsc{ugwt}(\boldsymbol{p})$ with $p_i = 0$ for $i \geq d$. The random variable $\mu_{(T,o)}(\{0\})$ has distribution $\Theta_{\boldsymbol{p},\widehat{\boldsymbol{p}}}(\nu)$, where $\nu \in \mathcal{P}([0,1])$ is the largest solution to the RDE*

$$\nu = \Theta_{\widehat{\boldsymbol{p}},\widehat{\boldsymbol{p}}}(\nu). \tag{2.13}$$

In the above sentence, the word largest refers to the usual stochastic order on $\mathcal{P}([0,1])$: $\nu_1 \leq \nu_2$ if for any continuous function $\varphi : [0,1] \to \mathbb{R}$,

$$\int_{[0,1]} \varphi d\nu_1 \leq \int_{[0,1]} \varphi d\nu_2.$$

We now solve explicitly the distributional fixed point equation (2.13). We assume that $p_0 + p_1 < 1$, otherwise $\widehat{p_0} = 1$ and the only solution to (2.13) is clearly $\nu = \delta_1$. Let $\phi(x) = \sum_{r\geq 0} p_r x^r$ be the generating function of $\boldsymbol{p}$ and for any $x \in [0,1]$, we define:

$$M(x) = \phi'(1)x\overline{x} + \phi(1-x) + \phi(1-\overline{x}) - 1 \text{ where, } \overline{x} = \frac{\phi'(1-x)}{\phi'(1)}. \tag{2.14}$$

First observe that $M'(x) = \phi''(1-x)(\overline{\overline{x}} - x)$, and therefore any $x \in [0,1]$ where M admits a local extrema must satisfy $x = \overline{\overline{x}}$. We will say that M admits a historical record at x if $x = \overline{\overline{x}}$ and $M(x) > M(y)$ for any $0 \leq y < x$. Since $[0,1]$ is compact and M is analytic, there are only finitely many such records. In fact, they are in one-to-one correspondence with the solutions to the RDE (2.13).

Theorem 2.10. *If $x_1 < \cdots < x_r$ are the locations of the historical records of M, then the RDE (2.13) admits exactly r solutions; moreover, these solutions can be stochastically ordered, say $\nu_1 < \cdots < \nu_r$, and for any $i \in \{1, \ldots, r\}$:*

(i) $\nu_i(\{0\}) = 1 - x_i$;

(ii) $\Theta_{\boldsymbol{p},\widehat{\boldsymbol{p}}}(\nu_i)$ *has mean* $M(x_i)$.

In particular, $\mathbb{E}_{(T,o)}[\mu_{(T,o)}(\{0\})] = \max_{x\in[0,1]} M(x)$.

We can now summarize our results as follows:

Theorem 2.11. *Let $(G_n)_{n\geq 1}$ be a sequence of finite graphs with uniformly bounded degrees such that $G_n \rightsquigarrow (T, o)$ with $(T, o) \sim \mathrm{UGWT}(\boldsymbol{p})$ for a certain $\boldsymbol{p} \in \mathcal{P}(\mathbb{N})$. We have*

$$\begin{aligned} \lim_{n\to\infty} \mu_{G_n}(\{0\}) &= \max_{x\in[0,1]} M(x) \\ \lim_{n\to\infty} \frac{\nu(G_n)}{n} &= \frac{1 - \max_{x\in[0,1]} M(x)}{2}, \end{aligned}$$

where $M(x)$ is defined in (2.14).

In [29][1] and [30], it is shown how to relax the assumption of uniformly bounded degree to the weaker assumption that $\boldsymbol{p}$ has a finite second moment.

So far, we stated our results for a sequence of deterministic graphs but it is easy to extend them to random graphs introduced in Section 1.1 of the previous chapter. Under the assumption that the degree sequence $\boldsymbol{d}$ satisfies Condition 1 with limiting distribution $\boldsymbol{p}$, the local weak limit of $G(n, \boldsymbol{d})$ (or $G^*(n, \boldsymbol{d})$) is $\mathrm{UGWT}(\boldsymbol{p})$ almost-surely, see [26]. Hence, we can now summarize our previous results in this setting in the following theorem. Note that so far, we only considered sequence of finite graphs with uniformly bounded degrees but in the setting of random graphs, this assumption can be relaxed to $\sum_{i\in[n]} d_i^2 = O(n)$ which is exactly Condition 1 (iii).

Theorem 2.12. *Let $G(n, \boldsymbol{d})$ be a sequence of random graphs for a sequence $\boldsymbol{d}$ satisfying Condition 1 with limiting distribution $\boldsymbol{p}$. Let $(T, o) \sim \mathrm{UGWT}(\boldsymbol{p})$ and $\mu_T = \mathbb{E}_{(T,o)}[\mu_{(T,o)}]$. Then we have as $n \to \infty$,*

$$\begin{aligned} \mu_{G_n} &\Rightarrow \mu_T, \ \textit{a.s.} \\ \rho_{G_n} &\Rightarrow \mu_T, \ \textit{a.s.} \\ \frac{1}{n} \ln P_{G_n}(z) &\to \frac{1}{2} \int_{\mathbb{R}} \ln(1 + z\lambda^2) d\mu_T(\lambda), \ \textit{a.s.} \\ \frac{1}{n} \nu^z(G_n) &\to \frac{1}{2} \int_{\mathbb{R}} \frac{z\lambda^2}{1 + z\lambda^2} d\mu_T(\lambda), \ \textit{a.s.} \\ \frac{1}{n} \nu(G_n) &\to \frac{1 - \max_{x\in[0,1]} M(x)}{2}, \ \textit{a.s.} \end{aligned}$$

where $M(x)$ is defined in (2.14).

2.7 Extensions

In previous sections, we put emphasis on the matching measure and in this approach, Heilmann-Lieb Theorem 2.1 is crucial. This theorem does not extend to the other combinatorial optimization problems that we will discuss now [97]. However, using a slightly different approach, we now show that most of previous results extend to a much more general setting.

Let start with a finite graph G with a distinguished root o. We define the root exposure probability $r_o(z) = \frac{P_{G-o}(z)}{P_G(z)}$ where $G - o$ is the graph obtained from G by removing vertex o (and all its incident edges). We see that for $z > 0$, $r_o(z)$ is the probability for the root o to be exposed (i.e., such that $\sum_{e\in\partial o} B_e = 0$) in a random matching $\mathbf{B}$ sampled according to the Gibbs distribution μ_G^z defined by

[1] The result in [29] is weaker than Theorem 2.11 because we were not aware of Lück's approximation, see Section 2.5 in [27] for completeness.

(2.4). A simple calculation shows that this root exposure probability can be expressed as a function of the Borel-Stieltjes transform of $\mu_{(T_o(G),o)}$ as follows:

$$r_o(z) = -\frac{\sqrt{-1}}{\sqrt{z}} F_{\mu_{(T_o(G),o)}}\left(\frac{\sqrt{-1}}{\sqrt{z}}\right). \tag{2.15}$$

In particular, instead of focusing on the matching measure and its Borel-Stieljes transform, we could have obtained the results concerning the convergence of the various quantities $\frac{1}{n}\ln P_{G_n}(z)$, $\frac{1}{n}\nu^z(G_n)$ or $\frac{1}{n}\nu(G_n)$ in term of the root exposure probability which satisfies now the local equations for $z > 0$:

$$r_i(z) = \frac{1}{1 + z\sum_{j\to i} r_j(z)}.$$

These equations can be obtained by doing the appropriate change of variables to translate results from $F_{\mu_{(T_o(G),o)}}$ to $r_o(z)$ as defined by (2.15) in (2.9). This is the approach followed in [30].

This approach is slightly modified in [71] by focusing on the probability for an edge to be covered (instead of the vertex exposure probability). Of course, this is essentially similar but it allows to extend easily to the problem of spanning subgraphs with degree constraints [88], [70] and to c-capacited b-matching in [67]. Note however, that in all these works, the existence and uniqueness of a solution to the cavity recursion relies heavily on the assumption that (G, o) is unimodular (which is not required for the monomer-dimer problem as seen in Theorem 2.5). Finally, recently I explored the algorithmic aspects of this line of work. In [73], we develop a sublinear-time algorithm for estimating $\ln P_G(z)$ at an arbitrary value $z > 0$ within additive error ϵn with high probability. In [72], I present an iterative loopy annealing message passing algorithm which is shown to converge and to solve a Linear Programming relaxation of the vertex cover or matching problem on general graphs.

Chapter 3

Statistics in high dimension

This chapter is made of two independent parts. The first part follows closely my joint work with Mohsen Bayati and Andrea Montanari on high-dimensional geometry [14]. This work is a good illustration of the fruitful dialog between applied and theoretical scientific disciplines. The second part presents my joint work with Laurent Massoulié and Jiaming Xu about community detection [68]. This work started our collaboration in a relatively new field for the three of us. We raised very interesting open questions in [68]. We are still actively working on these questions and we hope to give rigorous answers in a near future!

3.1 Compressed sensing and the geometry of random polytope in high-dimension

Let $A \in \mathbb{R}^{N\times N}$ be a random Wigner matrix, i.e. a symmetric random matrix with i.i.d. entries A_{ij} satisfying $\mathbb{E}\{A_{ij}\} = 0$ and $\mathbb{E}\{A_{ij}^2\} = 1/N$. Considerable effort has been devoted to studying the distribution of the eigenvalues of such a matrix [11, 12, 93]. The *universality phenomenon* is a striking recurring theme in these studies. Roughly speaking, many asymptotic properties of the joint eigenvalues distribution are independent of the entries distribution as long as the latter has the prescribed first two moments, and satisfies certain tail conditions. We refer to [11, 12, 93] and references therein for a selection of such results. Universality is extremely useful because it allows to compute asymptotics for one entries distribution (typically, for Gaussian entries) and then export the results to a broad class of distributions.

In this paper we are concerned with random matrix universality, albeit we do not focus on eigenvalues properties. Given $A \in \mathbb{R}^{N\times N}$, and an initial condition $x^0 \in \mathbb{R}^N$ independent of A, we consider the sequence $(x^t)_{t\geq 0}$ $t \in \mathbb{N}$ defined by letting, for $t \geq 0$,

$$x^{t+1} = A\, f(x^t; t) - \mathsf{b}_t\, f(x^{t-1}; t-1)\,, \qquad \mathsf{b}_t \equiv \frac{1}{N}\mathrm{div}(f(x; t))\big|_{x=x^t}, \tag{3.1}$$

where, by convention, $\mathsf{b}_0 = 0$. Here for each $t \geq 0$, $f(\,\cdot\,; t) : \mathbb{R}^N \to \mathbb{R}^N$ is a separable function, i.e. $f(z; t) = (f_1(z_1; t), \ldots, f_N(z_N; t))$. We also assume that the functions $f_i(\,\cdot\,; t) : \mathbb{R} \to \mathbb{R}$ are polynomials of bounded degree. In addition, div denotes the divergence operator, and in particular, $\mathsf{b}_t = N^{-1}\sum_{i=1}^N f_i'(x_i^t; t)$.

The present paper is concerned with the asymptotic distribution of x^t as $N \to \infty$ with t fixed, and establishes the following results:

Universality. As $N \to \infty$, the finite-dimensional marginals of the distribution of x^t are asymptotically insensitive to the distribution of the entries of A_{ij}.

State evolution. The entries of x^t are asymptotically Gaussian with zero mean, and variance that can be explicitly computed through a one-dimensional recursion, that we will refer to as *state evolution*

Phase transitions in polytope geometry. As an application, we use state evolution to prove universality of a phase transition on polytope geometry, with connections to compressed sensing. This

solves –for a broad class of random matrices with independent entries– a conjecture put forward by David Donoho and Jared Tanner in [37, 43].

In order to illustrate the usefulness of the first two technical results, we will start the presentation of our results from the third one.

Before stating our results, it is useful to comment on the special form of the iteration (3.1), and in particular on the role of the memory term $\mathsf{b}_t\, f(x^{t-1}; t-1)$ (which is inspired from the so-called 'Onsager correction' in statistical physics [94, 78, 25]). The function of this term is to cancel, to leading order, the correlations between x_i^{t+1} and $\{x_i^s : s \le t\}$. This cancellation is particularly transparent in our proof technique (see [14]), whereby x_i^t is expressed as a sum of monomials in A_{jk}, with $1 \le j, k \le n$, and is indexed by labeled trees. The memory term effectively cancels the contribution of 'one-step reversing' trees. We illustrate this idea on a simple example in Section 3.1.4

Without such memory term, the properties of the resulting iteration change crucially. In particular, it is no longer true that x_i^t is approximately Gaussian as $N \to \infty$ (see section 3.1.4 for further clarification on this point).

3.1.1 Universality of polytope neighborliness

In this work, we only consider convex polytopes. A polytope Q is said to be *centrosymmetric* if $x \in Q$ implies $-x \in Q$. Following [37, 38] we say that such a polytope is k-*neighborly* if the condition below holds:

(I) Every subset of k vertices of Q which does not contain an antipodal pair, spans a $(k-1)$ dimensional face.

The *neighborliness* of Q is the largest value of k for which this condition holds. The prototype of neighborly polytope is the ℓ_1 ball also called the cross-polytope $C^n \equiv \{x \in \mathbb{R}^n : \|x\|_1 \le 1\}$, whose neighborliness is indeed equal to n.

It was shown in a series of papers [38, 37, 41, 40, 42] that polytope neighborliness has tight connections with the geometric properties of random point clouds, and with sparsity-seeking methods to solve underdetermined systems of linear equations. The latter are in turn central in a number of applied domains, including model selection for data analysis and compressed sensing. For the reader's convenience, these connections will be briefly reviewed in Section 3.1.5.

Intuitive images of low-dimensional polytopes suggest that 'typical' polytopes are not neighborly: already selecting $k = 2$ vertices, does lead to a segment that connects them and passes through the interior of Q. This conclusion is spectacularly wrong in high dimension. Natural random constructions lead to polytopes whose neighborliness scales *linearly* in the dimension. Motivated by the above applications, and following [38, 37, 41, 40], we focus here on a weaker notion of neighborliness. Roughly speaking, this corresponds to the largest k such that *most* subsets of k vertices of Q span a $(k-1)$-dimensional face. In order to formalize this notion, we denote by $\mathfrak{F}(Q;\ell)$ the number of $\lfloor \ell \rfloor$-dimensional faces of Q. Note in particular that for $\ell \in \mathbb{N}$, we have $\mathfrak{F}(C^n;\ell) = 2^{\ell+1}\binom{n}{\ell+1}$ which is the maximum possible number of ℓ faces for a centrosymmetric polytope with $2n$ vertices.

Definition 1. *Let $\mathcal{Q} = \{Q^n\}_{n\ge 0}$ be a sequence of centrosymmetric polytopes indexed by n where Q^n has $2n$ vertices and has dimension $m = m(n)$: $Q^n \subseteq \mathbb{R}^m$. We say that $\mathcal{Q}$ has* weak neighborliness $\rho \in (0,1)$ *if for any $\xi > 0$,*

$$\begin{aligned} \lim_{n\to\infty} \frac{\mathfrak{F}(Q^n; m(n)\rho(1-\xi))}{\mathfrak{F}(C^n; m(n)\rho(1-\xi))} &= 1, \\ \lim_{n\to\infty} \frac{\mathfrak{F}(Q^n; m(n)\rho(1+\xi))}{\mathfrak{F}(C^n; m(n)\rho(1+\xi))} &= 0\,. \end{aligned}$$

If the sequence $\mathcal{Q}$ is random, we say that $\mathcal{Q}$ has weak neighborliness ρ (in probability) *if the above limits hold* in probability.

In other words, a sequence of polytopes $\{Q^n\}_{n\ge 0}$ has weak neighborliness ρ, if for large n the m dimensional polytope Q^n has close to the maximum possible number of k faces, for all $k < m\rho(1-\xi)$.

The existence of weakly neighborly polytope sequences is clear when $m(n) = n$ since in this case we can take $Q^n = C^n$ with $\rho = 1$, but the existence is highly non-trivial when m is only a fraction of n.

It comes indeed as a surprise that this is a generic situation as demonstrated by the following construction. For a matrix $A \in \mathbb{R}^{m\times n}$, and $S \subseteq \mathbb{R}^n$, let $AS \equiv \{Ax \in \mathbb{R}^m : x \in S\}$. In particular, AC^n is the centrosymmetric m-dimensional polytope obtained by projecting the n-dimensional ℓ_1 ball to m dimensions. The following result was proved in [37].

Theorem 3.1 (Donoho, 2005)**.** *There exists a function $\rho_* : (0,1) \to (0,1)$ such that the following holds. Fix $\delta \in (0,1)$. For each $n \in \mathbb{N}$, let $m(n) = \lfloor n\delta \rfloor$ and define $A(n) \in \mathbb{R}^{m(n)\times n}$ to be a random matrix with i.i.d. Gaussian entries.*

Then, the sequence of polytopes $\{A(n)C^n\}_{n\geq 0}$ has weak neighborliness $\rho_*(\delta)$ in probability.

A characterization of the curve $\delta \mapsto \rho_*(\delta)$ was provided in [37], but we omit it here since a more explicit expression will be given below.

The proof of Theorem 3.1 is based on exact expressions for the number of faces $\mathfrak{F}(A(n)C^n; \ell)$. These are in turn derived from earlier works in polytope geometry by Affentranger and Schneider [4] and by Vershik and Sporyshev [96]. This approach relies in a fundamental way on the invariance of the distribution of $A(n)$ under rotations.

Motivated by applications to data analysis and signal processing, Donoho and Tanner [43] carried out extensive numerical simulations for random polytopes of the form $A(n)C^n$ for several choices of the distribution of $A(n)$. They formulated a *universality hypothesis* according to which the conclusion of Theorem 3.1 holds for a far broader class of random matrices. The results of their numerical simulations were consistent with this hypothesis.

In [14], we establish the first rigorous result indicating universality of polytope neighborliness for a broad class of random matrices. Define the curve $(\delta, \rho_*(\delta))$, $\delta \in (0,1)$, parametrically by letting, for $\alpha \in (0,\infty)$:

$$\delta = \frac{2\phi(\alpha)}{\alpha + 2(\phi(\alpha) - \alpha\Phi(-\alpha))}, \tag{3.2}$$

$$\rho = 1 - \frac{\alpha\Phi(-\alpha)}{\phi(\alpha)}, \tag{3.3}$$

where $\phi(z) = e^{-z^2/2}/\sqrt{2\pi}$ is the Gaussian density and $\Phi(x) \equiv \int_{-\infty}^{x} \phi(z)\,\mathrm{d}z$ is the Gaussian distribution. Explicitly, if the above functions on the right-hand side of Eqs. (3.2), (3.3) are denoted by $f_\delta(\alpha)$, $f_\rho(\alpha)$, then[1] $\rho_*(\delta) \equiv f_\rho(f_\delta^{-1}(\delta))$.

Here we extend the scope of Theorem 3.1 from Gaussian matrices to matrices with independent subgaussian[2] entries (not necessarily identically distributed).

Theorem 3.2. *Fix $\delta \in (0,1)$. For each $n \in \mathbb{N}$, let $m(n) = \lfloor n\delta \rfloor$ and define $A(n) \in \mathbb{R}^{m(n)\times n}$ to be a random matrix with independent subgaussian entries, with zero mean, unit variance, and common scale factor s independent of n. Further assume $A_{ij}(n) = \tilde{A}_{ij}(n) + \nu_0 G_{ij}(n)$ where $\nu_0 > 0$ is a constant independent of n and $\{G_{ij}(n)\}_{i\in[m],j\in[n]}$ is a collection of i.i.d. $\mathsf{N}(0,1)$ random variables independent of $\tilde{A}(n)$.*

Then the sequence of polytopes $\{A(n)C^n\}_{n\geq 0}$ has weak neighborliness $\rho_*(\delta)$ in probability.

It is likely that this theorem can be improved in two directions. First, a milder tail condition than subgaussianity is probably sufficient. Second, we are assuming that the distribution of A_{ij} has an arbitrarily small Gaussian component. This is not necessary for the upper bound on neighborliness, and appears to be an artifact of the proof of the lower bound.

By comparison, the most closely related result towards universality is by Adamczak, Litvak, Pajor, and Tomczak-Jaegermann [3]. For a class of matrices $A(n)$ with i.i.d. columns, these authors prove that $A(n)C^n$ has neighborliness scaling linearly with n. This however does not suggest that a limit weak neighborliness exists, and is universal, as established instead in Theorem 3.2.

At the other extreme, universality of compressed sensing phase transitions can be conjectured from the results of the non-rigorous replica method [64, 83].

[1] It is easy to show that $f_\delta(\alpha)$ is strictly decreasing in $\alpha \in [0,\infty)$, with $f_\delta(0) = 1$, $\lim_{\alpha\to\infty} f_\delta(\alpha) = 0$, and hence f_δ^{-1} is well defined on $[0,1]$. Further properties of this curve can be found in [39, 46].

[2] See Eq. (3.7) for the definition of subgaussian random variables.

3.1.2 Universality of iterative algorithms

We will consider here and below a setting that is somewhat more general than the one described by Eq. (3.1). Following the terminology of [39], we will refer to such an iteration as to the approximate message passing (AMP) iteration/algorithm.

We generalize the iteration (3.1) to take place in the vector space $\mathcal{V}_{q,N} \equiv (\mathbb{R}^q)^N \simeq \mathbb{R}^{N\times q}$. Given a vector $x \in \mathcal{V}_{q,N}$, we shall most often regard it as an N-vector with entries in $\mathbb{R}^q$, namely $x = (\mathbf{x}_1, \dots, \mathbf{x}_N)$, with $\mathbf{x}_i \in \mathbb{R}^q$. Components of $\mathbf{x}_i \in \mathbb{R}^q$ will be indicated as $(x_i(1), \dots, x_i(q)) \equiv \mathbf{x}_i$.

There are several motivations for considering such a generalization. On one hand, it is necessary for the application to high-dimensional polytope geometry presented in the previous section. The reader might have noticed that the random matrix in Theorem 3.2 is rectangular. This is a different setting from the one of iteration (3.1), whereby the random matrix A is square and symmetric. The generalization to $x \in \mathcal{V}_{q,N}$ introduced here, with A square and symmetric, covers the case of rectangular matrices as well through a suitable reduction. In a nutshell, given a rectangular random matrix A', the reduction consists in constructing a symmetric matrix that has A' as submatrix.

Additional motivations for the generalization introduced here come from the application of AMP algorithms to a variety of problems in signal processing. For instance [66, 44] study compressed sensing reconstruction for 'spatially coupled' sensing matrices. These are random matrices with independent but not identically distributed entries. As already discussed in [44, 62] for the case of Gaussian entries, a rigorous analysis of this algorithm requires generalizing the setting of (3.1).

Several other applications require a generalization of the iteration (3.1), including the analysis of generalized AMP algorithms [82], AMP reconstruction of block-sparse signals [45], the analysis of phase retrieval algorithms [90] and so on. All of these applications can be treated within the setting introduced here, although our rigorous analysis requires the use of polynomial non-linearities.

A brief sketch of some proof ideas for the 'scalar' case of Eq. (3.1) can be found in Section 3.1.4.

Given a matrix $A \in \mathbb{R}^{N\times N}$, we let it act on $\mathcal{V}_{q,N}$ in the natural way, namely for $v', v \in \mathcal{V}_{q,N}$ letting $v' = Av$ be given by $\mathbf{v}'_i = \sum_{j=1}^N A_{ij}\mathbf{v}_j$ for all $i \in [N]$, i.e. $v'_i(\ell) = \sum_{j=1}^N A_{ij} v_j(\ell)$. Here and below $[N] \equiv \{1, \dots, N\}$ is the set of first N integers. In other words we identify A with the Kronecker product $A \otimes \mathrm{I}_{q\times q}$.

Definition 2. *An* AMP instance *is a triple $(A, \mathcal{F}, x^0)$ where:*

1. *$A \in \mathbb{R}^{N\times N}$ is a symmetric matrix with $A_{i,i} = 0$ for all $i \in [N]$.*
2. *$\mathcal{F} = \{f^k : k \in [N]\}$ is a collection of mappings $f^k : \mathbb{R}^q \times \mathbb{N} \to \mathbb{R}^q$, $(\mathbf{x}, t) \mapsto f^k(\mathbf{x}, t)$ that are locally Lipschitz in their first argument;*
3. *$x^0 \in \mathcal{V}_{q,N}$ is an initial condition.*

Given $\mathcal{F} = \{f^k : k \in [N]\}$, we define $f(\,\cdot\,; t) : \mathcal{V}_{q,N} \to \mathcal{V}_{q,N}$ that maps v to $v' = f(v; t)$, and is given by $\mathbf{v}'_i = f^i(\mathbf{v}_i; t)$ for all $i \in [N]$.

Definition 3. *The* approximate message passing orbit *corresponding to the instance $(A, \mathcal{F}, x^0)$ is the sequence of vectors $\{x^t\}_{t\ge 0}$, $x^t \in \mathcal{V}_{q,N}$ defined as follows, for $t \ge 0$,*

$$x^{t+1} = A\, f(x^t; t) - \mathsf{B}_t\, f(x^{t-1}; t-1)\,. \tag{3.4}$$

Here $\mathsf{B}_t : \mathcal{V}_{q,N} \to \mathcal{V}_{q,N}$ is the linear operator that maps v to $v' = \mathsf{B}_t v$, and is defined by,

$$\mathbf{v}'_i = \left(\sum_{j\in[N]} A_{ij}^2 \frac{\partial f^j}{\partial \mathbf{x}}(\mathbf{x}_j^t, t)\right) \mathbf{v}_i\,, \tag{3.5}$$

with $\frac{\partial f^j}{\partial \mathbf{x}}$ denoting the Jacobian matrix of $f^j(\,\cdot\,; t) : \mathbb{R}^q \to \mathbb{R}^q$.

The above definition can also be summarized by the following expression for the evolution of a single coordinate under AMP

$$\mathbf{x}_i^{t+1} = \sum_{j\in[N]} A_{ij} f^j(\mathbf{x}_j^t, t) - \sum_{j\in[N]} A_{ij}^2 \frac{\partial f^j}{\partial \mathbf{x}}(\mathbf{x}_j^t, t) f^i(\mathbf{x}_i^{t-1}, t-1)\,. \tag{3.6}$$

Notice that Eq. (3.1) corresponds to the special case $q = 1$, in which we replaced A_{ij}^2 by $\mathbb{E}\{A_{ij}^2\} = 1/N$ for simplicity of exposition. The term $\mathsf{B}_t\, f(x^{t-1}; t-1)$ in Eq. (3.4) is the correct generalization of the term $\mathsf{b}_t\, f(x^{t-1}; t-1)$ introduced in the $q = 1$ case, cf. Eq. (3.1). Namely it cancels –to leading order– the correlations between $\mathbf{x}_i^{t+1}$ and $\{\mathbf{x}_i^s,\ s \le t\}$.

Recall that a centered random variable X is *subgaussian* with scale factor σ^2 if, for all $\lambda > 0$, we have

$$\mathbb{E}\left(e^{\lambda X}\right) \le e^{\frac{\sigma^2\lambda^2}{2}}. \tag{3.7}$$

Definition 4. *Let $\{(A(N), \mathcal{F}_N, x^{0,N})\}_{N\ge1}$ be a sequence of AMP instances indexed by the dimension N, with $A(N)$ a random symmetric matrix and $x^{0,N}$ a random vector. For constants C, d (independent of N), we say that the sequence is* (C,d)-regular *(or, for short,* regular*)* polynomial *sequence if*

1. *For each N, the entries $(A_{ij}(N))_{1\le i<j\le N}$ are independent centered random variables. Further they are subgaussian with common scale factor C/N (see Eq. (3.7)).*
2. *For each N, the functions $f^i(\,\cdot\,; t)$ in $\mathcal{F}_N$ (possibly random, as long as they are independent from $A(N)$, $x^{0,N}$) are polynomials with maximum degree d and coefficients bounded by C.*
3. *For each N, $A(N)$ and $x^{0,N}$ are independent. Further, we have $\sum_{i=1}^N \exp\{\|\mathbf{x}_i^{0,N}\|_2^2/C\} \le NC$ with probability converging to one as $N \to \infty$.*

We state now our universality result for the algorithm (3.4).

Theorem 3.3. *Let $(A(N), \mathcal{F}_N, x^{0,N})_{N\ge1}$ and $(\tilde{A}(N), \mathcal{F}_N, x^{0,N})_{N\ge1}$ be any two (C,d)-regular polynomial sequences of instances, that differ only in the distribution of the random matrices $A(N)$ and $\tilde{A}(N)$.*

Denote by $\{x^t\}_{t\ge0}$, $\{\tilde{x}^t\}_{t\ge0}$ the corresponding AMP orbits. Assume further that for all N and all $i < j$, $\mathbb{E}\{A_{ij}^2\} = \mathbb{E}\{\tilde{A}_{ij}^2\}$. Then, for any set of polynomials $\{p_{N,i}\}_{N\ge0,1\le i\le N}$ $p_{N,i} : \mathbb{R}^q \to \mathbb{R}$, with degree bounded by d and coefficients bounded by a constant B for all N and $i \in [N]$, we have

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \left\{\mathbb{E}p_{N,i}(\mathbf{x}_i^t) - \mathbb{E}p_{N,i}(\tilde{\mathbf{x}}_i^t)\right\} = 0\,. \tag{3.8}$$

3.1.3 State evolution

Theorem 3.3 establishes that the behavior of the sequence $\{x^t\}_{t\ge0}$ is, in the high dimensional limit, insensitive to the distribution of the entries of the random matrix A. In order to characterize this limit, we need to make some assumption on the collection of functions $\mathcal{F}_N$. In particular, we need to relate the functions $\mathcal{F}_N$ to the functions $\mathcal{F}_{N'}$ in order to have a high-dimensional ($N \to \infty$) limit.

Informally, we define a *converging sequence* by requiring that for each N there exists a partition $[N] = C_1^N \cup C_2^N \cup \cdots \cup C_k^N$ (with k a fixed integer independent of N), and independent random variables $Y(i)$ taking values in $\mathbb{R}^{\tilde{q}}$, indexed by $i \in [N]$, such that:

- The function f^i only depends on the partition index of $i \in [N]$, and on the value of $Y(i)$.
- The distribution of $Y(i)$ only depends on the partition index of $i \in [N]$.
- The fractional size $|C_a^N|/N$ is N-independent for large N.

There are a few points to make precise, and this is done in the definition below.

Definition 5. *We say that the sequence of AMP instances $\{(A(N), \mathcal{F}_N, x^{0,N})\}_{N\ge0}$ is* polynomial and converging *(or simply* converging*) if it is (C,d)-regular and there exists: (i) An integer k; (ii) A symmetric matrix $W \in \mathbb{R}^{k\times k}$ with non-negative entries; (iii) A function $g : \mathbb{R}^q \times \mathbb{R}^{\tilde{q}} \times [k] \times \mathbb{N} \to \mathbb{R}^q$, with $g(\mathbf{x}, Y, a, t) = (g_1(\mathbf{x}, Y, a, t), \dots, g_q(\mathbf{x}, Y, a, t))$ and, for each $r \in [q]$, $a \in [k]$, $t \in \mathbb{N}$, $g_r(\,\cdot\,, Y, a, t)$ a polynomial with degree d and coefficients bounded by C; (iv) k probability measures P_1, ..., P_k on $\mathbb{R}^{\tilde{q}}$, with P_a a finite mixture of (possibly degenerate) Gaussians for each $a \in [k]$; (v) For each N, a finite partition $C_1^N \cup C_2^N \cup \cdots \cup C_k^N = [N]$; (vi) k positive semidefinite matrices $\widehat{\Sigma}_1^0, \dots \widehat{\Sigma}_k^0 \in \mathbb{R}^{q\times q}$, such that the following happens.*

1. *For each $a \in [k]$, we have $\lim_{N\to\infty} |C_a^N|/N = c_a \in (0,1)$.*

2. *For each $N \geq 0$, each $a \in [k]$ and each $i \in C_a^N$, we have $f^i(\mathbf{x}, t) = g(\mathbf{x}, Y(i), a, t)$ where $Y(1), \dots, Y(N)$ are independent random variables with $Y(i) \sim P_a$ whenever $i \in C_a^N$ for some $a \in [k]$.*

3. *For each N, the entries $\{A_{ij}(N)\}_{1 \leq i < j \leq N}$ are independent subgaussian random variables with scale factor C/N, $\mathbb{E}A_{ij} = 0$, and, for $i \in C_a^N$ and $j \in C_b^N$, $\mathbb{E}\{A_{ij}^2\} = W_{ab}/N$.*

4. *For each $a \in [k]$, in probability,*

$$\lim_{N \to \infty} \frac{1}{|C_a^N|} \sum_{i \in C_a^N} g\Big(\mathbf{x}_i^0, Y(i), a, 0\Big) g\Big(\mathbf{x}_i^0, Y(i), a, 0\Big)^{\mathsf{T}} = \widehat{\Sigma}_a^0 \,. \tag{3.9}$$

With a slight abuse of notation, we will sometimes denote a converging sequence by $\{(A(N), g, x^{0,N})\}_{N \geq 0}$. We use capital letters to denote the $Y(i)$'s to emphasize that they are random and do not change across iterations.

Our next result establishes that the low-dimensional marginals of $\{x^t\}$ are asymptotically Gaussian. *State evolution* characterizes the covariance of these marginals. For each $t \geq 1$, state evolution defines a set of k positive semidefinite matrices $\Sigma^t = (\Sigma_1^t, \Sigma_2^t, \dots, \Sigma_k^t)$, with $\Sigma_a^t \in \mathbb{R}^{q \times q}$. These are obtained by letting, for each $t \geq 1$

$$\Sigma_a^t = \sum_{b=1}^{k} c_b \, W_{ab} \, \widehat{\Sigma}_b^{t-1} \tag{3.10}$$

$$\widehat{\Sigma}_a^t = \mathbb{E}\left\{ g(Z_a^t, Y_a, a, t) g(Z_a^t, Y_a, a, t)^{\mathsf{T}} \right\} , \tag{3.11}$$

for all $a \in [k]$. Here $Y_a \sim P_a$, $Z_a^t \sim \mathsf{N}(0, \Sigma_a^t)$ and Y_a and Z_a^t are independent.

Theorem 3.4. *Let $(A(N), \mathcal{F}_N, x^0)_{N \geq 0}$ be a polynomial and converging sequence of AMP instances, and denote by $\{x^t\}_{t \geq 0}$ the corresponding AMP sequence. Then for each $t \geq 1$, each $a \in [k]$, and each locally Lipschitz function $\psi : \mathbb{R}^q \times \mathbb{R}^{\tilde{q}} \to \mathbb{R}$ such that $|\psi(\mathbf{x}, y)| \leq K(1 + \|y\|_2^2 + \|\mathbf{x}\|_2^2)^K$, we have, in probability,*

$$\lim_{N \to \infty} \frac{1}{|C_a^N|} \sum_{j \in C_a^N} \psi(\mathbf{x}_j^t, Y(i)) = \mathbb{E}\{\psi(Z_a, Y_a)\} \,, \tag{3.12}$$

where $Z_a \sim \mathsf{N}(0, \Sigma_a^t)$ is independent of $Y_a \sim P_a$.

We conclude by mentioning that, following [39], generalizations of the algorithm (3.4) were studied by several groups [89, 82, 74], for a number of applications. Universality results analogous to the one proved here are expected to hold for such generalizations as well.

3.1.4 Universality of iterative algorithms: Sketch of main ideas

In this section we sketch some key ideas in the proof of Theorems 3.3 and 3.4. For the sake of clarity, we shall focus on the special scalar recursion (3.1) with $f(x; t) = f(x)$ kept constant across iterations and $(A_{ij})_{i<j}$ independent centered subgaussian, with $\mathbb{E}\{A_{ij}^2\} = 1/N$. As in the statement of Theorem 3.3 and 3.4, we further assume that $f(\,\cdot\,)$ is separable and polynomial. Finally, we shall only consider the initial condition $x^0 = \mathbf{1}$ (the all-ones vector). While this setting is significantly more restrictive than the one of Theorems 3.3 and 3.4, it is sufficient to elucidate all the main ideas.

In order to clarify the role of the memory term in Eq. (3.1), it is instructive to first consider the case $k = 1$, $f(x)$ equal to the identity function i.e. $f(x) = ((\mathbf{x}_1), (\mathbf{x}_2), \dots, (\mathbf{x}_n))$, and drop the memory term, thus defining the sequence $\bar{x}^t \in \mathbb{R}^N$ by

$$\bar{x}^{t+1} = A \, \bar{x}^t \,. \tag{3.13}$$

Let us focus on –say– coordinate 1 of $\bar{x}^t$. An explicit calculation yields (recall that, by convention, $A_{ii} = 0$)

$$\bar{\mathbf{x}}_1^1 = \sum_{i \in [N]} A_{1i} \,, \tag{3.14}$$

$$\bar{\mathbf{x}}_1^2 = \sum_{i \in [N]} \sum_{j \in [N]} A_{1i} A_{ij} = \sum_{i \in [N]} A_{1i}^2 + \sum_{i \in [N]} \sum_{j \in [N] \setminus 1} A_{1i} A_{ij} \,. \tag{3.15}$$

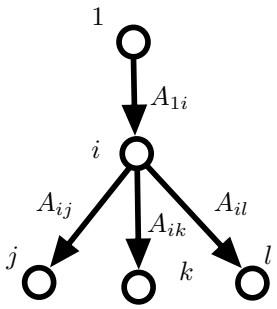

Figure 3.1: Graphical representation of a term in Eq. (3.18).

Consider first $t = 1$. Under our assumptions, $\bar{x}_1^1$ is a sum of i.i.d. random variables with mean 0 and variance $1/N$. By the central limit theorem, it converges in distribution to a standard Gaussian random variable, as predicted by Theorem 3.4.

Consider next $t = 2$. In Eq. (3.15) we decomposed the sum over $\{i, j\}$ in a sum over terms with $j = 1$, and a sum over terms with $j \neq 1$. The first sum $\sum_{i\in[N]} A_{1i}^2$ converges almost surely to 1 by the law of large numbers. It is easy to see that the second sum has expectation equal to zero and variance equal to $(N-1)/N$ that converges to 1. Indeed, a slightly more complicated calculation shows that it converges to a standard Gaussian. Overall, $\bar{\mathbf{x}}_1^2$ converges in distribution to a Gaussian with mean 1 and variance 1, unlike what is predicted by Theorem 3.4 for $\mathbf{x}_1^2$. (Theorem 3.4 always predicts $\mathbf{x}_i^t$ to have asymptotically zero mean.)

Notice that the terms in the sum (3.15) are indexed by an ordered triple $(1, i, j)$ with $i, j \in [N]$, $1 \neq i$, $i \neq j$. We can identify such a triple with a length 2 rooted (directed acyclic) path with vertices labeled by 1 (the root), i, j: $1 \to i \to j$. The terms that lead to a non-zero mean are those corresponding to $j = 1$, i.e. with a one-step reversal in the order in which they visit labels of $[N]$. These are paths of the form $1 \to i \to 1$.

Consider now adding back the memory term $\mathsf{b}_t\, f(x^{t-1}; t-1)$. It is easy to check that, in the present case (namely $f(x; t) = x$), Eq. (3.1) reduces to $x^{t+1} = Ax^t - x^{t-1}$ and, in particular

$$x^2 = A^2\mathbf{1} - \mathbf{1}\,.$$

Comparing with Eq. (3.15), we see that the memory term asymptotically cancels the effect of one-step reversing paths. The same analysis can be developed, with additional labor, to subsequent iterations. At each t, the memory term cancels the effect of one-step reversing paths, and the residual terms match the prediction of Theorem 3.4.

The proof follows a similar argument for a general polynomial $f(x)$. As in the linear case, each coordinate x_i^t can be expressed as a sum of monomials in the independent random variables $(A_{ij})_{i<j}$. The main difference is that now these monomials are indexed by rooted *trees* instead of rooted *paths* with vertex labels in $[N]$. To see this, consider the special case

$$f(x) = ((\mathbf{x}_1)^3, (\mathbf{x}_2)^3, \dots, (\mathbf{x}_n)^3)\,. \tag{3.16}$$

Then, a direct calculation of iteration (3.1) yields

$$\mathbf{x}_1^1 = \sum_{i\in[N]} A_{1i}\,, \tag{3.17}$$

$$\mathbf{x}_1^2 = \sum_{i\in[N]} \sum_{j,k,l\in[N]} A_{1i}A_{ij}A_{ik}A_{il} - \mathsf{b}_1\,, \qquad \mathsf{b}_1 = \frac{1}{N}\sum_{i\in[N]} 3\Big(\sum_{j\in[N]} A_{ij}\Big)^2\,. \tag{3.18}$$

The monomials in the sum appearing in the expression for $\mathbf{x}_1^2$ in Eq. (3.18) can be associated to rooted directed trees as per Fig. 3.1. In this simple example it is easy to check that the memory term exactly cancels the contribution of one-step reversing trees, i.e. the terms in the sum with labels $j = 1$, or $k = 1$,

or $l = 1$. The other terms in the sum correspond to non-reversing trees and their total contribution is asymptotically Gaussian with mean 0 and variance as predicted[3] by state evolution, Eq. (3.10). Since this sum is a polynomial in the independent random variables $(A_{ij})_{i<j}$, the method of moments provides a natural path to prove the last statement.

3.1.5 Proof of universality of polytope neighborliness

In this section we show how previous results allow us to prove Theorem 3.2. The approach we will follow is based on the equivalence between weak neighborliness and compressed sensing reconstruction developed in [38, 37, 41, 40]. Within compressed sensing, one considers the problem of reconstructing a vector $x \in \mathbb{R}^n$ from a vector of linear 'observations' $y = Ax$ with $y \in \mathbb{R}^m$ and $m \le n$. The measurement matrix $A \in \mathbb{R}^{m\times n}$ is assumed to be known. When $m \le n$, the problem is ill-posed because the system $At = y$ is highly under-determined. Thus to recover x, we need some information on its nature. The a priori hypothesis that we investigate is sparsity.

For $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, we denote $\mathrm{supp}(x) = \{i,\ 1 \le i \le n,\ x_i \neq 0\}$ the support of x. A vector x is said to be k-sparse if $|\mathrm{supp}(x)| \le k$. We denote by Σ_k the set of k-sparse vectors in $\mathbb{R}^n$. The exact reconstruction problem is then to reconstruct exactly any k-sparse vector $x \in \Sigma_k$ from the given data $y = Ax$.

An interesting approach towards reconstructing x from the linear observations y consists in solving a convex program:

$$\widehat{x}(y) = \arg\min \left\{ \|t\|_1 \text{ such that } t \in \mathbb{R}^n\,,\ \ y = At\,, \right\}. \tag{3.19}$$

One says that the matrix A has the exact reconstruction property of order k by ℓ_1-minimization if for every $x \in \Sigma_k$, the problem $\min_{t\in\mathbb{R}^n} \|t\|_1$ subject to $At = Ax$ has a unique solution equal to x. Donoho in [38] proved the following remarkable result :

Proposition 3.5. *Let A be an $m \times n$ matrix with columns $X_1, \dots, X_n \in \mathbb{R}^m$. The matrix A has the exact reconstruction property of order k by ℓ_1-minimization iff its columns $\pm X_1, \dots, \pm X_n$ are the $2n$ vertices of AC^n and AC^n is centrosymmetric k-neighborly.*

More precisely, he observed that if A and x are such that the arg min in (3.19) is uniquely defined and $\widehat{x}(y) = x$, then this will still be true for any x' with $\mathrm{supp}(x') = \mathrm{supp}(x)$. Hence for a given matrix A, the success of ℓ_1 minimization only depends on $\mathrm{supp}(x)$. Moreover the success of ℓ_1 minimization can be read from the geometry of the polytope $P = AC^n$ if it has $2n$ vertices. In this case, ℓ_1 minimization will succeed for $I = \mathrm{supp}(x)$ iff the polytope $P_I = P \cap \mathrm{Span}(X_i,\ i \in I)$ is a cross-polytope. In view of this result, Theorem 3.2 follows from the following result on compressed sensing with random sensing matrices.

Theorem 3.6. *Fix $\delta \in (0,1)$. For each $n \in \mathbb{N}$, let $m(n) = \lfloor n\delta \rfloor$ and define $A(n) \in \mathbb{R}^{m(n)\times n}$ to be a random matrix with independent subgaussian entries, with mean 0, variance $1/m$, and common scale factor s/m. Further assume $A_{ij}(n) = \tilde{A}_{ij}(n) + \nu_0\, G_{ij}(n)$ where $\nu_0 > 0$ is independent of n and $\{G_{ij}(n)\}_{i\in[m],j\in[n]}$ is a collection of i.i.d. $\mathsf{N}(0,1/m)$ random variables independent of $\tilde{A}(n)$. Assume that the matrix $A(n)$ has independent but not identically distributed entries. The vectors $x_0(n)$ have i.i.d. entries independent of $A(n)$, with $\mathbb{P}\{x_{0,i}(n) \neq 0\} = \rho\delta$. Then the following holds. If $\rho < \rho_*(\delta)$ then ℓ_1 minimization succeeds with high probability. Viceversa, if $\rho > \rho_*(\delta)$, then ℓ_1 minimization fails with high probability. (Here probability is with respect to the realization of the random matrix $A(n)$ and $x_0(n)$.)*

Notice that these matrices differ by a factor $1/\sqrt{m}$ from the matrices in the statement of Theorem 3.2. Since neighborliness is invariant under scale transformations, this change is immaterial. Let us now turn to the proof of Theorem 3.6. The following Lemma provides a useful sufficient condition for successful reconstruction. Here and below, for a convex function $F : \mathbb{R}^q \to \mathbb{R}$, $\partial F(x)$ denotes the subgradient of F at $x \in \mathbb{R}^q$. In particular $\partial\|x\|_1$ denotes the subgradient of the ℓ_1 norm at x. Further, for $R \subseteq [n]$, A_R denotes the submatrix of A formed by columns with index in R. The singular values of a matrix $M \in \mathbb{R}^{d_1\times d_2}$ are denoted by $\sigma_{\max}(M) \equiv \sigma_1(M) \ge \sigma_2(M) \ge \cdots \ge \sigma_{\min(d_1,d_2)}(M) \equiv \sigma_{\min}(M)$.

[3] In the present case, since $q = k = 1$, state evolution is a recursion for the single scalar Σ^t. We have $\Sigma^1 = 1$, $\Sigma^2 = \mathbb{E}\{g(Z^1)^2\}$ for $Z^1 \sim \mathsf{N}(0,1)$ and $g(x) = x^3$, whence $\Sigma^2 = 15$.

Lemma 1. *For any $c_1, c_2, c_3 > 0$, there exists $\varepsilon_0(c_1, c_2, c_3) > 0$ such that the following happens. If $x_0 \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, $y = Ax_0 \in \mathbb{R}^m$, are such that*

1. *There exists $v \in \partial \|x_0\|_1$ and $z \in \mathbb{R}^m$ with $v = A^{\mathsf{T}} z + w$ and $\|w\|_2 \le \sqrt{n}\,\varepsilon$, with $\varepsilon \le \varepsilon_0(c_1, c_2, c_3)$.*
2. *For $c \in (0,1)$, let $S(c) \equiv \{i \in [n] : |v_i| \ge 1 - c\}$. Then, for any $S' \subseteq [n]$, $|S'| \le c_1 n$, the minimum singular value of $A_{S(c_1) \cup S'}$ satisfies $\sigma_{\min}(A_{S(c_1) \cup S'}) \ge c_2$.*
3. *The maximum singular value of A satisfies $c_3^{-1} \le \sigma_{\max}(A)^2 \le c_3$.*

Then x_0 is the unique minimizer of $\|x\|_1$ over $x \in \mathbb{R}^n$ such that $y = Ax$.

The proof of Theorem 3.6 consists in two parts. For $\rho > \rho_*(\delta)$, we shall exhibit a vector x with $\|x\|_1 < \|x_0\|_1$ and $y = Ax$. For $\rho < \rho_*(\delta)$ we will show that assumptions of Lemma 1 hold. In particular, we will construct a subgradient v as per assumption 1. In both tasks, we will use an iterative message passing algorithm analogous to the one in Section 3.1.2. The algorithm is defined by the following recursion initialized with $x^0 = 0$:

$$x^{t+1} = \eta(x^t + A^{\mathsf{T}} z^t; \alpha\sigma_t)\,, \tag{3.20}$$

$$z^t = y - Ax^t + \mathsf{b}_t\, z^{t-1}\,, \tag{3.21}$$

where $\eta(u;\theta) = \mathrm{sign}(u)\,(u-\theta)_+$, α is a non-negative constant, and b_t is a diagonal matrix whose precise definition is immaterial here. Notice that the nonlinear mapping $\eta(\,\cdot\,;\alpha\sigma_t)$ is not a polynomial. This point should be addressed by constructing suitable *polynomial approximations* of η. Note also, that the matrix A is non-symmetric and we need to embed the rectangular matrix into a larger symmetric matrix to apply our results about iterative algorithms (see Section 4 in [14]).

3.2 Community detection in the labeled stochastic block model

Community detection aims to identify underlying communities of similar characteristics in an overall population from the observation of pairwise interactions between individuals [51]. The stochastic block model, also known as *planted partition model*, is a popular random graph model for analyzing the community detection problem [86, 35], in which pairwise interactions are binary: an edge is either present or absent between two individuals. In its simplest form, the stochastic block model consists of two communities of approximately equal size, where the within-community edge is present at random with probability p; while the across-community edge is present with probability q. If $p > q$, it corresponds to assortative communities where interactions are more likely within rather than across communities; while $p < q$ corresponds to disassortative communities.

In practice, interactions can be of various types and these types reveal more information on the underlying communities than the mere existence of the interaction itself. For example, in recommender systems, interactions between users and items come with user ratings. Such ratings contain far more information than the interaction itself to characterize the user and item types. Similarly, protein-protein chemical interactions in biological networks can be exothermic and endothermic; email exchanges in a club may be formal or informal; friendship in social networks may be strong or weak. The labeled stochastic block model was recently proposed in [55] to capture rich interaction types. In this model interaction types are described by labels drawn from an arbitrary collection. In particular, for the simple two communities case, the within-community edge is labeled at random with distribution μ; while the across-community edge is labeled with a different distribution ν. In this context an important question is how to leverage the labeling information for detecting underlying communities.

3.2.1 Information-Scarce Regime

Here, we focus on the sparse labeled stochastic block model in which every node has a limited average degree. We will provide results for the two cases where either $p, q = O(\mathrm{polylog}(n)/n)$ or $p, q = O(1/n)$, where n is the number of nodes. It corresponds to the information-scarce regime where only $O(n\mathrm{polylog}(n)$ or $O(n)$ edges and labels are observed in total. This regime is of practical interest, arising in several contexts. For example, in recommender systems, users only give ratings to few items; in

biological networks, only few protein-protein interactions are observed due to cost constraints; in social networks, a person only has a limited number of friends.

For the stochastic block model in the regime $p, q = O(1/n)$, there are $\Theta(n)$ isolated nodes, as in Erdős-Rényi random graphs with bounded average degree. For these isolated nodes, it is impossible to do better than a random guess and the probability to guess correctly their community membership is $1/2$. In particular, exact reconstruction of communities is impossible. Therefore, we resort to finding a partition into communities positively correlated to the true community partition (see Definition 6 below).

3.2.2 Main Results

Focusing on the two communities scenario, we show that a positively correlated reconstruction is fundamentally impossible when below a threshold. This establishes one half of the conjecture in [55]. In the positive direction, we establish the following results. We introduce a graph weighted by a suitable function of observed labels, on which we show that:

(1) Minimum bisection gives a positively correlated partition when above the threshold by a factor of $32 \ln 2$.

(2) A spectral method combined with removal of nodes of high degree gives a positively correlated partition when above the threshold by a constant factor.

3.2.3 Related Work

For the stochastic block model, most previous work focuses on the "dense" regime with an average degree diverging as the size of the graph n grows and "exact" reconstruction, such as [47, 63, 34]. McSherry [76] showed that the spectral method works as long as $p - q \geq \Omega(\sqrt{p \log n/n})$ with average degree as low as $\Omega(\log^6 n)$. Massoulié and Tomozei [75] reduced this lower bound on the average degree to $\Omega(\log n)$. More recently, it was shown in [31] that a matrix completion approach works when $p - q \geq \Omega(r\sqrt{p/n} \log^2 n)$ where the number of communities r could scale with n.

For the "sparse" regime with bounded average degrees, a sharp phase transition threshold for reconstruction (as defined in Definition 6) is conjectured in [35] by analyzing the belief propagation algorithm. The converse part of the conjecture was rigorously proved in [81]. In the converse direction, it is shown in [33] that a variant of spectral method gives a positively correlated partition when above the threshold by a large constant factor.

The labeled stochastic block was first proposed and studied in [55] and a new reconstruction threshold that incorporates the extra labeling information was conjectured. Simulations further indicate that the belief propagation algorithm works when above the threshold, but reconstruction algorithms that provably work are still unknown.

3.2.4 Model and Notation

This section formally defines the labeled stochastic block model with two symmetric communities and introduces the key notations and definitions used in the rest of this section.

The labeled stochastic block model $\mathcal{G}(n, p, q, \mu, \nu)$ is a random graph with n nodes of $\{\pm 1\}$ types and $\{\ell \in \mathcal{L}\}$-labeled edges, where $\mathcal{L}$ is a finite set of labels. To generate a particular realization (G, L, σ), first assign type $\sigma_u \in \{\pm 1\}$ to each node u uniformly and independently at random. Then, for every node pair (u, v), independently of everything else, draw an edge between u and v with probability p if $\sigma_u = \sigma_v$ and with probability q otherwise. Finally, each edge $e = (u, v)$ is labeled as $\ell \in \mathcal{L}$ independently at random with probability $\mu(\ell)$ if $\sigma_u = \sigma_v$ and with probability $\nu(\ell)$ otherwise.

When $\mu = \nu$, it reduces to the classical stochastic block model without labels. This paper focuses on the sparse case where $p = a_n/n$ and $q = b_n/n$ with $\max(a_n, b_n) = o(n)$. If we write a, b instead of a_n and b_n, we implicitly assume that we are in the setting where a and b are fixed constants. The goal is to reconstruct the true underlying types of nodes σ by observing the graph structure G and the labels on edges L.

Our goal is to reconstruct a type assignment which is positively correlated to the true type assignment. More formally, we adopt the following definition.

Definition 6. *A type assignment $\hat{\sigma}$ is said to be correlated with the true type assignment σ if a.a.s.*

$$\frac{1}{n}\min\{d(\sigma,\hat{\sigma}),d(\sigma,-\hat{\sigma})\}<1/2, \tag{3.22}$$

where d is the Hamming distance.

The shorthand a.a.s. denotes *asymptotically almost surely*. A sequence of events A_n holds a.a.s. if the probability of A_n converges to 1 as $n\to\infty$.

Define τ as

$$\tau=\sum_{\ell\in\mathcal{L}}\frac{(a\mu(\ell)-b\nu(\ell))^2}{2\,(a\mu(\ell)+b\nu(\ell))}. \tag{3.23}$$

It was conjectured in [55] that τ is the threshold for positively correlated reconstruction and more precisely:

Conjecture 3.7. *(i) If $\tau>1$, then it is possible to find a type assignment correlated with the true assignment a.a.s.*

(ii) If $\tau<1$, then it is not possible to find a type assignment correlated with the true assignment a.a.s.

In the sequel, we give a high-level view of the proof of (ii) and propose a simple spectral algorithm able to find a type assignment correlated with the true assignment for τ large enough.

3.2.5 Minimum Bisection

To recover the community partition, a natural way is via maximum likelihood estimation, i.e., $\hat{\sigma}=\arg\max_\sigma \log\mathbb{P}(G,L|\sigma)$. The log-likelihood can be written as:

$$\begin{aligned}\log\mathbb{P}(G,L|\sigma) \quad\propto\quad & \sum_{(u,v)\in E(G)}\log\frac{a\mu(L_{uv})}{b\nu(L_{uv})}\sigma_u\sigma_v\\ & +\log\left(\frac{1-a/n}{1-b/n}\right)\sum_{(u,v)\notin E(G)}\sigma_u\sigma_v\end{aligned}$$

Using the constraint $\sum_u\sigma_u=0$, maximum log likelihood estimation reduces to

$$\begin{aligned}&\max_\sigma \sum_{(u,v)\in E(G)}\left(\log\frac{a\mu(L_{uv})}{b\nu(L_{uv})}+\log\frac{1-a/n}{1-b/n}\right)\sigma_u\sigma_v\\ &\text{s.t.}\ \sum_u\sigma_u=0,\ \sigma_u\in\{\pm1\},\end{aligned}$$

If we ignore the $o(n)$ term in the sum above, this is equivalent to the minimum bisection on the weighted graph with a specific weight function $w(\ell)=\log\frac{a\mu(\ell)}{b\nu(\ell)}$. For a general weighing function $w:\mathcal{L}\to\mathbb{R}$, the minimum bisection finds the balanced bipartite subgraph in G with the minimum weighted cut, i.e.,

$$\begin{aligned}&\min_\sigma \sum_{(u,v):\sigma_u\neq\sigma_v}W_{uv}\\ &\text{s.t.}\ \sum_u\sigma_u=0,\ \sigma_u\in\{\pm1\},\end{aligned} \tag{3.24}$$

where $W_{uv}=A_{uv}w(L_{uv})$ and A is the adjacency matrix of G.

Theorem 3.8. *Assume the technical condition: $\sum_\ell a\mu(\ell)w^2(\ell),\sum_\ell b\nu(\ell)w^2(\ell)>8\ln 2$. Then if*

$$\frac{\left(\sum_\ell(a\mu(\ell)-b\nu(\ell))w(\ell)\right)^2}{\sum_\ell(a\mu(\ell)+b\nu(\ell))w^2(\ell)}>64\ln 2, \tag{3.25}$$

solutions of the minimum bisection (3.24) are positively correlated to the true type assignment $\sigma^\star$ a.a.s. Moreover, the left hand side of (3.25) is maximized when $w(\ell)=\frac{a\mu(\ell)-b\nu(\ell)}{a\mu(\ell)+b\nu(\ell)}$, in which case (3.25) reduces to $\tau>32\ln 2$.

Note that surprisingly, the weight function suggested by the maximum likelihood estimation is not optimal.

3.2.6 Spectral Method

The minimum bisection is NP-hard [52]. In this section, we present a simple spectral algorithm based on the matrix W introduced above (see [32] for a similar approach in the unlabeled case). We show that this algorithm finds an assignment correlated with the true assignment provided τ is large enough.

First note that we have $\mathbb{E}[W|\sigma] = \frac{\alpha}{n}\mathbf{1}\mathbf{1}^\top + \frac{\beta}{n}\sigma\sigma^\top - \frac{\alpha+\beta}{n}\mathbf{I}$ with

$$\alpha = \frac{1}{2}\sum_\ell w(\ell)(a\mu(\ell) + b\nu(\ell)),$$
$$\beta = \frac{1}{2}\sum_\ell w(\ell)(a\mu(\ell) - b\nu(\ell)). \tag{3.26}$$

The term $\frac{\alpha+\beta}{n}\mathbf{I}$ is irrelevant to the main results (thanks to Weyl's perturbation theorem) and neglected for simplicity. Let $D = W - \frac{\alpha}{n}\mathbf{1}\mathbf{1}^\top$ and then $\mathbb{E}[D|\sigma] = \frac{\beta}{n}\sigma\sigma^\top$ has rank one with singular value β. Hence, it makes sense to define $\hat{D}$ as the best rank-1 approximation of the matrix D. In other words, if $D = \sum_i v_i x_i y_i^\top$ is the singular value decomposition of D with singular values $v_1 \geq v_2 \geq \ldots$, we define $\hat{D} = v_1 x_1 y_1^\top$. Then if the matrix D is "close" to its mean $\mathbb{E}[D|\sigma]$, we expect v_1 to be "close" to β and $\mathrm{sign}(x_1)$ to be correlated with σ, where $\mathrm{sign}(x)$ gives the sign of x component-wise. Unfortunately, in the regime where a and b are constants, there are vertices of degree $\Omega(\frac{\log n}{\log\log n})$ and thus the largest singular value of W could reach $\Omega(\sqrt{\frac{\log n}{\log\log n}})$ which is much higher than β.

In order to take care of this issue, we begin with a preliminary step to "clean" the spectrum of W: we remove all vertices in the graph with degree larger than $\frac{3}{2}\frac{a+b}{2}$. To summarize, for a given weight function $w(\ell)$, our algorithm Spectral − Reconstruction has the following structure:

1. Remove vertices with degree larger than $\frac{3}{2}\frac{a+b}{2}$ and assign a random type to these vertices.
2. Define W' by setting to zero the rows and columns of W corresponding to vertices removed.
3. Let $\hat{x}$ be the left-singular vector associated with the largest singular value of $D' = W' - \frac{\alpha}{n}\mathbf{1}\mathbf{1}^\top$. Output $\mathrm{sign}(\hat{x})$ for the types of the remaining vertices.

Observe that step 3) can be seen as a relaxation of the minimum bisection (3.24) by replacing the integer constraint with the unit-norm constraint and relaxing the hard constraint $\sum_u \sigma_u = 0$ to be a regularized term in the objective function. Spectral − Reconstruction needs estimates of α and $a+b$, which can be well approximated by $\frac{1}{n}\mathbf{1}^\top W\mathbf{1}$ and $\frac{2}{n}\mathbf{1}^\top A\mathbf{1}$ respectively. To simplify the analysis, we will assume that the exact value of α and $a+b$ is known.

Theorem 3.9. *Assume $|w(\ell)| \leq 1$ for all ℓ. There exists a universal constant C (i.e. not depending on a, b, μ or ν) such that if $\beta > C\sqrt{a+b}$, where β is defined in (3.26) then* Spectral − Reconstruction *outputs a type assignment correlated with the true assignment a.a.s. In the particular case, where $w(\ell) = \frac{a\mu(\ell)-b\nu(\ell)}{a\mu(\ell)+b\nu(\ell)}$, the condition $\beta > C\sqrt{a+b}$ reduces to $\tau > C'\sqrt{a+b}$.*

Compared to point (i) in the Conjecture 3.7, our result does not give the right order of magnitude when a and b are large. Indeed, we are able to improve it if we allow a_n and b_n to grow with n.

Theorem 3.10. *Assume that $\min(a_n, b_n) = \Omega(\log^6 n)$. If*

$$\liminf_{n\to\infty} \frac{\left(\sum_\ell (a_n\mu(\ell) - b_n\nu(\ell))w(\ell)\right)^2}{\sum_\ell (a_n\mu(\ell) + b_n\nu(\ell))w^2(\ell)} > 128, \tag{3.27}$$

then Spectral − Reconstruction *outputs a type assignment correlated with the true assignment a.a.s. Moreover, the left hand side of (3.27) is maximized when $w(\ell) = \frac{a_n\mu(\ell)-b_n\nu(\ell)}{a_n\mu(\ell)+b_n\nu(\ell)}$, in which case (3.27) reduces to $\tau > 64$. With this choice of $w(\ell)$, as soon as $\tau \to \infty$,* Spectral − Reconstruction *outputs the true assignment for all vertices except $o(n)$ a.a.s.*

Note that in the regime $\min(a_n, b_n) = \Omega(\log^6 n)$, the degrees are very concentrated and step 1) of the algorithm can be removed without harm. The simulation results, depicted in Fig. 3.2, further indicate that Spectral − Reconstruction leaving out step 1) outputs a positively correlated assignment

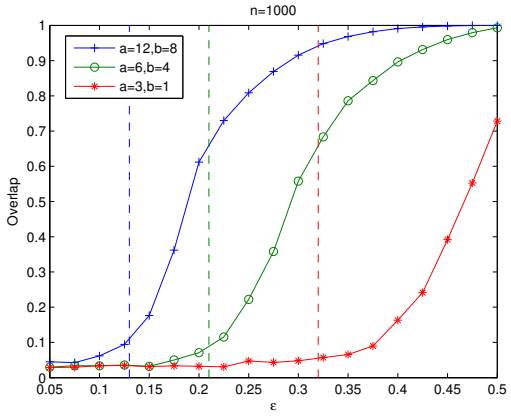

Figure 3.2: The overlap Q against ϵ from 0.05 to 0.5.

when above the threshold. In the simulation, we assume for simplicity only two labels: r and b, and define $\mu(r) = 0.5 + \epsilon$ and $\nu(r) = 0.5 - \epsilon$. We generate the graph from the labeled stochastic block model with $n = 1000$ nodes for various a, b, ϵ. We define the overlap as:

$$Q = 1/2 - \frac{1}{n} \min\{d(\sigma, \hat{\sigma}), d(\sigma, -\hat{\sigma})\} > 0.$$

Fix a, b, we plot the overlap Q against ϵ and indicate the threshold $\tau = 1$ as a vertical dash line. All plotted values are averages over 100 trials.

Note that our algorithm is most efficient when the parameters (a, b, μ and ν) of the model are known as the optimal weight function depends on these parameters. In the case where the labels are uninformative, i.e. $\mu = \nu$, our algorithm is very simple, does not require to know the values a and b and is (in the range of Theorem 3.10) the algorithm with the best performance guarantees (see [31]).

3.2.7 Converse Result

In this section, we present the main steps for a proof of part (ii) in Conjecture 3.7: when $\tau < 1$, asymptotically it is impossible to tell whether any two nodes are more likely to belong to the same community. Since reconstructing a positively correlated type assignment is harder than telling whether any two nodes are more likely to belong to the same community, it further implies that reconstructing a positively correlated type assignment is fundamentally impossible.

Theorem 3.11. *If $\tau < 1$, then for any fixed vertices ρ and v,*

$$\mathbb{P}_n(\sigma_\rho = +1 | G, L, \sigma_v = +1) \to 1/2 \ \textit{a.a.s.} \tag{3.28}$$

Theorem 3.11 is related to the Ising spin model in the statistical physics [50, 80], and it essentially says that there is no long range correlation in the type assignment when $\tau < 1$. The main idea in the proof of Theorem 3.11 is borrowed from [81] and works as follows: (1) pick any two fixed vertices ρ, v and consider the local neighborhood of ρ up to distance $O(\log(n))$. The vertex v lies outside of the local neighborhood of ρ a.a.s.. (2) conditional on the type assignment at the boundary of the local neighborhood, σ_ρ is asymptotically independent with σ_v. (3) the local neighborhood of ρ looks like a Markov process on a labeled Galton-Watson tree rooted at ρ. (4) For the Markov process on the labeled Galton-Watson tree, the types of leaves provide no information about the type of the root ρ when the depth of tree goes to infinity.

3.2.8 Conclusion

Our results show that when $\tau < 1$ it is fundamentally impossible to give a positively correlated reconstruction; when $\tau > 1$, the labeling information can be effectively exploited through the suitably weighted graph. An interesting future work is to prove the positive part of Conjecture 3.7.

Bibliography

[1] Miklos Abért, Péter Csikvári, Péter Frenkel, and Gábor Kun. Matchings in benjamini-schramm convergent graph sequences. *arXiv preprint arXiv:1405.3271*, 2014.

[2] Miklós Abért, Péter Csikvári, and Tamás Hubai. Matching measure, benjamini-schramm convergence and the monomer-dimer free energy. *arXiv preprint arXiv:1405.6740*, 2014.

[3] R. Adamczak, A.E. Litvak, A. Pajor, and N. Tomczak-Jaegermann. Restricted isometry property of matrices with independent columns and neighborly polytopes by random sampling. *Constructive Approximation*, pages 61–88, 2011.

[4] R. Affentranger and R. Schneider. Random projections of regular simplices. *Discr. and Comput. Geometry*, 7:219–226, 1992.

[5] D. Aldous and R. Lyons. Processes on unimodular random networks. *Electronic Journal of Probability*, 12:1454–1508, 2007.

[6] David Aldous and J. Michael Steele. The objective method: probabilistic combinatorial optimization and local weak convergence. In *Probability on discrete structures*, volume 110 of *Encyclopaedia Math. Sci.*, pages 1–72. Springer, Berlin, 2004.

[7] Hamed Amini. Bootstrap percolation and diffusion in random graphs with given vertex degrees. *Electron. J. Combin.*, 17(1):Research Paper 25, 20, 2010.

[8] Hamed Amini, Moez Draief, and Marc Lelarge. Flooding in weighted sparse random graphs. *SIAM J. Discrete Math.*, 27(1):1–26, 2013.

[9] Hamed Amini and Marc Lelarge. Upper deviations for split times of branching processes. *J. Appl. Probab.*, 49(4):1134–1143, 2012.

[10] Hamed Amini and Marc Lelarge. The diameter of weighted random graphs. *Ann. Appl. Probab.*, under revision.

[11] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices*. Cambridge University Press, 2009.

[12] Z. Bai and J. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*. Springer, 2005.

[13] József Balogh and Boris G. Pittel. Bootstrap percolation on the random regular graph. *Random Structures Algorithms*, 30(1-2):257–286, 2007.

[14] Mohsen Bayati, Marc Lelarge, and Andrea Montanari. Universality in polytope phase transitions and message passing algorithms. *Ann. Appl. Probab.*, to appear.

[15] Edward A. Bender and E. Rodney Canfield. The asymptotic number of labeled graphs with given degree sequences. *J. Combinatorial Theory Ser. A*, 24(3):296–307, 1978.

[16] Itai Benjamini and Oded Schramm. Recurrence of distributional limits of finite planar graphs. *Electron. J. Probab.*, 6:no. 23, 13 pp. (electronic), 2001.

[17] S. Bhamidi and G Hooghiemstra R. van der Hofstad. First passage percolation on the Erdös-Rényi random graph. *Combinatorics, Probability & Computing*, 20(5):683–707, 2011.

[18] S. Bhamidi, R. van der Hofstad, and G. Hooghiemstra. Extreme value theory, Poisson-Dirichlet distributions and first passage percolation on random networks. *Advances in applied probability*, 42(3):706–738, 2010.

[19] S. Bhamidi, R. van der Hofstad, and G. Hooghiemstra. First passage percolation on random graphs with finite mean degrees. *Annals of Applied probability*, 20(5):1907–1965, 2010.

[20] Shankar Bhamidi, Remco van der Hofstad, and Gerard Hooghiemstra. Universality for first passage percolation on sparse random graphs. *arXiv preprint arXiv:1210.6839*, 2012.

[21] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.

[22] G Biroli and R Monasson. A single defect approximation for localized states on random lattices. *Journal of Physics A: Mathematical and General*, 32(24):L255–L261, 1999.

[23] Béla Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European J. Combin.*, 1(4):311–316, 1980.

[24] Béla Bollobás. *Random graphs*, volume 73 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2001.

[25] Erwin Bolthausen. An iterative construction of solutions of the tap equations for the sherrington-kirkpatrick model. *arXiv preprint arXiv:1201.2891*, 2012.

[26] Charles Bordenave. Lecture notes on random graphs and probabilistic combinatorial optimization!! draft in construction!! 2012.

[27] Charles Bordenave. Spectral measures of random graphs. 2014.

[28] Charles Bordenave and Marc Lelarge. Resolvent of large random graphs. *Random Structures Algorithms*, 37(3):332–352, 2010.

[29] Charles Bordenave, Marc Lelarge, and Justin Salez. The rank of diluted random graphs. *Ann. Probab.*, 39(3):1097–1121, 2011.

[30] Charles Bordenave, Marc Lelarge, and Justin Salez. Matchings on infinite graphs. *Probability Theory and Related Fields*, 157(1-2):183–208, 2013.

[31] Yudong Chen, Sujay Sanghavi, and Huan Xu. Clustering sparse graphs. In *Advances in neural information processing systems*, pages 2204–2212, 2012.

[32] Amin Coja-Oghlan. A spectral heuristic for bisecting random graphs. *Random Structures Algorithms*, 29(3):351–398, 2006.

[33] Amin Coja-oghlan. Graph partitioning via adaptive spectral techniques. *Comb. Probab. Comput.*, 19(2):227–284, 2010.

[34] Anne Condon and Richard M. Karp. Algorithms for graph partitioning on the planted partition model. *Random Struct. Algorithms*, 18(2):116–140, Mar 2001.

[35] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborova. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physics Review E*, 84:066106, 2011.

[36] J. Ding, J. Han Kim, E. Lubetzky, and Y. Peres. Diameters in supercritical random graphs via first passage percolation. *Combinatorics, Probability and Computing*, 19:729–751, 2010.

[37] D. L. Donoho. High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension. *Discrete Comput. Geom.*, page 617652, 2005.

[38] D. L. Donoho. Neighborly polytopes and sparse solution of underdetermined linear equations. Technical Report, Statistics Department, Stanford University, 2005.

[39] D. L. Donoho, A. Maleki, and A. Montanari. Message Passing Algorithms for Compressed Sensing. *Proceedings of the National Academy of Sciences*, 106:18914–18919, 2009.

[40] D. L. Donoho and J. Tanner. Neighborliness of randomly-projected simplices in high dimensions. *Proceedings of the National Academy of Sciences*, 102(27):9452–9457, 2005.

[41] D. L. Donoho and J. Tanner. Sparse nonnegative solution of underdetermined linear equations by linear programming. *Proceedings of the National Academy of Sciences*, 102(27):9446–9451, 2005.

[42] D. L. Donoho and J. Tanner. Counting faces of randomly projected polytopes when the projection radically lowers dimension. *Journal of American Mathematical Society*, 22:1–53, 2009.

[43] D. L. Donoho and J. Tanner. Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Phil. Trans. R. Soc. A*, pages 4273–4293, 2011.

[44] D.L. Donoho, A. Javanmard, and A. Montanari. Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing. *Information Theory, IEEE Transactions on*, 59(11):7434–7464, 2013.

[45] D.L. Donoho, I. Johnstone, and A. Montanari. Accurate prediction of phase transitions in compressed sensing via a connection to minimax denoising. *Information Theory, IEEE Transactions on*, 59(6):3396–3433, 2013.

[46] D.L. Donoho, A. Maleki, and A. Montanari. The Noise Sensitivity Phase Transition in Compressed Sensing. *IEEE Trans. on Inform. Theory*, 57:6920–6941, 2011.

[47] M.E Dyer and A.M Frieze. The solution of some random NP-hard problems in polynomial expected time. *Journal of Algorithms*, 10(4):451 – 489, 1989.

[48] Gabor Elek. On the limit of large girth graph sequences. *Combinatorica*, 30(5):553–563, 2010.

[49] P Erdős and T Gallai. Graphs with prescribed degrees of vertices (in Hungarian), Mat. *Lapok*, 11:264274, 1960.

[50] William Evans, Claire Kenyon, Yuval Peres, and Leonard J. Schulman. Broadcasting on trees and the ising model. *The Annals of Applied Probability*, 10(2):410–433, 2000.

[51] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3):75–174, 2010.

[52] M. R. Garey, D. S. Johnson, and L. Stockmeyer. Some simplified NP-complete graph problems. *Theoret. Comput. Sci.*, 1(3):237–267, 1976.

[53] C. D. Godsil. Matchings and walks in graphs. *J. Graph Theory*, 5(3):285–297, 1981.

[54] Ole J. Heilmann and Elliott H. Lieb. Theory of monomer-dimer systems. *Comm. Math. Phys.*, 25:190–232, 1972.

[55] Simon Heimlicher, Marc Lelarge, and Laurent Massoulié. Community detection in the labelled stochastic block model. Nov. 2012, available at: http://arxiv.org/abs/1209.2910.

[56] Vojkan Jaksic. Topics in spectral theory. In *Open Quantum Systems I*, pages 235–312. Springer, 2006.

[57] Svante Janson. The probability that a random multigraph is simple. *Combin. Probab. Comput.*, 18(1-2):205–225, 2009.

[58] Svante Janson, Malwina Luczak, and Peter Windridge. Law of large numbers for the sir epidemic on a random graph with given degrees. *Random Structures & Algorithms*, 45(4):724–761, 2014.

[59] Svante Janson and Malwina J. Luczak. A simple solution to the k-core problem. *Random Structures Algorithms*, 30(1-2):50–62, 2007.

[60] Svante Janson and Malwina J. Luczak. A new approach to the giant component problem. *Random Structures Algorithms*, 34(2):197–216, 2009.

[61] Svante Janson, Tomasz Łuczak, and Andrzej Rucinski. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[62] Adel Javanmard and Andrea Montanari. State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Information and Inference*, 2013.

[63] Mark Jerrum and Gregory B. Sorkin. The metropolis algorithm for graph bisection. *Discrete Applied Mathematics*, 82(1-3):155 – 175, 1998.

[64] Y. Kabashima, T. Wadayama, and T. Tanaka. A typical reconstruction limit for compressed sensing based on lp-norm minimization. *J.Stat. Mech.*, page L09003, 2009.

[65] Olav Kallenberg. *Foundations of modern probability*. Probability and its Applications (New York). Springer-Verlag, New York, second edition, 2002.

[66] Florent Krzakala, Marc Mézard, François Sausset, YF Sun, and Lenka Zdeborová. Statistical-physics-based reconstruction in compressed sensing. *Physical Review X*, 2(2):021005, 2012.

[67] Mathieu Leconte, Marc Lelarge, and Laurent Massoulié. Convergence of multivariate belief propagation, with applications to cuckoo hashing and load balancing. In Sanjeev Khanna, editor, *SODA*, pages 35–46. SIAM, 2013.

[68] M. Lelarge, L. Massoulie, and Jiaming Xu. Reconstruction in the labeled stochastic block model. In *Information Theory Workshop (ITW), 2013 IEEE*, pages 1–5, Sept 2013.

[69] Marc Lelarge. Diffusion and cascading behavior in random networks. *Games Econom. Behav.*, 75(2):752–775, 2012.

[70] Marc Lelarge. A new approach to the orientation of random hypergraphs. In Yuval Rabani, editor, *SODA*, pages 251–264. SIAM, 2012.

[71] Marc Lelarge. Bypassing correlation decay for matchings with an application to xorsat. In *Information Theory Workshop (ITW), 2013 IEEE*, pages 1–5. IEEE, 2013.

[72] Marc Lelarge. Loopy annealing belief propagation for vertex cover and matching: convergence, LP relaxation, correctness and bethe approximation. *CoRR*, abs/1401.7923, 2014.

[73] Marc Lelarge and Hang Zhou. Sublinear-time algorithms for monomer-dimer systems on bounded degree graphs. *Theor. Comput. Sci.*, 548:68–78, 2014.

[74] A. Maleki, L. Anitori, A. Yang, and R. Baraniuk. Asymptotic Analysis of Complex LASSO via Complex Approximate Message Passing (CAMP). arXiv:1108.0477, 2011.

[75] Laurent Massoulié and Dan-Cristian Tomozei. Distributed user profiling via spectral methods. Sep. 2011, available at: http://arxiv.org/abs/1109.3318.

[76] F. McSherry. Spectral partitioning of random graphs. In *42nd IEEE Symposium on Foundations of Computer Science*, pages 529 – 537, Oct. 2001.

[77] Marc Mézard and Andrea Montanari. *Information, physics, and computation*. Oxford University Press, USA, 2009.

[78] Marc Mézard, Giorgio Parisi, and Miguel Angel Virasoro. *Spin glass theory and beyond*, volume 9. World scientific Singapore, 1987.

[79] Michael Molloy and Bruce Reed. A critical point for random graphs with a given degree sequence. *Random Structures Algorithms*, 6(2-3):161–179, 1995.

[80] E. Mossel. Survey - information flows on trees. *DIMACS series in discrete mathematics and theoretical computer science*, pages 155–170, 2004.

[81] E. Mossel, J. Neeman, and A. Sly. Stochastic block models and reconstruction. Feb. 2012, available at: http://arxiv.org/abs/1202.1499.

[82] S. Rangan. Generalized Approximate Message Passing for Estimation with Random Linear Mixing. In *IEEE Intl. Symp. on Inform. Theory*, St. Perersbourg, August 2011.

[83] S. Rangan, A. K. Fletcher, and V. K. Goyal. Asymptotic analysis of map estimation via the replica method and applications to compressed sensing. In *Neural Information Processing Systems (NIPS)*, Vancouver, 2009.

[84] Michael Reed and Barry Simon. *Methods of modern mathematical physics. vol. 1. Functional analysis*, volume 2. Academic press New York, 1972.

[85] G. J. Rodgers and A. J. Bray. Density of states of a sparse random matrix. *Phys. Rev. B (3)*, 37(7):3557–3562, 1988.

[86] Karl Rohe, Sourav Chatterjee, and Bin Yu. Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, 39(4):1878–1915, 2011.

[87] Justin Salez. Some implications of local weak convergence for large random graphs. *Ph.D. Thesis.*

[88] Justin Salez. Weighted enumeration of spanning subgraphs in locally tree-like graphs. *Random Structures & Algorithms*, 43(3):377–397, 2013.

[89] P. Schniter. Turbo Reconstruction of Structured Sparse Signals. In *Proceedings of the Conference on Information Sciences and Systems*, Princeton, 2010.

[90] Philip Schniter and Sundeep Rangan. Compressive phase retrieval via generalized approximate message passing. In *Communication, Control, and Computing (Allerton), 2012 50th Annual Allerton Conference on*, pages 815–822. IEEE, 2012.

[91] G Semerjian and L F Cugliandolo. Sparse random matrices: the eigenvalue spectrum revisited. *Journal of Physics A: Mathematical and General*, 35(23):4837–4851, 2002.

[92] Gerard Sierksma and Han Hoogeveen. Seven criteria for integer sequences being graphic. *J. Graph Theory*, 15(2):223–231, 1991.

[93] T. Tao and V. Vu. Random matrices: The Universality phenomenon for Wigner ensembles. arXiv:1202.0068, 2012.

[94] DJ Thouless, PW Anderson, and RG Palmer. Solution of'solvable model of a spin glass'. *Philosophical Magazine*, 35(3):593–601, 1977.

[95] Remco Van Der Hofstad. Random graphs and complex networks. *Available on http://www. win. tue. nl/rhofstad/NotesRGCN. pdf*, 2009.

[96] A. M. Vershik and P. V. Sporyshev. Asymptotic behavior of the number of faces of random polyhedra and the neighborliness problem. *Selecta Math. Soviet.*, 11:181201, 1992.

[97] David G. Wagner. Weighted enumeration of spanning subgraphs with degree constraints. *J. Combin. Theory Ser. B*, 99(2):347–357, 2009.

[98] Duncan J. Watts. A simple model of global cascades on random networks. *Proc. Natl. Acad. Sci. USA*, 99(9):5766–5771 (electronic), 2002.

[99] Lenka Zdeborová and Marc Mézard. The number of matchings in random graph. *Journal of Statistical Mechanics*, 2006(5):P05003, 2006.