Large Discrete Structures

Statistical Inference, Combinatorics and Limits

Dissertation

zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich 12
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Maximilian Grischa Hahn-Klimroth
aus Bad Soden

Frankfurt 2021
(D 30)

vom Fachbereich Informatik und Mathematik der

Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan:

Prof. Dr. Lars Hedrich

Gutachter:

Prof. Dr. Amin Coja-Oghlan

Prof. Dr. Uriel Feige

Datum der Disputation: 11.05.2021

Contents

Acknowledgement

Throughout my doctoral studies that ultimately led to the writing of this dissertation I have received a great deal of support.

At first, I would like to express my deep gratitude to my first supervisor, Prof. Dr. Coja-Oghlan, for his guidance and continued mentorship as well as fruitful collaboration on an equal footing. I am especially thankful that he gave me the opportunity to be part of the active research community by, for instance, letting me attend various conferences and workshops. Not to forget, he always had an open door and gave me helpful advice as well as honest feedback. Also, I appreciate his tireless efforts to sweeten our joint lunch breaks which always contained interesting and relaxed conversations. Furthermore, I am grateful for him making sure that all of us were able to enjoy the sights and landmarks around the aforementioned conference venues.

Second, I thank my second supervisor, Prof. Dr. Yury Person, for his guidance and mentorship. While being physically separated, he made sure that we stayed in contact, by, for instance, inviting me multiple times to Ilmenau. Moreover, it was he who initially excited me for the subject of probabilistic and extremal combinatorics during my Bachelor's studies and introduced me to the "Stiftung Polytechnische Gesellschaft" during my Master's degree. The latter institution played a key role in my academic career.

More precisely, this dissertation would not have been possible without the funding of Stiftung Polytechnische Gesellschaft Frankfurt am Main. Besides this financial support, I thank especially Tobias König for his never ending interest and support in a variety of ways. He meticulously organised some of the best seminar courses I could ever imagine which brought me a big step forward in terms of personal and professional skills.

Furthermore, I especially thank Prof. Dr. Uriel Feige whose very detailed comments led to numerous corrections of one of my publications and increased its quality significantly.

Additionally, I would like to thank all my esteemed colleagues in Frankfurt: Jean Bernoulli Ravelomanana, Joon Lee, Maurice Rolvien, Michèle Fellinghauer, Noëla Müller, Oliver Gebhard and Philipp Loick who made my time as a PhD student unforgettable. I thank you for our productive joint research as well as many stimulating conversations and welcome distractions to rest my mind. In especially, I thank my former office mate Philipp Loick that I could always count on his support. Furthermore, I am really glad having found two coffee lovers amongst my colleagues, Joon Lee and Oliver Gebhard, with whom I could enjoy short breaks on our joint office days. Last but not least, I especially thank Oliver Gebhard for his touristic guidance on our various joint trips. Moreover, I want to thank Olaf Parczyk for his patience in our joint projects and especially for giving me the possibility to visit him in London.

I am very grateful that my mother, Inge Hahn-Klimroth, gave me the opportunity to study mathematics in the first place and supported me throughout my whole academic training. Further, I thank Jennifer Gübert for her emotional support while I was writing this thesis.

Finally, I want to thank Jennifer Gübert, Inge Hahn-Klimroth, Joon Lee and Philipp Loick for proofreading parts of this thesis.

Abstract

Studying large discrete systems is of central interest in, non-exclusively, discrete mathematics, computer sciences and statistical physics. The study of *phase transitions*, e.g. points in the evolution of a large random system in which the behaviour of the system changes drastically, became of interest in the classical field of random graphs, the theory of spin glasses as well as in the analysis of algorithms [78, 82, 121].

It turns out that ideas from the statistical physics' point of view on spin glass systems can be used to study inherently combinatorial problems in discrete mathematics and theoretical computer sciences (for instance, satisfiability) or to analyse phase transitions occurring in inference problems (like the group testing problem) [68, 135, 168]. A mathematical flaw of this approach is that the physical methods only render mathematical conjectures as they are not known to be rigorous.

In this thesis, we will discuss the results of six contributions. For instance, we will explore how the theory of *diluted mean-field models* for spin glasses helps studying random constraint satisfaction problems through the example of the random 2−SAT problem. We will derive a formula for the number of satisfying assignments that a random 2−SAT formula typically possesses [2].

Furthermore, we will discuss how ideas from spin glass models (more precisely, from their *planted* versions) can be used to facilitate inference in the group testing problem. We will answer all major open questions with respect to non-adaptive group testing if the number of infected individuals scales sublinearly in the population size and draw a complete picture of phase transitions with respect to the complexity and solubility of this inference problem [41, 46].

Subsequently, we study the group testing problem under sparsity constrains and obtain a (not fully understood) phase diagram in which only small regions stay unexplored [88].

In all those cases, we will discover that important results can be achieved if one combines the rich theory of the statistical physics' approach towards spin glasses and inherent combinatorial properties of the underlying random graph.

Furthermore, based on partial results of Coja-Oghlan, Perkins and Skubch [42] and Coja-Oghlan et al. [49], we introduce a consistent limit theory for discrete probability measures akin to the graph limit theory [31, 32, 128] in [47]. This limit theory involves the extensive study of a special variant of the cut-distance and we obtain a continuous version of a very simple algorithm, the pinning operation, which allows to decompose the phase space of an underlying system into parts such that a probability measure, restricted to this decomposition, is close to a product measure under the cut-distance. We will see that this *pinning lemma* can be used to rigorise predictions, at least in some special cases, based on the physical idea of a *Bethe state decomposition* when applied to the Boltzmann distribution.

Finally, we study sufficient conditions for the existence of perfect matchings, Hamilton cycles and bounded degree trees in randomly perturbed graph models if the underlying deterministic graph is sparse [93].

1. Introduction

Large discrete systems play a central role in discrete mathematics, theoretical computer sciences, statistical physics as well as in statistics. Analysing such systems became of major interest in the last decades. Prominent examples are the study of phase transitions on classical random graphs, the modelling and analysis of spin glasses, creating and analysing limit theories of discrete structures (like the graph limit theory) as well as the asymptotic analysis of algorithms. While this list is far from being complete, researchers from different fields realised that the interdisciplinary application of certain methods can be used very profitably [135, 168].

This thesis gives examples in which methods inspired by the statistical physics analysis' of large spin glass systems are used to provide rigorous mathematical understanding of statistical inference problems as well as a classical random constraint satisfaction problem. The richness provided by this approach is due to combining ideas from physics like message passing with combinatorial arguments and interpretations. Different combinatorial insights will be used to study large perturbed graphs, thus deterministic graphs where a bit of randomness is added.

Finally, a rigorous mathematical approach to certain physics' intuitions is carried out by analysing a particular form of the cut-distance for discrete probability measures. We will create a theory of limiting objects comparable to the well known and rich theory of graph limits [31, 32, 128] and formalise the intuition of basic statistical physics' concepts like *pure states*. This limiting theory will provide an elegant algorithmic regularity lemma for probability measures which can be translated into an algorithmic version of the weak regularity lemma for graphs [84].

1.1. Message passing and a statistical physics' approach

In this section, we will briefly introduce some of the most basic concepts of statistical physics that will be used to describe mathematical problems in an elegant and uniform way. We will, as [135], describe statistical physics as a part of probability theory, and will only sometimes give a meaningful interpretation of the concepts in nature.

1.1.1. Physical systems and important quantities

Let Ω be a finite set and n be the size of a physical system. Then we say that this system contains n *particles* and call Ω^n the *configuration space*. Thus, a *configuration* $\sigma \in \Omega^n$ assigns each particle i a *spin* $\sigma_i \in \Omega$. Furthermore, for $k \geq 1$, we define the *Hamiltonian* (or *energy*) of a configuration $\sigma \in \Omega^n$ as a function

$$H(\sigma) = -\sum_{i_1,\dots,i_k} J_{i_1,\dots,i_k}(\sigma_{i_1},\dots,\sigma_{i_k}). \tag{1.1.1}$$

In the special case $k = 1$, the system is called a *non-interacting* system as the different particles do not interact. Conversely, if $k \geq 2$, the system is called *$k-$body interacting*.

The real numbers $J_{i_1,\dots,i_k}$ express the interactions of the particles. This setup is quite general and allows for a lot of modifications, i.e., for each choice of Ω, k and H a different system is described. Some systems of this family of systems are well known physical models, for example for magnetism or for glasses. Depending on the choice of the interaction between its particles, a model is called *ferromagnetic*, *anti-ferromagnetic* or a *spin glass*. Intuitively, a ferromagnetic system prefers the interaction of particles with the same spin while an anti-ferromagnetic system prefers the opposite. Finally, in a spin glass, we find ferromagnetic as well as anti-ferromagnetic interactions.

Some prominent models that fit into this generic setup are the Potts model [16] or the Edwards-Anderson model [69]. The former is frequently used to study phase transitions – we will come to phase transitions later – or to model systems with (easy) nearest neighbour interactions while the latter is a

widely accepted mathematical model for magnetism [135]. Both models are defined on the d-dimensional grid $L = (V, E)$ as an underlying graph structure. The configuration space of the Potts model is $\Omega_{\text{Potts}} = \{1, 2, ..., q\}$ and it is a 2−body interacting system with Hamiltonian

$$H_{\text{Potts}}(\sigma) = -\sum_{ij \in E} J\mathbf{1}\{\sigma_i = \sigma_j\}. \tag{1.1.2}$$

Therefore, if two neighboured particles have the same spin under a configuration σ, the correspondent summand vanishes. Clearly, for $J > 0$, this system is ferromagnetic and for $J < 0$ it is anti-ferromagnetic. On the other hand, the Edwards-Anderson model has only two spins (negative and positive), thus $\Omega_{\text{EA}} = \{-1, +1\}$. Furthermore, as it is a model for magnetism, it adds the possibility of the appearance of some external magnetic field of strength $B > 0$. Moreover, its Hamiltonian allows for different interactions between two neighboured spins, depending on where they are placed inside the system, therefore

$$H_{\text{EA}}(\sigma) = -\sum_{ij \in E} J_{ij}\sigma_i\sigma_j - B\sum_{i \in V} \sigma_i. \tag{1.1.3}$$

Again, depending on the choice of J_{ij} being positive or negative for all edges ij, the model is ferromagnetic or anti-ferromagnetic, and when allowing different signs, it is a spin glass model. In such cases, where the interaction between particles can be written as in (1.1.3), the interactions J_{ij} are called *coupling constants*. The most prominent variant of the Edwards-Anderson model is the spin glass case in which the coupling constants are chosen from a symmetric probability distribution, for instance as standard Gaussians. In contrast, if we set $J_{ij} > 0$ for all coupling constants, we get the ferromagnetic *Ising model* on the grid as a special case [100]. Analogously, setting all coupling constants to a negative value renders the anti-ferromagnetic Ising model.

Now, to capture the idea behind a spin glass, an important physical observation is that any system strives for being in a state of minimal energy. In the anti-ferromagnetic or ferromagnetic Ising model it is quite easy to construct the configuration σ minimising $H(\sigma)$. But in the spin glass situation of the general Edwards-Anderson model it happens that a particle receives contradicting constraints from its neighbours. If this happens, we call the system *frustrated* at this particle. It is computationally hard to find a configuration that minimises the energy of the system [22]. On the other hand, there are many configurations with comparable low energy that *almost* minimise the system's energy – those states are called *metastable* [167]. Besides being mathematically challenging, spin glasses are a very good model for many real world materials, e.g. window glass or polymers or even granular media. Recent models like spin glasses for these materials assume that certain atoms and molecules occur randomly at random positions, thus the single particles behave either anti-ferromagnetically or ferromagnetically or are frustrated. The emerging theory of spin glasses allows to analyse those materials although they are random when seen microscopically, because on a macroscopic scale, they show describable properties [167].

1.1.1.1. Mean field and diluted mean field models

The interactions in the Potts model as well as the Edwards-Anderson model are defined on the grid. Therefore, those models have direct physical interpretations. The Edwards-Anderson model, for instance, is a realistic mathematical model for magnetism [135]. Unfortunately, the interactions facilitating this geometric constraints are challenging to analyse mathematically. One approach of simplification are so-called *mean-field* models. A well known example is the Sherrington-Kirkpatrick model (SK-model) [158] in which all particles $x_1 \dots x_n$ of a system interact with each other. More precisely, with $\Omega = \{-1, 1\}$, and $(\boldsymbol{J}_{ij})_{i,j=1\dots n}$ being standard Gaussians, its Hamiltonian reads

$$H_{\text{SK}}(\sigma) = -\frac{1}{\sqrt{n}}\sum_{i,j=1}^{n} \boldsymbol{J}_{ij}\sigma_i\sigma_j. \tag{1.1.4}$$

An important feature of the SK-model (as well as of general mean-field models) is the fact that the (distribution) of the Hamiltonian is invariant under the permutation of the coordinates, thus the geometric constraints of similar models on the grid (like the Edwards-Anderson model) vanish in the mean-field theory. The way mean-field models are defined, they neglect local and long-range structure, thus they are not able to describe all physically necessary properties of a system properly [157]. On the other hand, the models are somewhat easy, thus studying mean-field models is physically as well as mathematically more accessible than analysing their corresponding grid-models. For instance, the physical predictions of the properties of the SK-model [143] were proven rigorously by Talagrand [162].

Diluted mean-field models try to overcome the weakness of mean-field models while staying accessible to mathematical analyses. We introduce the diluted version of the SK-model by Viana and Bray [166]. Let $\Omega = \{-1, 1\}$ and $\boldsymbol{k} \sim \mathbf{Po}(\alpha n)$ be a Poisson random variable. Given $\boldsymbol{k}$, let $\{\boldsymbol{i}_k, \boldsymbol{j}_k\}_{k=1\dots\boldsymbol{k}}$ be uniform samples from $\{1, \dots, n\}$ and choose $\{\boldsymbol{J}_k\}_{k=1\dots\boldsymbol{k}}$ from a symmetric distribution. All those random variables are supposed to be mutually independent. Then, the Hamiltonian of the Viana-Bray model reads

$$H_{\mathrm{VB}}(\sigma) = -\sum_{k=1}^{\boldsymbol{k}} \boldsymbol{J}_k \sigma_{\boldsymbol{i}_k} \sigma_{\boldsymbol{j}_k}. \tag{1.1.5}$$

Thus, the Viana-Bray interactions are defined on a sparse random graph where each particle interacts with a Poisson number of other particles. Therefore, the long-range structure of a grid-model is clearly missing, but in contrast to the mean-field approach, the random graph looks locally a bit more like a grid, i. e., a very important feature is the finite connectivity on the random graph [134]. It turns out that analysing appropriately chosen diluted mean-field models may yield exact solutions of spin glass like models [144]. Thus, diluted mean field models carry important features of realistic models but are still mathematically approachable.

Furthermore, it turns out that very important problems in theoretical computer sciences, like random satisfiability and other random constraint satisfaction problems, can by expressed in the framework of diluted mean field models [140]. We will introduce random constraint satisfaction problems in Section 1.1.2.

After this short excursion to some prominent examples of particle systems, we formalise the intuition behind finding a configuration of minimal energy within a large system. We may think of any physical system with a Hamiltonian defined as in (1.1.1), but we will focus on diluted mean-field models.

1.1.1.2. Boltzmann distributions and ground states

If a system is observed showing a configuration of minimal energy, we will call this state a *ground state*. We denote by $\Omega_0^n \subset \Omega^n$ the set of configurations of minimal energy. Let us define a probability measure on the configuration space. Intuitively, a suitable probability measure should output the probability of observing a certain configuration, preferring those configurations of low energy.

Therefore, let $\beta > 0$ be the *inverse temperature* of the system. Having fixed an energy function H and an inverse temperature β, we define the *Boltzmann distribution* as a probability distribution on Ω^n as

$$\mu_\beta(\sigma) = \frac{\exp\left(-\beta H(\sigma)\right)}{Z_\beta}, \qquad \text{where} \qquad Z_\beta = \sum_{\sigma \in \Omega^n} \exp\left(-\beta H(\sigma)\right). \tag{1.1.6}$$

In (1.1.6), the normalising constant Z_β is known as the *partition function* of the system and we will see in due course that the partition function itself carries a lot of information about the system. Clearly, the lower the Hamiltonian $H(\sigma)$, the more probable is σ under μ_β. This effect decreases with β being small (the temperature of the system is large) and increases with β being large (the system's energy is small). More precisely, with $\beta \to 0$ (high-temperature limit), the Boltzmann distribution becomes the uniform distribution on Ω^n. On the other hand, in the low-temperature limit, the Boltzmann distribution corresponds to the uniform distribution on the ground states, thus on the configurations minimising the

system's energy. Formally,

$$\lim_{\beta\to 0}\mu_\beta(\sigma)=\frac{1}{|\Omega^n|},\qquad\text{and}\qquad\lim_{\beta\to\infty}\mu_\beta(\sigma)=\frac{\mathbf{1}\{\sigma\in\Omega_0^n\}}{\left|\Omega_0^n\right|}. \tag{1.1.7}$$

As statistical physics' main intention is to study the macroscopic behaviour of very large systems consisting of single particles and their interactions, various properties of a statistical system are studied in the *thermodynamical limit* ($n\to\infty$) [135]. Such a physical system can be described by various thermodynamic quantities. In the following, we will introduce the most important ones for our purposes. As already mentioned, Z_β is the partition function of the system. We denote by $\phi_{n,\beta}=\ln(Z_\beta)$ the *free entropy* and by $\phi_\beta=\lim_{n\to\infty}\frac{\ln(Z_\beta)}{n}$ the *free entropy density*. For any $\beta>0$, ϕ_β is convex, therefore it is continuous in every point in which it exists. We call the non-analytic points of ϕ_β *phase transitions*. Of course, non-analytical points are interesting mathematical objects, but physically speaking, those phase transition points indicate qualitative changes in the physical system [135].

In particular, those phase transitions corresponding to the so-called *replica symmetry breaking* are of deeper interest and will be studied later.

1.1.2. Phase transitions in special systems: Random Constraint Satisfaction

In the very general setup of a physical system with a given energy function (1.1.1), it is possible to describe constraint satisfaction problems (CSPs) which are prominently studied in computer sciences and mathematics. One of the most important constraint satisfaction problem is the $k-SAT$ problem.

1.1.2.1. $k-$SAT and factor graphs

A $k-SAT$ formula Φ is a conjunction of m clauses

$$\Phi=\Phi_1\wedge\ldots\wedge\Phi_m$$

such that each clause is a disjunction of exactly k literals out of n variables $x_1\ldots x_n$. One of the most intriguing questions is, obviously, whether there is a mapping $\sigma:\{x_1,\ldots,x_n\}\to\{-1,+1\}^n$ that assigns the Boolean values TRUE (+1) and FALSE (-1) to each variable such that each clause (and therefore the formula Φ) is satisfied. Without loss of generality, we suppose that a variable appears only once in a clause (i.e., x and $\neg x$ do never belong to one clause). Let us describe a $k-$SAT formula as a *factor graph*.

A factor graph $G=(V\cup F,E)$ is a bipartite graph with vertex classes V (variable nodes) and F (factor nodes) and edges E [81]. Sometimes, it is convenient to call the factor nodes *constraints*. We follow Mézard and Montanari [135] for construction of a factor graph G^Φ corresponding to a $k-$SAT formula Φ. Define

$$V=\{x_1,\ldots,x_n\}\qquad\text{and}\qquad F=\left\{a_1^\Phi,\ldots,a_m^\Phi\right\}.$$

Furthermore, we introduce two different types of edges $E=E^+\cup E^-$ such that $x_i a_j^\Phi\in E^+$ if and only if x_i appears in Φ_j and $x_i a_j^\Phi\in E^-$ if and only if $\neg x_i$ appears in Φ_j. The resulting factor graph has a fixed degree of k at each factor node while the variable node degree is not fixed in general.

While deciding whether a given $k-$SAT formula is satisfiable is known to be $\boldsymbol{NP}$ hard for $k\geq 3$ [109], it is easily possible to describe the problem as a physical system. Similarly as Kirkpatrick et al. [111], we interpret the n variables as particles of a system and assign each particle a spin from $\Omega=\{-1,+1\}$. A specific assignment is expressed by a configuration $\sigma\in\Omega^n$. The Hamiltonian turns out to be

$$H_{k-SAT}(\sigma)=\sum_{a_j^\Phi\in F}\mathbf{1}\left\{a_j^\Phi\text{ is not satisfied under }\sigma\right\}=\sum_{a_j^\Phi\in F}\mathbf{1}\left\{\sum_{x_i a_j^\Phi\in E^+}\sigma_i-\sum_{x_i a_j^\Phi\in E^-}\sigma_i=0\right\}.$$

Therefore, $H_{k-SAT}(\sigma)$ equals the number of unsatisfied clauses of Φ under σ. Let us now take the low-temperature limit and observe that the resulting Boltzmann distribution $\mu_\infty=\lim_{\beta\to\infty}\mu_\beta$ is the uniform

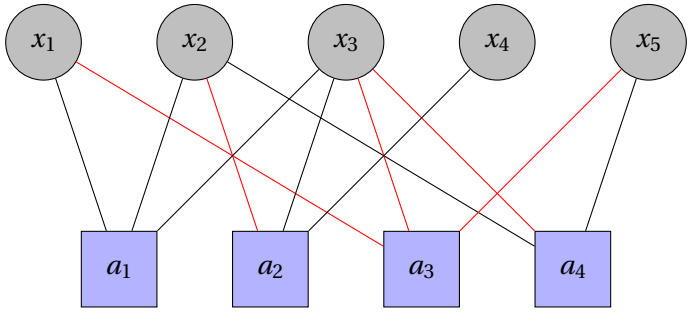

Figure 1.1.: The factor graph G^Φ corresponding to the 3−SAT formula $\Phi : (x_1 \vee x_2 \vee x_3) \wedge (\neg x_2 \vee x_3 \vee x_4) \wedge (\neg x_1 \vee \neg x_3 \vee \neg x_5) \wedge (x_2 \vee \neg x_3 \vee x_5)$. The $n = 5$ variable nodes are represented as circles while the $m = 4$ factor nodes are drawn as rectangles. The edge color represents the sign of the literal x_i in clause a_j. Thus, $s_{ij} = -1$ if edge $x_i a_j$ is red and $s_{ij} = 1$ otherwise. At each factor node a_j we have a local function $\Psi_{a_j} : \{-1,+1\}^{|\partial a_j|} \to \{-1,+1\}$ such that we have $\Psi_{a_j}(\sigma_{\partial a_j}) = \mathbf{1}\left\{\max_{x_i \in \partial a_j} \{\sigma_i s_{ij} = 1\}\right\}$ for an assignment $\sigma \in \{-1,+1\}^5$.

distribution on all configurations that satisfy the most clauses of Φ. Now let $\boldsymbol{\sigma} \sim \mu_\infty$ be a random sample, then the formula Φ is satisfiable if and only if $H_{k-SAT}(\boldsymbol{\sigma}) = 0$. The ***NP*** hardness of k−SAT for $k \geq 3$ immediately confirms the fact that it is in general computationally hard to find configurations of minimal energy in a physical system. We observe that

$$\mu_\infty(\sigma) = \lim_{\beta \to \infty} \frac{\prod_{a_j^\Phi \in F} \exp\left(-\beta \mathbf{1}\left\{a_j^\Phi \text{ is not satisfied under } \sigma\right\}\right)}{Z(\Phi)}. \tag{1.1.8}$$

The possibility to write the Boltzmann distribution in a factorised form as in (1.1.8) yields a few insights that directly generalise to other constraint satisfaction problems.

- Each factor of (1.1.8) corresponds to a factor node in the factor graph G^Φ, therefore, the factor graph perfectly describes the factorisation of the Boltzmann distribution into local constraints.
- For each $\beta > 0$ a non-satisfied clause gives a penalty of $\exp(-\beta)$ to the probability of observing a certain configuration. Given that Φ is satisfiable, in the low-temperature limit (or sometimes called *at zero temperature*), the Boltzmann distribution is supported on the satisfying assignments.
- Therefore, given that a formula is satisfiable, the partition function of the Boltzmann distribution in the low-temperature limit of the k-SAT problem ($Z_0(\Phi) = \left|\Omega_0^n\right|$) just equals the number of satisfying assignments and is an object of major interest in solving CSPs.

1.1.2.2. Random CSPs

In the general context of CSPs *random* constraint satisfaction problems gained a lot of attention [156]. Let us briefly sketch, what *random* means in this context. As we already discussed, a CSP can be expressed by a factor graph. Thus, given n variables and $\boldsymbol{m}$ factors (where the number of factors might be random itself), we randomly connect variable nodes and factor nodes. Depending on the problem itself, we can have arbitrary distributions of degree sequences for both, factor nodes as well as variable nodes. Given specific degree sequences, a random factor graph is chosen uniformly at random among all possible factor graphs satisfying the degree sequences.

Let us explain this very general description as before with the example of a random k-SAT formula. Given n variables and m clauses, initialise a factor graph on n variable nodes $x_1, \ldots, x_n$ and m factor nodes $a_1, \ldots, a_m$. Each factor node has degree k and variable x_i has a random degree $\boldsymbol{d}_i \sim \mathbf{Po}(mk/n)$.

Given the event that $\sum_{i=1}^{n} \boldsymbol{d}_i = mk$, select one (simple) graph with the given degree sequences uniformly at random.

In this setting, a very natural question arises: Is a formula obtained by this process satisfiable? Obviously, such questions can only be answered with high probability, thus with a probability tending to 1 with $n \to \infty$. This question has been studied for a long time and various tools of statistical physics have been applied to this problem [37, 111] that led to a (non-rigorously proven) prediction for a critical ratio $\alpha_s = m_s/n$, such that in the thermodynamic limit, each random $k-SAT$ formula with a smaller clause-to-variable ratio than α_s is satisfiable with high probability, whilst each formula with a larger ratio is not satisfiable with high probability. Such a phenomenon is called a *phase transition* and will be further discussed in a moment. This conjecture attained a lot of attention within the mathematical community and a lot of important steps towards proving this conjecture were done [4, 51, 54, 91] until Ding, Sly and Sun [64] managed to prove the existence of the phase transition for large enough k at

$$\alpha_s = 2^k \ln 2 - \frac{1+\ln 2}{2} + O(2^{-k}).$$

The existence of such a *satisfiability threshold* is not limited to the random k-SAT but is a genuine phenomenon of random constraint satisfaction problems [136]. The satisfiability threshold, however, is not the only interesting threshold regarding random CSPs. Indeed, let $\mathscr{S}$ be the solution space of a random constraint satisfaction problem (we may think about $\mathscr{S}$ being the set of all configurations satisfying a random k-SAT formula). We say that a pair of solutions is connected if its Hamming distance equals 1 and call a subset of Ω^n consisting of connected solutions a *cluster*. It turns out that the geometry of $\mathscr{S}$ has a highly complicated structure, but fortunately, the so-called *1-Replica Symmetry Breaking (1-RSB) Ansatz* from statistical physics draws a non-rigorously proven but fine grained picture. We will discuss this Ansatz in a moment. It conjectures that, with growing clause-to-variable ratio α, the solution space $\mathscr{S}$ undergoes four phase transitions (see Figure 1.2). While being a non-rigorous tool, the existence of the predicted phase transitions could be already proven rigorously for some models [161]. We let α start at 0 and let it increase continuously, then we observe four critical values $\alpha_u \le \alpha_{\text{clus}} \le \alpha_{\text{cond}} \le \alpha_s$ at which $\mathscr{S}$ changes dramatically [121, 138, 169, 170].

1. For $\alpha < \alpha_u$, there is exactly one cluster of solutions. Therefore, this phase is called the *unique* phase.

2. Once α exceeds α_u, the system is in the *extremal* phase. (Very) few and exponentially small clusters of satisfying configurations appear besides one cluster containing almost all solutions.

3. At the *clustering* threshold α_{clus}, the set of solutions shatters into exponentially many exponentially small clusters. Withing this phase, the size and and number of clusters reduces further with α increasing.

4. Subsequently, when α passes the *condensation* threshold, the solutions condense into a bounded number of clusters.

5. Finally, at the satisfiability threshold α_s, $\mathscr{S}$ becomes empty.

While an intuitive explanation what a phase transition is was already given, until now, we lacked a formal definition. We distinguish between *strict* and *coarse* phase transitions in random discrete systems. Let $\mathscr{P}$ be a property, and α be a parametrisation of the system. For instance, $\mathscr{P}$ could be the property that a random k-SAT formula is unsatisfiable and α is the clause-to-variable ratio. Then the system (the random factor graph) $\mathscr{G}$ undergoes a strict phase transition at a threshold α^* if for any $\varepsilon > 0$ we have

$$\mathbb{P}\left(\mathscr{G} \in \mathscr{P} \mid \alpha \le (1-\varepsilon)\alpha^*\right) = o(1) \qquad \text{and} \qquad \mathbb{P}\left(\mathscr{G} \in \mathscr{P} \mid \alpha \ge (1+\varepsilon)\alpha^*\right) = 1 - o(1).$$

Similarly, a coarse phase transition occurs if

$$\mathbb{P}\left(\mathscr{G} \in \mathscr{P} \mid \alpha = o(\alpha^*)\right) = o(1) \qquad \text{and} \qquad \mathbb{P}\left(\mathscr{G} \in \mathscr{P} \mid \alpha = \omega(\alpha^*)\right) = 1 - o(1).$$

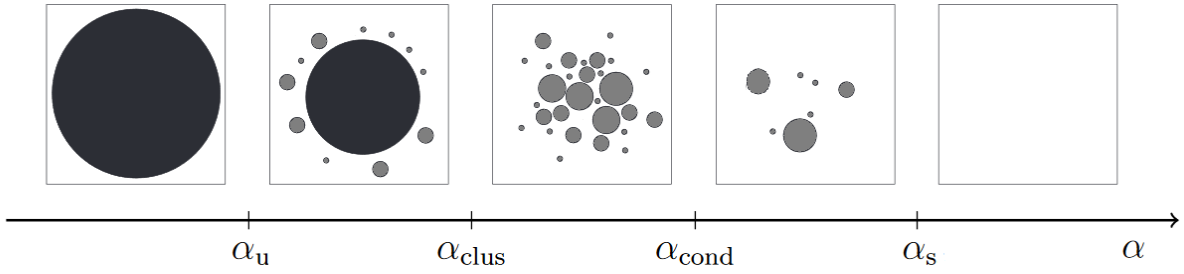

Figure 1.2.: Overview over the geometry of the solution space in random constraint satisfaction problems undergoing four important phase transitions obtained by the physics' 1-RSB Ansatz. The graphic is modified after [138, 169, 170].

(Clearly, the definition above requires $\mathscr{P}$ to become more probable for increasing α, it can analogously be defined for a property whose probability increases with decreasing α.)

Due to Friedgut and Bourgain [82] it is very well understood that non-uniform strict phase transitions exist in random structures if we look at monotonously increasing or decreasing properties. For example, the property for a random $k-$SAT formula to be unsatisfiable is clearly of this kind, as adding more clauses only increases the probability to be unsatisfiable. Of course, Friedgut and Bourgain's result only tells us that there *are* phase transitions, but it cannot tell us, *where* they are. For many graph properties regarding the random graph $\mathscr{G}(n,p)$[1] the exact position of the phase transitions is known [103]. For random constraint satisfaction problems, things are much more unexplored, even if recently the satisfiability thresholds could be pinned down rigorously for many random CSPs [25, 43, 55, 64]. The large temporal gap between the beginnings of studying random CSPs and the discovery of the exact satisfiability thresholds might be due to the reason that a simple first-moment calculation overestimates the number of satisfying assignments in many random CSPs dramatically, thus $\mathbb{E}[Z(\mathbf{\Phi})] \gg Z(\mathbf{\Phi})$ w.h.p. [55]. In the statistical physics' interpretation, this is due to the existence of the condensed phase in which the solution space is dominated by few large clusters. Therefore, the expected number of solutions is blown up by solutions that do not occur in typical formulas frequently and therefore, obviously, the first moment method that calculates the critical value α_{avg} such that $\mathbb{E}[Z(\mathbf{\Phi})] = 0$, fails.

Dealing with such effects is non-trivial, but the physics' 1-RSB Ansatz is a very promising (non-rigorous) approach towards understanding the condensed phase.

1.1.2.3. Phase transitions in physical systems and replica symmetry

Let us start with explaining major concepts of this Ansatz using a fairly easy model. The Curie-Weiss model [71] is a simple model for ferromagnetism. Let $\Omega = \{-1,1\}$ and define the corresponding Hamiltonian as

$$H_{CW}(\sigma) = -\frac{1}{n}\sum_{i\neq j\in[n]}\sigma_i\sigma_j - B\sum_{i=1}^{n}\sigma_i. \tag{1.1.9}$$

We directly observe by above's discussion that the Curie-Weiss model is a prime example of a *mean-field* model as all pairs of variables interact with each other. Again, B is the strength of an external magnetic field and the first sum in (1.1.9) reduces the energy of such configurations that do not exhibit many different spins. Following Mézard and Montanari [135], we are going to compute the partition function of the Curie-Weiss model non-rigorously, but each step can in principle be turned into a rigorous argument. We emphasise that all formulas (and their derivations) can be found in [135]. First, we need a

[1] We define the binomial random graph as Gilbert [90], thus every edge is present independently of the rest with probability p.

simple global property of a given configuration σ: the magnetisation. Let

$$m(\sigma) = \frac{1}{n}\sum_{i=1}^{n}\sigma_i$$

be the magnetisation of σ, then the Hamiltonian (1.1.9) can be written as

$$H_{CW}(\sigma) = n\left(\frac{1}{2} - \frac{1}{2}m(\sigma)^2 - Bm(\sigma)\right), \tag{1.1.10}$$

thus can be expressed in terms of the magnetisation. Therefore, in the partition function, we can cluster the summands by their magnetisation $m \in \mathscr{M} = \{-1+2\ell/n : \ell = 0\dots n\}$ and find [135, Eq. (2.75)]

$$Z_{CW}(\beta,B) = \exp\left(-n\beta/2\right)\sum_{m\in\mathscr{M}}\binom{n}{n(1+m)/2}\exp\left(\frac{n\beta}{2}m^2 + n\beta Bm\right). \tag{1.1.11}$$

Using the standard asympotical behaviour of the binomial coefficient $\binom{n}{\alpha n} \sim \exp\left(nH(\alpha)\right)$ where $H(\alpha)$ is the entropy of a $\mathbf{Be}(\alpha)$ variable and the definition

$$\phi_m(\beta,B) = -\frac{\beta}{2}(1-m^2) + \beta Bm + H\left(\frac{1+m}{2}\right),$$

(1.1.11) can be simplified to [135, Eq. (2.77)]

$$Z_{CW}(\beta,B) \sim \int_{-1}^{1}\exp\left(n\phi_m(\beta,B)\right)\mathrm{d}m. \tag{1.1.12}$$

Here and subsequently, we use $\sim$ for asymptotic equality. Then the Laplace method applied to (1.1.12) yields [135, Eq. (2.79)]

$$\frac{1}{n}\ln\left(Z_{CW}(\beta,B)\right) \sim -\beta\max_{m\in[-1,1]}\phi_m(\beta,B). \tag{1.1.13}$$

The maximum turns out to be achieved away from the boundary and is dependent on the inverse temperature β as well as the external magnetic field B. Let us, for the moment, suppose that there is no external magnetic field, thus $B = 0$. Then for $0 \le \beta \le 1$, $\phi_m(\beta,0)$ is concave and takes its unique maximum at $m = 0$. Therefore, for $n \to \infty$, if we sample a configuration from the underlying Boltzmann distribution μ_{CW} defined as in (1.1.6), we do not know the exact configuration, but in almost all cases, its magnetisation will turn out to be 0, as all other configurations have exponentially smaller probability mass under μ_{CW}. In other words, almost all created instances of the system will exhibit the same property - thus we find a high degree of symmetry between different instances.

If now, on the other hand, β exceeds the critical value of 1, $\phi_m(\beta,0)$ is symmetric with respect to the y-axis and exhibits two global maxima $m^*(\beta) > 0$ and $-m^*(\beta) < 0$ whose peaks become larger with increasing β. Therefore, if we sample sufficiently often from the Boltzmann distribution, half of the configurations will have magnetisation $m^*(\beta)$ and the other half will show a magnetisation of $-m^*(\beta)$. In this sense, the symmetry is *broken*.

Therefore, in the thermodynamic limit, the space of all configurations of minimal energy Ω_0^n can be decomposed into two *pure states* $\Omega_0^n = \Omega_+^n \cup \Omega_-^n$ where

$$\Omega_+^n = \left\{\sigma \in \Omega_0^n : m(\sigma) = m^*(\beta)\right\} \qquad \text{and} \qquad \Omega_-^n = \left\{\sigma \in \Omega_0^n : m(\sigma) = -m^*(\beta)\right\}.$$

In the setting of Figure 1.2, this corresponds to a condensed phase, where the configurations of one cluster have significantly more particles of spin -1 than the configurations in the second cluster.

This observation corresponds intuitively to a fairly important feature of condensation: the marginal probabilities of the single particles under the Boltzmann distribution are dependent. Indeed, if σ is a random element of Ω_0^n and we get the information that particle x_1 has spin $\sigma_1 = -1$, it is more likely for

σ to be an element of Ω_-^n and therefore, in expectation, the other particles prefer spin -1 as well.

Let us at this point shortly discuss the so-called *replica symmetry* and *replica symmetry breaking*. Two replicas can be just seen as identically distributed instances of the same system. If such replicas are symmetric, the following effect takes place. Suppose we sample $\boldsymbol{\sigma}, \boldsymbol{\sigma}'$ from the Boltzmann distribution of a physical system and calculate the *overlap* of $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$ which is, for the Curie-Weiss model, defined as

$$\langle \boldsymbol{\sigma}, \boldsymbol{\sigma}' \rangle = n^{-1} \sum_{i=1}^{n} \boldsymbol{\sigma}_i \boldsymbol{\sigma}'_i.$$

Now, if the system is replica symmetric, we suppose that the overlap between two randomly sampled configurations is concentrated around exactly one value. If this is not the case, we say that *replica symmetry breaking* is present. The last formulation has to be read very cautiously because its definition is actually a bit more complicated. We say that a system is still replica symmetric if the pure states are related by *global symmetries* [131, 135]. The latter is clearly the case in the Curie-Weiss model as Ω_-^n can be turned into Ω_+^n (and vice versa) by flipping the spin of each particle.

Overall, we learned the intuition behind replica symmetry and replica symmetry breaking as well as that we defined pure states. In the following, we will formalise this intuition a bit.

1.1.2.4. Pure state decomposition, replica symmetry and the cut-distance

Let us suppose that we analyse the solution space and the Boltzmann distribution of a random constraint satisfaction problem (or a diluted mean-field model). Thus, the underlying graph corresponding to the interactions of the n particles is a sparse random graph. The physics' prediction suggests that the Boltzmann distribution has a fairly comfortable property before condensation (thus, in the replica symmetric phase): there are no long-range correlations present. More specifically, suppose we know the prevalence for a specific spin of the $i-$th particle x_i. If particle x_j is far away (in the sparse random graph) from x_i, then we do not gain any information about the spin of x_j from this knowledge. Formally, given a probability measure μ on Ω^n, we denote by μ_i the marginal of μ on spin i. More generally, for a set $I \subset [n]$, μ_I is the marginal probability on the particles in I. Then, we say that μ is ε*-symmetric* if

$$\mathbb{E}_{\boldsymbol{i},\boldsymbol{j}} \left[\left\| \mu_{\{\boldsymbol{i},\boldsymbol{j}\}} - \mu_{\boldsymbol{i}} \otimes \mu_{\boldsymbol{j}} \right\|_{\mathrm{TV}} \right] \leq \varepsilon.$$

Thus, ε-symmetry formalises the intuition that far-apart particles do not influence each other.

Now, after the condensation threshold, such long-range correlations start to appear. Even if the graph still looks locally tree-like, like every sparse random graph, the knowledge about a spin of a far-away particle might influence the probability of observing a specific spin on another particle. Recall that in the condensed phase of a random constrained satisfaction problem, the solution space is predicted to consist of clusters of solutions. Knowing the spin of a single particle might contain information on which of those clusters the configuration at hand lies in and therefore, this information pushes the probability to observe a certain spin on a different particle. But, if $1-$RSB is present, it is predicted that within one of those clusters, those long-range correlations disappear. More precisely, if $\Xi_1, \ldots, \Xi_\ell$ is the partition into clusters of Ω^n in the condensation phase and the physicist's prediction of $1-$RSB takes place, the Boltzmann distribution conditioned on one of these *pure states*, is symmetric. Formally, if μ is the Boltzmann distribution, we have for all $1 \leq k \leq \ell$

$$\mathbb{E}_{\boldsymbol{i},\boldsymbol{j}} \left[\left\| \mu_{\{\boldsymbol{i},\boldsymbol{j}\}}[\cdot|\Xi_k] - \mu_{\boldsymbol{i}}[\cdot|\Xi_k] \otimes \mu_{\boldsymbol{j}}[\cdot|\Xi_k] \right\|_{\mathrm{TV}} \right] = o(1).$$

We stress at this point that we only described the so-called $1-$RSB Ansatz. In physics' literature, the replica symmetry breaking might be of different orders. Intuitively, the clusters of solutions described here might themselves decompose into clusters of solutions ($2-$RSB) iteratively (until $\infty-$RSB) before the pure states (and thus absence of long-range correlations) appear. If the higher order RSB is interpreted via the overlap distribution as previously 1-RSB, we find that $k-$RSB corresponds to the concentration of the overlap onto $k+1$ values up to symmetry. Of course, this perfectly fits into the picture drawn before with clusters of clusters of clusters... Indeed, suppose that $2-$RSB is present and suppose we have *big*

clusters C_j and contained *small* clusters C_{jk}. If we draw two configurations from different big clusters C_j and $C_{j'}$ (independent of the small clusters), we observe a certain overlap value concentrated around q_0 (this are two configurations which are very far apart). If, on the other hand, we sample configurations within the same small cluster C_{ij}, the overlap will be concentrated around q_1 and finally, if the configurations come from the same big cluster but from different small clusters, we find an overlap concentrated around a third value q_2.

Thus, the geometrical interpretation directly shows that the overlap distribution facilitates a kind of *ultra-metricity*. What is an ultra-metric? We say that a metric d is an ultra-metric if the triangle inequality holds in a stronger version, e.g. $d(a,b) \leq \max\{d(a,c), d(b,c)\}$. This analytical property has one advantage if compared to the geometric interpretation of clusters of clusters of... Indeed, in the case of $\infty-$RSB, the single clusters would not be well separated but the analytical property of ultra-metricity still holds and describes the relevant behaviour [135, Section 8.2.2]. The interested reader can find a detailed discussion of (approximate) ultra-metricity and the replica symmetry breaking in an article of Jagannath [101].

Let us leave the interpretation of replica symmetry breaking for the moment, we will come back to it in a moment. Mathematically spoken, there is a decomposition of the Boltzmann distribution which kind of resembles this pure state decomposition. It comes along with a very simple algorithm: the pinning operation [49]. Extending and analysing results in context of the pure-state decomposition is part of one contribution of this thesis (c.f. Section 2.3). To be more precise, for an arbitrary probability measure on Ω^n, we can express the property of being $o(1)$-symmetric in terms of being close to a specific measure under a carefully chosen metric - the *cut-distance*. For two probability measures η, ν on a finite set Ξ, let $\Gamma(\eta,\nu)$ denote the set of all couplings γ of η and ν, thus γ is a probability measure on $\Xi \times \Xi$ with marginals η and ν. Further, let $\mathbb{S}_n$ denote the set of permutations $\phi : [n] \to [n]$. Then, the cut-distance is defined as

$$\Delta_{\boxtimes}(\mu,\nu) = \inf_{\substack{\gamma\in\Gamma(\mu,\nu),\\ \phi\in\mathbb{S}_n}} \sup_{\substack{S\subset\Omega^n\times\Omega^n,\\ X\subset[n],\\ \omega\in\Omega}} \left| \sum_{\substack{(\sigma,\tau)\in S,\\ x\in X}} \gamma(\sigma,\tau)\left(\mathbf{1}\{\sigma_x=\omega\} - \mathbf{1}\left\{\tau_{\phi(x)}=\omega\right\}\right)\right|. \tag{1.1.14}$$

The definition (1.1.14) can be interpreted as a kind of a two-player game. The first player chooses a coupling and a permutation of the particles under which μ and η look as much alike as possible. Now, given the choices γ and ϕ, the second player tries to find a subset of coordinates and configurations, on which the two measures differ as much as possible. We stress that the cut-distance is indeed a very weak metric [49]. But it suffices to gather the idea of $\varepsilon-$symmetry. Let μ be a probability measure on Ω^n and define

$$\bar{\mu}(\sigma) = \prod_{i=1}^{n} \mu_i(\sigma_i) \tag{1.1.15}$$

as the product measure on the marginals of μ. Clearly, $\bar{\mu}$ is a probability measure on Ω^n as well and if μ is a Boltzmann distribution such that the spins on all particles are completely independent, we have $\mu \equiv \bar{\mu}$. Thus, if most pairs of particles do not influence each other, we would expect that μ and $\bar{\mu}$ are close to each other under a certain distance. The cut-distance turns out to exactly express this connection. As proven by Coja-Oghlan and Perkins [53, Proposition 2.5], we have

$$\Delta_{\boxtimes}(\mu,\bar{\mu}) \leq \varepsilon^3 \Longrightarrow \mathbb{E}_{\boldsymbol{i},\boldsymbol{j}}\left[\left\|\mu_{\{\boldsymbol{i},\boldsymbol{j}\}} - \mu_{\boldsymbol{i}}\otimes\mu_{\boldsymbol{j}}\right\|_{\mathrm{TV}}\right] \leq O(\varepsilon)$$

and

$$\mathbb{E}_{\boldsymbol{i},\boldsymbol{j}}\left[\left\|\mu_{\{\boldsymbol{i},\boldsymbol{j}\}} - \mu_{\boldsymbol{i}}\otimes\mu_{\boldsymbol{j}}\right\|_{\mathrm{TV}}\right] \leq \varepsilon^3 \Longrightarrow \Delta_{\boxtimes}(\mu,\bar{\mu}) \leq O(\varepsilon).$$

Thus, the cut-distance is a perfectly suiting mathematical tool to formalise the idea of pure states and the absence of long range correlations. Furthermore, as already mentioned, the cut-distance can be used to find a decomposition of any configuration space Ω^n for a large enough system ($n \geq n_0 \in \mathbb{N}$) into

finitely many parts $E_1, \dots, E_\ell$ such that any probability measure conditioned on a specific part $\mu[\cdot \mid E_j]$ is close to the product of its marginals under the cut-distance, hence $\Delta_{\boxtimes}\left(\mu[\cdot \mid E_j], \bar{\mu}[\cdot \mid E_j]\right) \leq \varepsilon$ (for any $\varepsilon > 0$). This result is a type of a *regularity lemma* and will be discussed in more detail in Section 1.3.1.

The cautious reader might ask how this finite decomposition fits into the predictions of replica theory. Indeed, the physics' prediction is that the phase space decomposes into infinitely many pure states (if $n \to \infty$) if some kind of replica symmetry breaking is present [131] while the pinning operation – and regularity lemmas which will be discussed later – yield a finite decomposition. But the pinning operation (as well as the regularity lemmas) only guarantee the existence of states $E_1, \dots, E_\ell$ with above's property such that $\mu(E_1 \cup \dots \cup E_\ell) \geq 1 - \varepsilon$ for any small but constant $\varepsilon > 0$. Therefore, there can be infinitely many clusters which carry very little probability mass under the Boltzmann distribution.

After having described how the solution space of random constraint satisfaction problems (respectively, diluted mean field models) might look like, we will study a very specific random constraint satisfaction problem in more detail which plays the leading role in one of this thesis's contributions.

1.1.3. The random 2-SAT problem

While we already introduced the $k-$SAT problem for arbitrary k, we will focus on the very special case $k = 2$. This setup plays a very specific role, as it is the only k-SAT problem in which deciding whether a given formula is satisfiable and, if so, finding a satisfying assignment is computationally easy [120]. For the random $2-$SAT problem, the satisfiability threshold is well known since the early 1990's by independent works of Chvatal and Reed [40] and Goerdt [91], which both link the problem to the well studied percolation phase transition in a random directed graph. Subsequently, Bollobás et al. [30] managed to analyse the scaling window of the satisfiability threshold in more detail which turns out to correspond to the scaling window of the giant component in the binomial random graph [29].

Even if the satisfiability threshold itself is well understood, things were completely different for a question that seems quite innocuous at first. If a random $2-$SAT formula is satisfiable w.h.p., how many satisfying assignments exist? Indeed, this question was posed prominently by Fernandez de la Vega [78] and it actually turns out that counting the number of satisfying assignments is computationally hard, this counting problem lies in $\#\mathbf{P}$ [165].

Describing and analysing the partition function of a random $2-$SAT formula would answer this question as the partition function exactly counts the number of satisfying assignments, at least, if the formula is satisfiable in the first place. Luckily, there is a non-rigorous heuristics from statistical physics which leads to a precise prediction of how this partition function looks like. One of this thesis's contribution verifies this prediction rigorously as will be seen in Section 2.2. Intuitively, the non-rigorous physics' approach calculates the marginals of the Boltzmann distribution, thus for each variable node x they get the probability that x is assigned the value 0 or 1 under a randomly chosen satisfying assignment $\boldsymbol{\Phi}$. Now, it is possible to connect $Z(\boldsymbol{\Phi})$ with those marginal probabilities through an operator called *Bethe functional*. We will subsequently describe how to (non-rigorously) calculate marginals of the Boltzmann distribution in Section 1.1.4 and how to link them to the partition function in Section 1.1.5. We start by finding the marginal probabilities of the Boltzmann distribution. In physics' literature, specific message passing algorithms are conjectured to be able to do so.

1.1.4. Getting the marginals through message passing

The key idea behind message passing algorithms is the following. Given a factor graph $\mathcal{G}$, an algorithm computes *messages* for each edge of $\mathcal{G}$ such that in each round of the algorithm, a set of messages is sent on every edge of the graph in parallel. Those messages can be sent from the variables to the factors and vice versa. A computationally intriguing advantage of message passing algorithms is that all messages are computed through local functions at the vertices of $\mathcal{G}$. Probably the most prominent message passing algorithm is Belief Propagation BP. It was discovered under various names in different fields of research, for instance, in statistical physics it is known as the *Bethe-Peierls approximation* [27], while coding theorists developed the *sum-product algorithm* [122]. The terminology *Belief Propagation* which we will use subsequently, has its origins in the research towards artificial-intelligence [145].

1.1.4.1. Belief Propagation

Let $\mathcal{G} = (V \cup F, E)$ be a given factor graph. BP is used to estimate the marginal distribution for each variable, for instance, if $\mathcal{G}$ represents a $k-$SAT formula, this corresponds to the probability of a specific variable being set to 1 under a random satisfying assignment. While its first occurrence can be traced back to the 1930's in statistical physics [27], it was Pearl [145] who proved that BP correctly calculates the marginals if $\mathcal{G}$ is a tree. Furthermore, it is a widespread conjecture that BP performs well if the graph is at least locally tree-like and there are almost no long-range correlations as discussed in Section 1.1.2.4 [146]. Probably due to its performance as well as being an efficient algorithm, BP finds its applications in artificial intelligence and information theory. Empirically it was shown that BP can be used in various applications including ldpc-codes and turbo-codes in coding theory, free energy approximation in statistical physics and satisfiability in theoretical computer science [38].

Let us now introduce the messages sent by BP formally. In order to do so, we require a few definitions. Let Ω be a finite set (the set of spins as before) and $\mathcal{G} = (V \cup F, E)$ a factor graph. Denote by $x_1, \dots, x_n \in V$ the n variables and by $a_1, \dots, a_m \in F$ the m factors of $\mathcal{G}$. Furthermore, let $\sigma \in \Omega^n$ be a configuration that assigns each variable a specific spin from Ω. Further, for $I = \{i_1, \dots, i_k\} \subset [n]$, we set $\sigma_I = (\sigma_{i_1}, \dots, \sigma_{i_k}) \in \Omega^k$ and denote by Ω^I the subspace of Ω^n given by the coordinates of I. Moreover, for a vertex $v \in V \cup F$, we denote by $\partial_{\mathcal{G}} v \subset V \cup F$ the set of neighbours of v in $\mathcal{G}$. If the context clarifies what the underlying graph structure is, we write ∂v for the sake of brevity. Finally, we use $\propto$ to express equality up to a normalisation constant and write $f(n) \sim g(n)$ if $f(n)/g(n) \to 1$ for $n \to \infty$.

Before stating the BP messages, let us revisit the definition of a factor graph. As already introduced, one merit of a factor graph is that the Hamiltonian of the corresponding physical system factorises such that the Boltzmann distribution can be written as

$$\mu(\sigma) \propto \prod_{i=1}^{m} \psi_{a_i}(\sigma_{\partial a_i}). \tag{1.1.16}$$

Here, $\psi_{a_1}, \dots, \psi_{a_m}$, are the local contributions of the single factor nodes to the system's energy. Of course, given a factor graph $\mathcal{G} = (V \cup F, E, \Psi)$ with *weight functions* $\Psi = \{\psi_{a_1}, \dots, \psi_{a_m}\}$, we can associate a corresponding Hamiltonian such that the system's energy is calculated via (1.1.16). It turns out that sometimes it is more convenient to use the factor graph notation and talk about weight functions instead of referring to the physical interpretations.

We are now in position to state the BP messages on such a factor graph $\mathcal{G} = (V \cup F, E, \Psi)$. For $y \in \Omega$, a variable $x \in V$ and a test $a \in F$ they read

$$\nu^{(t+1)}_{x \to a}(y) \propto \prod_{b \in \partial x \setminus a} \nu^{(t)}_{b \to x}(y) \qquad \text{and} \qquad \nu^{(t)}_{a \to x}(y) \propto \sum_{\sigma_{\partial a \setminus x}} \psi_a(\sigma_{\partial a}) \prod_{x' \in \partial a \setminus x} \nu^{(t)}_{x' \to a}(y). \tag{1.1.17}$$

More precisely, the set $\left\{\nu^{(t)}_{x \to a}\right\}_{x \in V, a \in F}$ is called the set of *variable-to-factor messages* at time t while the *factor-to-variable messages* are $\left\{\nu^{(t)}_{a \to x}\right\}_{x \in V, a \in F}$. Clearly, each message itself is a probability distribution on Ω. Intuitively speaking, $\nu^{(t)}_{x \to a}$ represents the marginal distribution of spins on variable x in a model that does not contain factor a and analogously, $\nu^{(t)}_{a \to x}$ is the marginal distribution of x in a model where all factors in which x is contained except of a have been deleted.

Observing those messages, various natural questions arise. Probably the most intriguing are the following.

- Do these messages converge to a fixed-point?
- If so, is the fixed-point unique?
- If these messages indeed converge to a unique fixed-point, how are those messages related to the marginals of the Boltzmann distribution?

These questions can be answered rigorously for tree-factor graphs, thus for graphs $\mathcal{G}$ that do not contain cycles. Furthermore, it is conjectured that similar results hold for locally tree-like graphs if the solution

space exhibits certain properties. This conjecture was proven rigorously in some special cases [121, 50]. We will discuss those questions in Section 1.1.5. BP is designed to be applied to any distribution that can be written like (1.1.16), but in many CSPs, the weight functions have a very specific form, as the Hamiltonian just counts the number of unsatisfied constraints (clauses, in above's terminology on satsifiability). Therefore,

$$\psi_a(\sigma_{\partial a}) = 1 - \mathbf{1}\{a \text{ is satisfied under } \sigma\}. \tag{1.1.18}$$

In this special case, the messages (1.1.17) simplify remarkably, i.e., if the set of messages at time zero is initialised with values in $\{0,1\}$, then this condition still holds after an arbitrary number of updates [135, Proposition 14.5]. The resulting update rules are known as *Warning Propagation* WP [75].

1.1.4.2. Warning Propagation

The semantic interpretation of those integer-valued BP messages is the reason why WP holds its name. More precisely, for a spin $y \in \Omega$, a variable x and a factor a, we have

$$\begin{aligned} \nu_{x\to a}(y) = 1 \quad &\leftrightarrow \quad \text{according to all factors } b \in \partial x \setminus a \text{ variable } x \\ &\qquad \text{should not have spin } y \text{ under a satisfying assignment.} \\ \nu_{x\to a}(y) = 0 \quad &\leftrightarrow \quad \text{according to all factors } b \in \partial x \setminus a \text{ variable } x \\ &\qquad \text{may take spin } y \text{ under a satisfying assignment.} \end{aligned} \tag{1.1.19}$$

Thus, the variable-to-factor messages $\{\nu_{x\to a}\}$ *warn* a factor node a if a variable can probably not be used to satisfy a taking certain values. WP itself is a broadly used algorithm in random CSPs and it is known that it finds all direct implications of a partial assignment of the variables, based on the local structure [135]. Furthermore, it is known to find satisfying assignments of different satisfiability problems under mild conditions, for instance, it can be used to solve certain instances of random 3-SAT [75].

Let us suppose for the moment that BP or WP are exact on a given problem. We will now discuss what we actually understand by *being exact*.

1.1.5. Boltzmann marginals and the partition function

Let, as above, $\left\{\nu^{(t)}_{x\to a}\right\}$ and $\left\{\nu^{(t)}_{a\to x}\right\}$ be the BP messages at time t. Then we define the *BP estimation of the marginals at time t* as

$$\nu^{(t)}_x(y) \propto \prod_{a\in\partial x} \nu^{(t)}_{a\to x}(y). \tag{1.1.20}$$

Hence, $\nu^{(t)}_x$ is the product measure over all incoming messages at variable x. If μ is the corresponding Boltzmann distribution, we have that $\nu^{(t)}_x$ is an estimation for μ_x. If we say that BP is exact (for instance, on factor graphs that are acyclic), then it is rigorously proven that

$$\nu^{(t)}_x = \mu_x$$

if t is large enough. This, of course, already implies that the BP messages converge to a fixed-point. We call the set of fixed-point messages $\{\nu^*_{x\to a}\}$ and $\{\nu^*_{a\to x}\}$ respectively.

Now, let us discuss, how this helps understanding the partition function of the underlying system. Given a factor graph $\mathcal{G} = (V \cup F, E, \Psi)$ and a set of messages $\nu = (\{\nu_{x\to a}\}, \{\nu_{a\to x}\})$, we define the *Bethe free*

entropy as follows. Recall that those messages are probability distributions on Ω and define

$$\Xi_a(\nu) = \ln\left(\sum_{\sigma_{\partial a}\in\Omega^{\partial a}} \psi_a(\sigma_{\partial a}) \prod_{x\in\partial a} \nu_{x\to a}\right), \qquad \Xi_x(\nu) = \ln\left(\sum_{\omega\in\Omega} \prod_{b\in\partial x} \nu_{b\to x}\right),$$
$$\Xi_{x,a} = \ln\left(\sum_{\omega\in\Omega} \nu_{x\to a}(\omega)\nu_{a\to x}(\omega).\right) \tag{1.1.21}$$

Thus $(\Xi_a)_{a\in F}$ and $(\Xi_x)_{x\in V}$ describe the entropy on the factor nodes and on the variables respectively, while $\Xi_{x,a}$ measures their interaction. If we put in the set of fixed-point messages ν^*, we get the Bethe free entropy $\mathbf{\Phi}$ as

$$\mathbf{\Phi} = \sum_{a\in F} \Xi_a(\nu^*) + \sum_{x\in V} \Xi_x(\nu^*) - \sum_{ax\in E} \Xi_{x,a}(\nu^*). \tag{1.1.22}$$

If BP is exact on a model, then we find

$$\ln Z = \mathbf{\Phi}. \tag{1.1.23}$$

Thus, (1.1.23) gives a recipe, how to calculate the partition function of a physical system, for instance, a random constraint satisfaction problem, by using an easy to implement efficient algorithm. Due to the ***NP***-hardness of calculating $\ln Z$ in general, as discussed earlier, this can of course only be true for very specific instances of random CSPs (if we suppose $\boldsymbol{P} \neq \boldsymbol{NP}$).

Indeed, it turns out that the *Bethe approximation by Belief Propagation*, which we will call the procedure above, yields the correct value of the partition function sometimes [50, 52, 145] and sometimes it does not [126]. The latter is assumed to happen, if the system's solution space had undergone a phase transition at which replica symmetry breaking occurred. A statistical physics' approach suggests to run BP on a modified problem in this case.

1.1.5.1. The (1-RSB) Cavity Method

Before going deeper into the *1-RSB cavity approach*, we shortly stress that the method itself is a highly non-rigorous technique. Furthermore, this presentation's focus is solely to draw an intuitive idea behind the 1-RSB cavity method and not to carry out technical details.

Intuitively, BP depends on a central assumption, namely, whenever we delete a factor node, the spins on the affected variables become roughly independent. This assumption is clearly violated when the factor graph either contains short cycles, or whenever long-range correlations appear [135]. As (sparse) random factor graphs which we deal with in this thesis (or: diluted mean-field models) look locally tree-like, the failure of BP needs to be due to the appearance of such long-range correlations. In the replica symmetric phase, those are supposed to be negligible, thus we study systems in a phase in which replica symmetry is broken.

We recall that in the presence of 1-RSB, the configuration space Ω^n is supposed to decompose into pure states (c.f. Section 1.1.2.4). Let us take a different look on those pure states. Recall that BP is known to render the correct marginals of the Boltzmann distribution on some replica symmetric factor graphs and recall that given $\mathcal{G} = (V \cup F, E, \Psi)$, we can write the Boltzmann distribution as

$$\mu(\sigma) \propto \prod_{a\in F} \psi_a(\sigma_{\partial a}).$$

Let $U \subset V$ be a subset of the variables, then we define $\mathcal{G}[U] = (U \cup F[U], E[U], \Psi)$ as the induced factor graph on U as follows. It contains

- all variables of U,
- all factor nodes $a \in F$ such that $\partial a \subset U$,
- all edges a, x with $a \in F[U], x \in U$,

- all half-edges x, a with $a \notin F[U]$ but $x \in U$.

The half-edges play an important role. Let us denote the set of those half-edges by $\mathscr{H}(U)$. Suppose that BP renders the correct marginals, then the messages $\{\nu_{a\to x}\}_{(x,a)\in\mathscr{H}(U)}$ on the half-edges can be seen as *boundary conditions* which represent the influence of the factors and variables outside of U. We will call such a set U coming with $\mathscr{G}[U]$ a *cavity*. Observe that some variables in U miss factor nodes (constraints) compared to their distribution in $\mathscr{G}$. If the system is replica symmetric, we would suppose that we can express the Boltzmann marginal on U – at least approximately – by the weight functions at the factor nodes inside of U and the boundary condition, thus

$$\mu_U(\sigma_U) \propto \prod_{a\in F[U]} \psi_a(\sigma_{\partial a}) \prod_{ax\in\mathscr{H}(U)} \nu_{a\to x}(\sigma_x) + \varepsilon_n. \tag{1.1.24}$$

Led by this intuition, we call a probability measure μ on Ω^n a *Bethe measure* or *Bethe state* if in the thermodynamic limit, $n\to\infty$, there is a set of messages $\{\nu_{a\to x}\}$ such that μ satisfies (1.1.24) for almost all finite subsets U [135, Definition 19.1]. We highlight that there has to be one set of BP messages satisfying (1.1.24) for (almost) all finite subsets of variables and in fact it turns out that messages satisfying the equation for almost all U are actually very close to valid Belief Propagation messages (1.1.17). To be a bit more precise, an *almost-solution* of the BP equations is a set of messages that satisfy almost all BP equations up to a vanishing error term of $o(1)$.

What does this imply? Any set of messages satisfying (1.1.24) for almost all U, thus a Bethe measure, corresponds to an almost-solution of the Belief Propagation messages. It is far from true that the converse is correct as well [135, Example 19.2] in every situation, but a core assumption of the 1-RSB cavity method is that this is indeed correct in the problem at hand. Suppose we have a set of almost-solutions of the BP equations $\{\nu^i = (\nu^i_{x\to a}, \nu^i_{a\to x})\}_{i=1\dots\ell}$, then we can associate with each of those (almost) fixed-points a corresponding Bethe measure μ^i (1.1.24) as well as the corresponding Bethe free entropy $\mathbf{\Phi}_i$ via (1.1.22). We call a Bethe measure *extremal* if it does not exhibit long-range correlations. We denote the set of all extremal Bethe measures out of $\{\mu^i\}_{i=1\dots\ell}$ as $\{\tilde\mu^i\}_{i=1\dots\tilde\ell}$ corresponding to BP messages $\{\tilde\nu^i\}_{i=1\dots\tilde\ell}$ Now, there are three basic assumptions on a model that make the 1-RSB method work heuristically. To this end, let $\mathbf{\Sigma}:\mathbb{R}\to\mathbb{R}_+$ be a function, which is called *complexity function* in physics' language.

1. For any interval $[\phi,\phi']$, the number of almost-solutions with Bethe free entropy $\mathbf{\Phi}_i \in [n\phi, n\phi']$ equals $\exp(n\mathbf{\Sigma}^* + o(n))$ where $\mathbf{\Sigma}^* = \sup_{\sigma^*\in[\phi,\phi']}\{\mathbf{\Sigma}(\phi^*)\}$. Thus, roughly speaking, the number of almost-solutions with Bethe free entropy of approximately $n\phi$ is given by $\exp\big(n\mathbf{\Sigma}(\phi)\big)$.

2. The Boltzmann distribution can be expressed as a convex combination of extremal Bethe measures
$$\mu(\sigma) = \sum_{i=1}^{\tilde\ell} \omega_i \tilde\mu^i(\sigma)$$
with weights $\omega_i = \exp(\mathbf{\Phi}_i)/\mathscr{Z}$ such that $\mathscr{Z} = \sum_{i=1}^{\ell}\exp(\mathbf{\Phi}_i)$.

3. The number of extremal Bethe measures $\tilde\ell$ equals, up to the leading order, the number of almost-solutions to the BP equations, thus the number of extremal Bethe measures with Bethe free entropy $\sim n\phi$ equals approximately $\exp\big(n\mathbf{\Sigma}(\phi)\big)$.

Now suppose that the three assumptions are satisfied. Then we build a new physical system (an auxiliary model) from the original system we started with. More precisely, we interpret the BP messages as variables of the auxiliary model and the corresponding Bethe measures become the configurations. To this end, let ζ be the set of the extremal Bethe measures and let Λ be a probability measure on ζ defined as follows. Let κ be the inverse temperature, sometimes called *Parisi 1-RSB parameter*, of the auxiliary model. Then we define

$$\Lambda(\tilde\mu^i) = \omega_i(\kappa) = \frac{\exp(\kappa\mathbf{\Phi}_i)}{\mathscr{Z}(\kappa)}. \tag{1.1.25}$$

At a first glance, (1.1.25) strongly resembles the definition of the Boltzmann distribution (1.1.6) and indeed, as it turns out, this resemblance has a deeper meaning. But first observe that for $\kappa = 1$, the auxiliary model equals the original system. Nevertheless, studying $\mathscr{Z}(\kappa)$ for general κ turns out to be the key for calculating $\boldsymbol{\Sigma}$, which ultimately allows understanding the original model. Without going too much into detail, we will shortly describe how to translate a physical model with an underlying sparse random factor graph which undergoes 1-RSB, into an auxiliary model which can then, itself, be solved using Belief Propagation. We stress at this point that the Boltzmann distribution (1.1.25) of the auxiliary model is a probability distribution on probability distributions, thus a difficult object to analyse.

Let us assume that we start with a problem described by the factor graph $\mathscr{G} = (V \cup F, E, \Psi)$. Then, following Mézard and Montanari [135, Section 19.2.1], we create an auxiliary model in three straightforward steps. Suppose that $\{\nu_{x\to a}\}$ and $\{\nu_{a\to x}\}$ are the BP messages on $\mathscr{G}$. We highlight that we leave out technical details (for instance, we suppose implicitly that those messages were discrete measures) as we are only interested in sketching the main idea of the 1-RSB cavity method. Recall Ξ_a, Ξ_x, Ξ_{ax} from (1.1.21). Then, we proceed as follows in the construction of the factor graph $\boldsymbol{G} = (\boldsymbol{V} \cup \boldsymbol{F}, \boldsymbol{E}, \boldsymbol{\Psi})$ representing the auxiliary model.

1. For $ax \in E$, we create a variable node $\boldsymbol{xa}$ representing $(\nu_{x\to a}, \nu_{a\to x})$ and a factor node connected to this variable with weight function $\boldsymbol{\Psi}_{xa} = \exp(-\kappa \Xi_{xa}(\nu))$.

2. For any factor $a \in F$, we introduce a factor node $\boldsymbol{a}$ in the auxiliary model and connect it to all variable nodes $\boldsymbol{xa}$ with $x \in \partial a$ in the original model. The corresponding weight function reads

$$\boldsymbol{\Psi}_a = \prod_{x\in\partial a} \mathbf{1}\left\{\nu_{a\to x} \propto \sum_{\sigma_{\partial a}\in\Omega^{\partial a}} \Psi_a(\sigma_{\partial a}) \prod_{x'\in\partial a\setminus x} \nu_{x'\to a}\right\} \exp\left(-\kappa \Xi_a\left(\left\{\nu_{y\to a}\right\}_{y\in\partial a}\right)\right).$$

   (Here, we write $\Xi_a\left(\left\{\nu_{y\to a}\right\}_{y\in\partial a}\right)$ instead of $\Xi_a(\nu)$ in order to highlight the local dependencies.) The purpose of the weight function is two-fold. First, it makes sure that the correct BP equations are observed and second, it weighs the contribution by a factor of $\exp(\kappa \Xi_a)$.

3. Each variable $x \in V$ produces a factor node $\boldsymbol{x} \in \boldsymbol{F}$ in the auxiliary model. This factor connects to all variables $\boldsymbol{xa}$ if $a \in \partial x$ in the original model. The corresponding weight function is defined as

$$\boldsymbol{\Psi}_x = \prod_{a\in\partial x} \mathbf{1}\left\{\nu_{x\to a} \propto \prod_{b\in\partial x\setminus a} \nu_{b\to x}\right\} \exp\left(\kappa \Xi_x\left(\{\nu_{b\to x}\}_{b\in\partial x}\right)\right).$$

   Again, this weight function guarantees observing the valid BP messages as well as it weighs the factor node's contribution.

Now it is possible to run BP on this auxiliary model. If the physics' intuition is correct, this auxiliary model is replica symmetric and BP yields a valid estimate of $\ln \mathscr{Z}(\kappa)$ via the Bethe free entropy. Recall that we have by definition

$$\mathscr{Z}(\kappa) = \sum_{i=1}^{\tilde{\ell}} \exp(\kappa \boldsymbol{\Phi}_i).$$

Leaving out technical details and justifications, we obtain

$$\mathscr{Z}(\kappa) \sim \int \exp\left(n\left(x\phi + \boldsymbol{\Sigma}(\phi)\right)\right) \mathrm{d}\phi.$$

Then, if $\mathscr{Z}(\kappa) \sim \exp\left(n f(\kappa)\right)$ for some function f, we find $\boldsymbol{\Sigma}$ through a Legendre transformation [135, Section 19.2] as

$$f(\kappa) = x\kappa + \boldsymbol{\Sigma}(\phi) \qquad \text{such that} \qquad \frac{\partial \boldsymbol{\Sigma}}{\boldsymbol{\phi}} = -\kappa.$$

In summary, if a model shows 1-RSB, Belief Propagation will not render the correct estimate for the partition function, i.e. it will have various (almost) fixed-points. In this case, the physics' intuition

proposes to create a statistical auxiliary model with the almost-solutions being variables and analysing the partition function of the auxiliary model by BP yielding to a (hopefully) unique fixed-point.

To put this section into context, we recall the already discussed phase diagram of random CSPs. Suppose that the Boltzmann distribution was concentrated on finitely many pure states (which is called *static 1-RSB*). Then the solution space has undergone the condensation phase transition. In the case that the Boltzmann distribution is not concentrated on a small amount of pure states but there are exponentially many pure states each with exponentially small probability mass (*dynamical 1-RSB*), the solution space finds itself in the clustering phase. We furthermore stress that one basic assumption of the 1-RSB Ansatz is that the Boltzmann distribution can be written as a convex combination of *extremal Bethe measures*, thus probability distributions which lack long-range correlations and which describe the local behaviour of the Boltzmann distribution based only on a finite neighbourhood with a boundary condition given by (almost)-solutions to the BP messages. This strongly resembles the finding of a partition of the configuration space Ω^n into clusters $\mathscr{E}_1,\dots,\mathscr{E}_\ell$ such that the Boltzmann distribution conditioned on those clusters is $\varepsilon-$extremal (c.f. Section 1.1.2.4). And indeed, the cut-distance formalism turned out to be a key tool in verifying the results predicted by the cavity method in some special problems [21, 48, 49].

Until now, we studied sparse random CSPs from a specific point of view which ultimately results in calculating the number of satisfying assignments. Thus, given a degree distribution of the factor nodes $(\deg(a_1),\dots,\deg(a_m))$ and a degree distribution of the variable nodes $(\deg(x_1),\dots,\deg(x_n))$ (which might be obtained by a distribution of choice), we first draw the set of weight functions $\Psi_1,\dots,\Psi_m$ from a distribution (which might, in principle, turn out to be deterministic choices like in the random $k-$SAT problem). Then, given the degree sequences as well as the weight functions, a factor graph is drawn uniformly at random from all factor graphs on those degree distributions. We can now ask the question, whether this random CSP is satisfiable w.h.p., or determine the number of solutions or study the geometry of the solution space. There is yet another interesting model closely related, which will be discussed in the next section.

1.2. Statistical Inference

The task of statistical inference can be modelled by the so-called *teacher-student scheme* quite intuitively. The scheme itself was introduced by Gardner and Derrida [87] in the context of studying the perceptron, a fairly easy binary classifier. In this section, we follow an introduction into statistical inference based on the study of physical systems by Zdeborová and Krzakala [168].

1.2.1. A statistical physics' approach

1.2.1.1. The teacher-student scheme

As a first step, the teacher generates some *ground-truth* $\boldsymbol{\sigma}$ from an arbitrary probability distribution μ^{TP} - this is the *teacher's* prior. Now he generates some observable data $\hat{\boldsymbol{\sigma}}$ from $\boldsymbol{\sigma}$ by a statistical model. This model is characterised by a distribution $\mu^{TM}(\hat{\boldsymbol{\sigma}} \mid \boldsymbol{\sigma})$, which expresses the likelihood of observing $\hat{\boldsymbol{\sigma}}$ given that the ground-truth was $\boldsymbol{\sigma}$ - this likelihood distribution is the *teacher's model*. And finally, the teacher conveys the data $\hat{\boldsymbol{\sigma}}$ and some information about μ^{TP} as well as μ^{TM} to the student.

The student's ultimate goal is to infer as much information as possible about $\boldsymbol{\sigma}$ from the given data and the (probably limited) information about the teacher's prior as well as the teacher's model. If the teacher gives the full information about the prior as well as the model to the student, we call this setting *Bayes optimal*. Let us focus on statistical inference problems which exhibit quite convenient properties. More precisely, we suppose the following. Let Ω be a finite set and sample the ground-truth $\boldsymbol{\sigma} \in \Omega^n$ from μ^{TP}. Furthermore, suppose that $\hat{\boldsymbol{\sigma}}$ is an m-dimensional vector with entries in a finite set χ and that the following assumptions hold.

- The prior distribution factorises, thus

$$\mu^{TP}(\sigma) = \prod_{i=1}^{n} \mu_i^{TP}(\sigma_i).$$

- The single observable data points $\hat{\boldsymbol{\sigma}}_1, \dots \hat{\boldsymbol{\sigma}}_m$ are independently generated based on a subset of variables $\partial\hat{\boldsymbol{\sigma}}_1, \dots, \partial\hat{\boldsymbol{\sigma}}_m$, hence

$$\mu^{TM}(\hat{\boldsymbol{\sigma}} \mid \boldsymbol{\sigma}) = \prod_{j=1}^{m} \mu_{\partial\hat{\boldsymbol{\sigma}}_j}(\hat{\boldsymbol{\sigma}}_j \mid \partial\hat{\boldsymbol{\sigma}}_j).$$

Let us now view the whole scenario through the student's eyes. As the student might not have full access to the teacher's prior as well as the model's statistics, we suppose that the known prior is described by a distribution ν on Ω^n and the information about data generation is modelled by a distribution $\tilde{\nu}$ on χ^m. If ν and $\tilde{\nu}$ satisfy above's assumptions, the student can employ Bayes' theorem and write the posterior distribution as

$$\nu(\boldsymbol{\sigma} \mid \hat{\boldsymbol{\sigma}}) \propto \prod_{i=1}^{n} \nu_i(\boldsymbol{\sigma}_i) \prod_{j=1}^{m} \tilde{\nu}_{\partial\hat{\boldsymbol{\sigma}}_j}(\hat{\boldsymbol{\sigma}}_j \mid \partial\hat{\boldsymbol{\sigma}}_j). \tag{1.2.1}$$

Now, the best the student can do in order to infer $\boldsymbol{\sigma}$ from $\hat{\boldsymbol{\sigma}}$, ν and $\tilde{\nu}$, is to sample a configuration $\tilde{\sigma}$ from $\nu(\boldsymbol{\sigma} \mid \hat{\boldsymbol{\sigma}})$. If $\tilde{\sigma} = \boldsymbol{\sigma}$, we say that the student succeeded in inferring the ground-truth completely. If we are in the convenient situation of Bayes optimality, thus the student gets the full information about the teacher's prior and the data generation, we can observe the following. Denote by $\boldsymbol{\tau}, \boldsymbol{\tau}', \boldsymbol{\tau}''$ three independent uniform samples from $\nu(\cdot \mid \hat{\boldsymbol{\sigma}})$. Furthermore, let $f : \Omega^n \times \Omega^n \to \mathbb{R}$ be some arbitrary function, then we have [168, Eq. (15)]

$$\mathbb{E}\left[f(\boldsymbol{\tau}', \boldsymbol{\tau}'')\right] = \mathbb{E}\left[f(\boldsymbol{\tau}, \boldsymbol{\sigma})\right].$$

Thus, at least with respect to the expectation, there is no difference between the ground-truth $\boldsymbol{\sigma}$ and uniform samples from the posterior distribution. This result in the Bayes optimal setting is called the *Nishimori property* and from now on, we will tacitly assume to satisfy this Nishimori property, thus being in the Bayes optimal setting. Next, we express the problem of inference in the spin glass language from the previous sections.

1.2.1.2. Planted models

Probably the first connection between statistical inference problems and statistical physics was observed by Jaynes [104] while the expression of statistical inference problems in terms of spin glass language was strongly influenced by pioneering work of Kirkpatrick, Gelatt and Vecchi [110] on simulated annealing. We first observe that (1.2.1) can be written as [168, Eq. (25)]

$$\nu(\boldsymbol{\sigma} \mid \hat{\boldsymbol{\sigma}}) \propto \exp\left(\sum_{i=1}^{n} \ln\left(\nu_i(\boldsymbol{\sigma}_i)\right) + \sum_{j=1}^{m} \ln\left(\tilde{\nu}_{\partial\hat{\boldsymbol{\sigma}}_j}\left(\hat{\boldsymbol{\sigma}}_j \mid \partial\hat{\boldsymbol{\sigma}}_j\right)\right)\right). \tag{1.2.2}$$

Now, we introduce an inverse temperature β, such that $\beta = 1$ recovers the original posterior distribution and introduce $\ln\left(\nu_i(\boldsymbol{\sigma}_i)\right)$ as an external magnetic field at particle i while $\ln\left(\tilde{\nu}_{\partial\hat{\boldsymbol{\sigma}}_j}\left(\hat{\boldsymbol{\sigma}}_j \mid \partial\hat{\boldsymbol{\sigma}}_j\right)\right)$ expresses the interaction between particles. Thus, as given through (1.1.1), we find a Hamiltonian $H(\boldsymbol{\sigma}, \hat{\boldsymbol{\sigma}})$ such that (1.2.2) becomes

$$\nu(\boldsymbol{\sigma} \mid \hat{\boldsymbol{\sigma}}) = \frac{\exp\left(-\beta H(\boldsymbol{\sigma}, \hat{\boldsymbol{\sigma}})\right)}{Z(\hat{\boldsymbol{\sigma}})} \tag{1.2.3}$$

with $Z(\hat{\boldsymbol{\sigma}})$ being the partition function of the described physical system. Such a physical system is called a *planted model*. Let us elaborate on this shortly. While the particle interactions $J_{i_1,\dots,i_k}(\sigma_{i_1}, \dots, \sigma_{i_k})$

in (1.1.1) can be any function on k spins, spin glasses exhibit positive as well as negative interactions between particles. One possibility to generate such glasses is to choose the interactions randomly from a (often symmetric) probability distribution, like in the case of Gaussian spin glass models [23]. In contrast, the planted model is a very special system, as the interactions between particles are random but mutually correlated as they are all generated given the ground-truth $\boldsymbol{\sigma}$. This is why we call this ground-truth $\boldsymbol{\sigma}$ the *planted configuration*. Following [168], we will introduce a short example on random linear equations that shows why this correlations of the particle interactions can influence the system's behaviour drastically. Suppose that $\boldsymbol{y} \in \{0,1\}^m$ is a random vector and $\boldsymbol{A}$ is a random $m \times n$ matrix over $\mathbb{F}_2$. If $m > n$, the random system of linear equations $\boldsymbol{A}x = \boldsymbol{y}$ has no solution x with high probability [18]. But if we sample a uniform binary vector $\boldsymbol{x}$, calculated $\boldsymbol{y}$ as $\boldsymbol{Ax}$, then the system of linear equations will always feature at least one solution, namely $\boldsymbol{x}$. In our setting, $\boldsymbol{x}$ corresponds to the planted ground-truth while the teacher shows the student $\boldsymbol{A}$ and $\boldsymbol{y}$ (and the information that $\boldsymbol{x}$ and $\boldsymbol{A}$ are independently and uniformly chosen). Now the student's inference task is to infer $\boldsymbol{x}$ from $(\boldsymbol{A}, \boldsymbol{y})$.

The principle of *planting* did not only appear in statistical physics. For instance, it has been used to prove a hardness result on the planted clique problem [74, 105] or within proofs on the geometry of the solution space of random CSPs [1]. Probably one of the most studied planted models is the *stochastic block model* which is a widely used model for community detection on networks [62, 85, 95]. To describe it in the framework at hand, given n individuals $V = \{x_1 \dots x_n\}$, a teacher generates q communities, thus a coloring $\boldsymbol{\sigma} \in [q]^n$ assigning each individual a community. Now, a random graph $\boldsymbol{G} = (V, E)$ is generated as follows. Each edge $x_i x_j$ is present with probability p_1 if $\boldsymbol{\sigma}_i = \boldsymbol{\sigma}_j$ and with probability p_2 if $\boldsymbol{\sigma}_i \neq \boldsymbol{\sigma}_j$. Clearly, if $p_1 = p_2$, this is just a binomial random graph containing no information about the communities. If $p_1 \gg p_2$, it is more likely to observe edges within one community and if $p_2 \gg p_1$, most edges are expected between different communities. Of course, there are various generalisations to this problem. Nevertheless, the student's task is, given $(\boldsymbol{G}, p_1, p_2, q)$ to infer a $q-$coloring $\tilde{\sigma}$ that has the highest possible overlap with $\boldsymbol{\sigma}$.

As one can conclude, planted models are a fairly common used technique in statistical inference. Studying the corresponding physical system whose Boltzmann distribution equals the posterior distribution obtained by the student might bring together powerful tools from different fields of research. For the sake of convenience and as this thesis's contributions on statistical inference require this setup, we tacitly assume that the Boltzmann distribution of the planted model can be expressed by a sparse random factor graph, such that it has the form (1.1.16). As we previously studied different phase transitions, mostly corresponding to the geometry of the solution space of such random CSPs, it is not very surprising that there are important phase transitions in statistical inference problems as well.

1.2.1.3. Phase transitions in statistical inference

Suppose being in the Bayes optimal setting, thus, the student's guess ν is exactly the teacher's prior and the student's model knowledge $\tilde{\nu}$ equals the teacher's model generating distribution. Furthermore, suppose that we observe the underlying physical system in the thermodynamic limit ($n \to \infty$). Moreover, let $\boldsymbol{\sigma} \in \Omega^n$ for some finite set Ω. In principle, the following discussion can be extended to more general ground-truth domains. For two configurations $\sigma, \tau \in \Omega^n$, we denote the overlap $\langle \sigma, \tau \rangle$ as the number of coordinates in which σ and τ coincide. Let q_0 denote the expected overlap between a uniformly at random chosen $\boldsymbol{\tau}$ from ν and $\boldsymbol{\sigma}$, formally

$$q_0 = \mathbb{E}_{\boldsymbol{\tau} \sim \nu} \langle \boldsymbol{\tau}, \boldsymbol{\sigma} \rangle.$$

Given data $\hat{\boldsymbol{\sigma}}$ as well as ν and $\tilde{\nu}$, the student's task might be evaluated with respect to two levels of reconstruction.

- Is the student able to guess $\tilde{\sigma}$ such that $\langle \tilde{\sigma}, \boldsymbol{\sigma} \rangle > q_0$? *(weak reconstruction)*
- Is the student able to infer $\boldsymbol{\sigma}$? *(strong reconstruction)*

With the statistical inference tasks being part of this thesis, we are only interested in the strong reconstruction scenario. Now, this question can be answered under two different sets of restrictions.

- Reconstruction is *information-theoretically* possible, if the student can infer $\boldsymbol{\sigma}$ from $(\hat{\boldsymbol{\sigma}}, \nu, \tilde{\nu})$ given unlimited computational power.
- Reconstruction is *algorithmically* possible, if there is a polynomial-time algorithm $\mathscr{A}$ that outputs $\boldsymbol{\sigma}$ on input $(\hat{\boldsymbol{\sigma}}, \nu, \tilde{\nu})$.

Clearly, the planted model undergoes phase transitions with respect to this questions. We suppose that the planted model comes as a factor graph $\mathscr{G}$ with n variables and m factor nodes and that the student has access to this graph. With a slight misuse of notation, we suppose that the student gains knowledge about $\tilde{\nu}$ as well, if she has access to the factor graph. Furthermore, on each factor node a we find a weight function ψ_a such that $\psi_a(\partial a) = \hat{\boldsymbol{\sigma}}_a$. As before, let $\alpha = m/n$ be the factor-to-variable ratio. In this case, the more factors there are (as α getting large), the more measurements of the ground-truth are available for gathering information and thus, the easier the task seems to become. We denote by α_{inf} the information-theoretic threshold and by α_{alg} the algorithmic threshold. Then we have the following, assuming tacitly that all thresholds might be either strict or coarse phase transitions as already discussed.

- For $\alpha < \alpha_{\mathrm{inf}}$, there is no algorithm (efficient or not) that is able to infer $\boldsymbol{\sigma}$ from $(\mathscr{G}, \hat{\boldsymbol{\sigma}}, \nu)$.
- For $\alpha_{\mathrm{inf}} < \alpha < \alpha_{\mathrm{alg}}$, there is no efficient algorithm that is able to infer $\boldsymbol{\sigma}$ from $(\mathscr{G}, \hat{\boldsymbol{\sigma}}, \nu)$.
- If $\alpha > \alpha_{\mathrm{alg}}$, there is a polynomial-time algorithm $\mathscr{A}$ that outputs $\boldsymbol{\sigma}$ on input $(\mathscr{G}, \hat{\boldsymbol{\sigma}}, \nu)$.

If α^* is some threshold, we will subsequently refer to negative results (*for $\alpha < \alpha^*$ inference is not possible*) as *converse statements* and to positive results (*for $\alpha > \alpha^*$ inference is possible*) as *achievability statements* respectively. When mentioning algorithmic achievability, a natural question arises: which class of algorithms is supposed to perform well on inference problems? We make use of the observation that the Boltzmann distribution of the planted model can be tackled by the message passing algorithms of Section 1.1.4. While Belief Propagation is indeed an efficient algorithm in a complexity theoretical way, in each iteration at each factor a, we need to compute roughly $\deg(a)$ messages. If the underlying graph is not too sparse, this might not be feasible computationally on large instances. Suppose we have weight functions that are not too sensitive to single messages' contributions. Then a family of algorithms called *approximate message passing algorithms* is supposed to perform well.

1.2.1.4. Approximate Message Passing AMP

Let us first discuss what a sensitive weight function is. A prime example would be the weight functions occurring in the random $k-$SAT problem where a single message can turn the evaluation from nearly zero to almost 1. On the other hand, a weight function that, for instance, counts the adjacent variables with spin 1, would be very insensitive – at least if the factor node degree is large. Thus, suppose the latter is the case and suppose further that the average degree of a factor $k = \omega(1)$ is large. Instead of calculating k different messages (depending on which variable is removed) at each factor as in Belief Propagation, one could compute one message where no variable is removed and send it to all neighbours. Donoho, Maleki and Montanari [66] introduced a family of message passing algorithms which is, intuitively speaking, an approximate version of Belief Propagation.

More precisely, let $\mathscr{G} = (V \cup F, E)$ be a factor graph representing a statistical inference problem on n variables and m factors with $\alpha = m/n$. For a vector $\tau \in \mathbb{R}^k$ we write $\langle \tau \rangle = k^{-1} \sum_{i=1}^{k} \tau_i$ for the average over all entries of τ. Denote by $\{\eta_t : \mathbb{R}^n \to \mathbb{R}^n\}_{t \geq 1}$ a family of coordinate-wise applied (non-linear) threshold functions and let $A \in \mathbb{R}^{m \times n}$ be the (normalised) adjacency matrix of the factor graph, thus normalise the columns to ℓ_2 norm 1. Then, approximate message passing starts with an initial guess $\sigma^{(0)}$ of the ground-truth $\boldsymbol{\sigma}$ and computes iteratively

$$\sigma^{(t+1)} = \eta_t\left(A^T z^{(t)} + \sigma^{(t)}\right) \qquad \text{and} \qquad z^{(t)} = \hat{\boldsymbol{\sigma}} - A\sigma^{(t)} + \alpha^{-1} z^{(t-1)} \left\langle \eta_t'\left(A^T z^{(t-1)} + \sigma^{(t-1)}\right)\right\rangle. \tag{1.2.4}$$

Let us only briefly sketch the meaning of the single parts of (1.2.4), a complete introduction and a formal justification of those equations is provided by [66, 129].

- $\sigma^{(t)}$ is the current estimate for the ground-truth $\boldsymbol{\sigma}$.
- $z^{(t)}$ can be interpreted as a current residual.
- The threshold function η pushes $\sigma^{(t)}$ towards the sparsest solution. In its absence, the algorithm would converge to a solution of $\hat{\boldsymbol{\sigma}} = A\boldsymbol{\sigma}$ of least ℓ_2 norm.
- $\alpha^{-1} z^{(t-1)} \left\langle \eta'\left(A^T z^{(t-1)} + \sigma^{(t-1)}\right)\right\rangle$ is derived from the Belief Propagation update rules on the corresponding factor graph. It improves the convergence towards sparse solutions even further.

While AMP is clearly fast to run and easy to implement, it also achieves the best algorithmic performance presently known in some of the most prominent inference problems like compressed sensing [67] or the pooled data problem [70].

After this excursion into the statistical physics' foundations of statistical inference, we will introduce the *group testing* problem in the next section. Group testing is a prime example of a statistical inference problem and is the protagonist of multiple contributions of this thesis.

1.2.2. Group Testing

In the group testing problem, one is given n individuals $x_1, \dots, x_n$ out of which a small number k is infected. We may employ a testing procedure that allows to pool various individuals into one *group test* that renders a positive result if and only if at least one infected individual is contained in the test. Given probes of those n individuals and the prevalence k/n, the ultimate goal is to find a testing strategy (we will refer to this as a *pooling scheme*) that is able to infer the infection status of each individual with the minimum number of tests possible. Group testing itself found its first appearance in literature in the early 1940's when Dorfman [68] proposed the following, fairly simple, inference algorithm.

(D1) Assign a group of Γ individuals to a test, such that each individual gets tested once.

(D2) If a test renders a negative result, all of the contained individuals are uninfected. If a test renders a positive result, test all individuals individually.

Supposing that the prevalence is k/n and each individual is infected independently of all other individuals, it is straightforward to calculate the expected number of tests $\boldsymbol{m}$ required in this testing scheme as

$$\mathbb{E}[\boldsymbol{m}] = \frac{n}{\Gamma} + n\left(1 - \left(1 - \frac{k}{n}\right)^{\Gamma}\right).$$

Given an estimation of the prevalence (this is the teacher's prior), it is possible to optimise the test size Γ in order to minimise the expected number of tests.

Even if the Dorfman scheme is, as we will see, a suboptimal design, it finds its applications in various medical applications which might be due to the very simple inference algorithm and the fact that it recovers the infection status of each individual correctly (supposing that each test outputs the correct result) [130].

Since its first appearance, group testing gained a lot of attention in various installments. In the first decades, the focus was lying on *combinatorial* group testing, where one aims to construct a pooling scheme that successfully recovers every possible ground-truth from the teacher's prior distribution. This problem was studied intensively, amongst others, by D'yachkov et al. [58], Erdős and Rényi [73], Fischer, Klasner and Wegener [80], Hwang [98] and Ungar [164]. But in the early 2000's, the focus changed to so-called *probabilistic* test-designs in which inference is only required with high probability with respect to the random choice of the ground-truth. Some of the most influential contributions are due to Aldridge, Baldassini and Johnson [8], Aldridge, Johnson and Scarlett [9], Damaschke [59], Gandikota et al. [86], Johnson, Aldridge and Scarlett [108], Mezard and Toninelli [132], Mézard, Tarzia and Toninelli [133] and Scarlett and Cevher [154]. Let us describe more systematically which kind of group testing problems exist in literature. The exact problem description may vary with respect to ...

- ... the teacher's prior $\boldsymbol{\sigma} \in \{0,1\}^n$:

  - $\boldsymbol{\sigma}$ can be a uniformly sampled configuration out of all configurations with exactly k non-zero entries *(hypergeometric group testing model)*,
  - Alternatively, $\boldsymbol{\sigma}$ might be a binomial random vector, such that each entry equals 1 independently of all other entries with probability k/n *(i.i.d. group testing model)*.
  - Finally, each individual x_i might be (independent of the other individuals) infected with probability p_i such that p_i and p_j do not need to be equal necessarily. This choice of the teacher's prior is called *group testing with priors*.

- ... the number of subsequent rounds a test-design may contain:
  - In the *non-adaptive* group testing problem all tests need to be conducted in parallel, thus one cannot use information gained in previous stages.
  - The *adaptive* group testing problem allows to design subsequent stages of tests based on the outcome of previous stages.

- ... the level of required certainty:
  - In *combinatorial group testing*, a testing scheme needs to output the correct infection status of all individuals on *any* ground-truth $\boldsymbol{\sigma}$.
  - On the other hand, in *probabilistic group testing*, it suffices to recover $\boldsymbol{\sigma}$ with high probability (with respect to the randomly generated ground-truth).

- ... the type of recovery:
  - If we demand *exact recovery*, each individual has to be assigned the correct infection status.
  - Otherwise, if we can tolerate a small number of falsely classified individuals, we call this criterion *partial recovery*.

- ... the correctness of each test:
  - In *noiseless group testing*, each test outputs the correct result.
  - *Noisy group testing* instances are characterised by a random flip of each test-result. This noise might be uniformly at random (binary symmetric channel), dependent of the tests' correct result (e.g. $Z-$channel and reverse $Z-$channel) or correlated with the number of infected and uninfected individuals in the test (diluted noise models).
  - In *threshold group testing*, a test-result is negative if the number of contained infected individuals is below a given threshold t_1 and positive if it exceeds a second threshold $t_2 \geq t_1$. In the range (t_1, t_2), the test-result might be randomly chosen.

- ... the constraints on individuals-per-test and tests-per-individual:
  - In the *unrestricted group testing problem* each individual might take place in an arbitrary number of tests. Furthermore, each test might contain between one and all individuals. For the sake of brevity, we will refer to the unrestricted group testing problem as the *group testing problem*.
  - In $\Delta-$*divisible sparsity constrained group testing*, each individual might take place in a maximum of Δ tests.
  - Analogously, in the $\Gamma-$*sparse group testing problem*, each tests may contain at most Γ individuals.

- ... the prevalence of infected individuals:
  - If the number of infected individuals k satisfies $k = \Theta(1)$, thus is independent of the number of individuals n, we call the setting the *ultra-sparse* regime.
  - If, on the other hand, $k \sim n^{\theta} (\theta \in (0,1))$, the regime is called *sublinear*.

– Finally, if we suppose $k = \alpha n$ for some constant $\alpha \in (0,1)$, the setting is denoted as the *linear* regime.

In this thesis's contributions, we will analyse phase transitions in **noiseless probabilistic hypergeometric group testing** instances with a focus on non-adaptive group testing. Nevertheless, some contributions contain results on adaptive group testing as well. We will discuss this thesis's results in detail in Section 2.1. Let us briefly express a non-adaptive group testing instance in terms of the statistical physics' framework. A pooling scheme for such an instance can be represented as a factor graph $\mathcal{G} = (V \cup F, E)$. We denote with $V = \{x_1, \dots, x_n\}$ the n individuals and the m tests are given by $F = \{a_1, \dots, a_m\}$. Furthermore, an edge $x_i a_j$ exists if and only if individual x_i takes part in test a_j. As we are in the setting of inference, suppose that the ground-truth $\boldsymbol{\sigma} \in \{0,1\}^n$ assigns each individual its infection status. Furthermore, let $\hat{\boldsymbol{\sigma}} \in \{0,1\}^m$ denote the sequence of test-results, hence

$$\hat{\boldsymbol{\sigma}}_a = \max_{x \in \partial a} \boldsymbol{\sigma}_x.$$

This completely describes the planted model introduced via (1.2.3), as we, as usual in group testing, suppose that we have complete knowledge about the model's generation and the teacher's prior (Bayes optimal setting). A visualisation can be found in Figure 1.3. In the following, we will use a slightly different notation as in the previous sections on phase transitions. While we previously denoted by α the factor-to-variable ratio and analysed the system's behaviour with respect to the size of α, it is usual in the group testing community to express phase transitions in terms of the required number of tests $m = m(n,k)$ in the large-system limit $n \to \infty$. Of course, those statements can directly be translated into a statement about the factor-to-variable ratio.

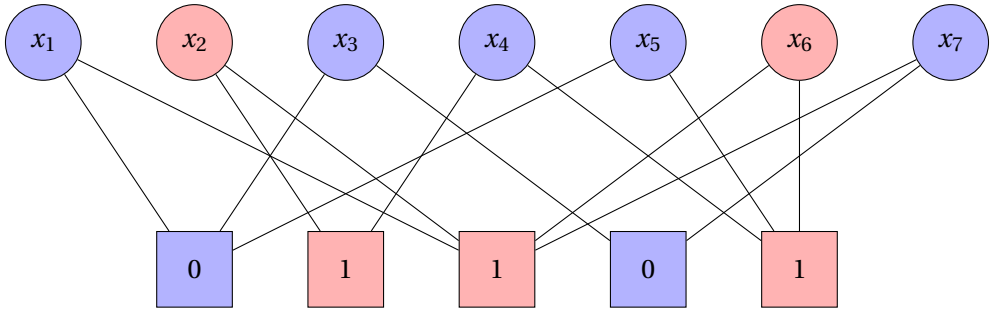

Figure 1.3.: The factor graph representation $G = (V \cup F, E)$ of a group testing instance with $n = 7$ individuals out of which $k = 2$ are infected on $m = 5$ tests. Blue individuals are uninfected while red individuals are infected. Furthermore, a test renders a positive result if and only if at least one infected individual is contained.

Clearly, if the problem was to be studied for finite n, the different levels of prevalence would not be well defined. It is, for instance, not possible to distinguish in a population of $n = 10^3$ individuals whether the occurrence of $k = 20$ infected individuals should be described as $k = 20, k \approx n^{0.43}$ or as $k = 0.02n$, thus results obtained in the large-system limit need to be verified empirically for small n if they should be applied in real laboratories.

Next, we will shortly describe known results and open problems ahead of this thesis's contributions. As this overview is with respect to noiseless probabilistic group testing, we refer the interested reader to the detailed overview article of Aldridge, Johnson and Scarlett [10] for an overview over the state of the play in different models.

1.2.2.1. Prior results on noiseless probabilistic hypergeometric group testing

We will tacitly assume throughout this section that all results are meant to be read with respect to exact recovery if not stated differently. Let us start this section by a folklore counting argument which is a good indication for how many tests we need to use to have a chance of exact recovery. Namely, independent of the choice of the pooling scheme (adaptive or non-adaptive), by making m tests, we can

generate 2^m different sequences of test-results. Clearly, this number needs to exceed the number of possible ground-truth values $\binom{n}{k}$, therefore, if m_{inf} is the minimum number of tests necessary to solve the problem information-theoretically, we find

$$2^{m_{\text{inf}}} \geq \binom{n}{k} \quad \Leftrightarrow \quad m_{\text{inf}} \gtrsim \frac{n\ln n - (n-k)\ln(n-k) - k\ln k}{\ln 2}. \tag{1.2.5}$$

Therefore, with $H(\alpha)$ denoting the entropy of $\mathbf{Be}(\alpha)$ and $\alpha, \theta \in (0,1)$, the counting bound (1.2.5) yields

$$m_{\text{inf}} \gtrsim \begin{cases} \frac{H(\alpha)}{\ln 2} n, & k = \alpha n \\ \frac{1-\theta}{\ln 2} k \ln n, & k \sim n^\theta. \end{cases}$$

Thus, if the prevalence is constant, we require linearly many group tests but if the spread of the disease scales sublinearly in the population size, we can do much better. The counting bound (1.2.5) gives a lower bound on the number of tests required in any testing scheme with respect to a hypergeometric problem setup, thus it applies for adaptive pooling schemes as well as for non-adaptive pooling schemes. Clearly, it should be easier to come along with m_{inf} tests adaptively rather than non-adaptively. Indeed, it turns out that this intuition is mostly correct.

Adaptive group testing Pretty soon after group testing found its way into mathematical literature, there were some negative results on group testing. More precisely, Ungar [164] proved that under the i.i.d. prior there is a phase transition at prevalence $p = \frac{3-\sqrt{5}}{2} \approx 0.38$. More precisely, there is no test design succeeding at inference of $\boldsymbol{\sigma}$ on less than n tests if the problem's prevalence is larger and on the other hand, if the prevalence becomes smaller, there is a test design achieving inference on at most $n-1$ tests. In the hypergeometric group testing problem, it is conversely conjectured that this phase transition (of course with n replaced by $n-1$) occurs at a prevalence of $1/3$ [80, 97] but until now it was only proven that for prevalence $p > \log^{-1.5}(3) \approx 0.369$ performing $n-1$ individual tests is optimal [149]. Clearly, this number of required tests is far from the counting bound. Assuming there are αn infected individuals, Hwang [98] provides a generalised binary splitting algorithm which was later improved by Allemann [13] such that inference of $\boldsymbol{\sigma}$ is possible with

$$m_{\text{Hwang}} \sim m_{\text{inf}} + \alpha n \qquad \text{and, respectively} \qquad m_{\text{Allemann}} \sim m_{\text{inf}} + 0.255\alpha n \tag{1.2.6}$$

tests.

We stress that in the linear regime it turns out that m_{Allemann} exceeds n at roughly $\alpha = 1/3$, underlying the conjecture of Hu, Hwang and Wang [97] and that in the *sublinear regime*, we find

$$m_{\text{Hwang}} \sim m_{\text{Allemann}} \sim m_{\text{inf}},$$

as the deviations are of lower order.

Unfortunately, the binary splitting approach comes with a practical flaw - in order to guarantee successful inference, one requires $\Omega(\ln n)$ rounds of adaptive tests. A natural question is if there might be algorithms succeeding at inference of $\boldsymbol{\sigma}$ with a bounded number of testing rounds. We underline that, with a slight misuse of wording, we tacitly assume that an algorithm is always efficient, thus runs in polynomial time, whenever we do not explicitly state differently. Up to the best of our knowledge there is no algorithm known which succeeds on an arbitrary group testing instance in less than $\Omega(\ln n)$ rounds on m_{Allemann} tests. The situation becomes much more convenient if the problem gets sparser.

More precisely, suppose we have a vanishing prevalence, thus $k \sim n^\theta$ for some $\theta \in (0,1)$. The binary splitting algorithm by Allemann clearly performs asymptotically optimal in this sublinear regime, as already discussed. But now, there are well known efficient algorithms achieving the same asymptotic performance in much less rounds, the probably best known algorithms are due to Damaschke and Muhammad [60] whose algorithm achieves inference in not more than 4 rounds and Scarlett [153] whose algorithm needs only two rounds in the described setting with no more than $(1+o(1))m_{\text{inf}}$ tests,

improving on the 2-stage algorithm of Mezard and Toninelli [132] requiring $\frac{(1+o(1))m_{\inf}}{\ln 2}$ tests.

Therefore, the adaptive group testing problem is, up to the exact phase transition point in hypergeometric linear group testing, well understood. Things turn out to be completely different for non-adaptive group testing.

Non-adaptive group testing Let us again start with discussing known results for a constant prevalence $k = \alpha n$. It turns out that this case is, from a mathematical viewpoint, completely uninteresting. Due to Bay, Price and Scarlett [24] it is known that recovery of the ground-truth is impossible with fewer than $(1-\varepsilon)n$ tests for any $\varepsilon, \alpha \in (0,1)$. The authors built up on work by Aldridge [7] who established a coarse phase transition, thus proved that recovery under the given circumstances fails with positive probability.

Therefore, we let our focus be the sublinear regime in which we suppose throughout that the prevalence is given by $k/n = n^{-(1-\theta)}$ for some density parameter θ. Thus, if θ becomes larger, the prevalence gets higher and the group testing problem is said to become denser.

We will first introduce two specific non-adaptive pooling schemes. Given the teacher's prior in the hypergeometric model, we might construct a random graph, thus the teacher's model, as follows.

- *Bernoulli testing:* Each individual takes place in any test with probability p independently.
- *Random (almost) regular testing:* Each individual chooses Δ tests uniformly at random without (or with, respectively) replacement.

It turns out that both model choices have some similarities but that the first is inferior to the second. We will verify this fact in Section 2.1, which is basically due to high fluctuations of the individual degree. Indeed, the information-theoretically optimal designs require each test to be positive with probability $1/2$ implying that $\Delta = \Theta(\ln n)$, or $p = \Theta\left(k^{-1}\right)$ respectively. The choice that any test needs to be positive with probability $1/2$ is, intuitively speaking, due to the fact that this choice maximises the system's entropy, thus the gain on information per test is maximised.

While there was no universal converse statement sharpening $m_{\inf}$ known prior to this thesis's contributions, there are several algorithmic and information-theoretical achievability as well converse statements on those two random models. We will discuss the most influential ones. Regarding Bernoulli testing, it were Scarlett and Cevher [152] who proved that it is information-theoretically possible to infer the ground-truth with $(1+o(1))m_{\inf}$ tests if the group testing instance is fairly sparse, thus $\theta \leq 1/3$. This result was strengthened by Aldridge [11] who pinned down the information-theoretic strict phase transition of the Bernoulli group testing design at

$$m_{\text{Bernoulli}} = \frac{1}{c_{\text{Ber}} \ln 2} k \ln \frac{n}{k} \qquad \text{where} \qquad c_{\text{Ber}} = \max_{\nu>0} \min \left\{ \frac{(1-\theta)\nu \exp(-\nu)}{\theta \ln 2}, \frac{H\left(\exp(-\nu)\right)}{\ln 2} \right\} k \ln \frac{n}{k} \tag{1.2.7}$$

which is strictly worse than (1.2.5) for all $\theta > 1/3$. Subsequently, Aldridge, Johnson and Scarlett [9] provided an information-theoretical converse statement (though, no achievability result) in the random regular model. More precisely, inference of $\boldsymbol{\sigma}$ fails with positive probability, if the test-design contains less than

$$m_{\text{rand-reg}} = \max \left\{ \frac{\theta}{(1-\theta)\ln^2 2}, \frac{1}{\ln 2} \right\} k \ln \frac{n}{k} \tag{1.2.8}$$

tests. A short calculation verifies that $m_{\text{rand-reg}} < m_{\text{Bernoulli}}$ for $\theta \in \left(\frac{1}{3}, 1\right)$ and $m_{\text{rand-reg}} = m_{\inf}$ for $\theta \leq \frac{\ln 2}{1+\ln 2} \approx 0.409$ but nevertheless, $m_{\text{rand-reg}} > m_{\inf}$ for larger θ. Thus, the random regular model might outperform the Bernoulli testing but it is clearly far from the counting bound for a high prevalence. Such an information-theoretic gap might be due to the model itself or it might be the case that non-adaptive pooling schemes cannot perform at $m_{\inf}$. We will discuss in Section 2.1 that the latter is the case and actually $m_{\text{rand-reg}}$ is a universal information-theoretic converse, independent of the pooling scheme, for any non-adaptive design. We will see furthermore that the converse statement (1.2.8) of [9] actually marks an information-theoretic phase transition point.

At this point, we make a very short excursion into the setting of partial recovery. The already mentioned paper of Scarlett and Cevher [152] actually proves that m_{inf} is an important threshold for partial recovery models. More precisely, the simple Bernoulli test design suffices to recover all but γk individuals correctly with m_{inf} tests. On the negative side, no test-design can come along with less than $(1-\gamma)m_{\text{inf}}$ tests when trying to recover all but γk individuals correctly.

Let us come back to the problem of exact recovery. While we already discovered the state of the play prior to this thesis's contributions with respect to information-theoretic aspects, we will now introduce three of the most prominent non-adaptive group testing algorithms and state known results about their performances. In detail, we will introduce the COMP algorithm as well as the DD algorithm and its greedy extension called SCOMP.

The two probably most basic algorithms are COMP and DD and their descriptions can be found in Algorithms 1 – 2.

Input: Pooling scheme $\mathcal{G} = (V \cup F, E)$, test-results $\hat{\boldsymbol{\sigma}} \in \{0,1\}^m$
Output: Estimate $\tilde{\sigma}$ of $\boldsymbol{\sigma}$
1 Mark all individuals occurring in a negative test as uninfected.
2 Declare all other individuals as infected.

Algorithm 1: The COMP algorithm as first introduced by Chan et al. [39].

Input: Pooling scheme $\mathcal{G} = (V \cup F, E)$, test-results $\hat{\boldsymbol{\sigma}} \in \{0,1\}^m$
Output: Estimate $\tilde{\sigma}$ of $\boldsymbol{\sigma}$
1 Mark all individuals occurring in a negative test as uninfected and remove them and the corresponding negative test from the graph.
2 Mark an individual as infected if it appears as the only individual in a positive test in this reduced graph.
3 Declare all other individuals as uninfected.

Algorithm 2: DD algorithm as defined by Aldridge, Baldassini and Johnson [8].

While COMP cannot produce any false negatives, thus $\tilde{\sigma}_i^{\texttt{COMP}} = 0 \Rightarrow \boldsymbol{\sigma}_i = 0$, DD guarantees that all declared infected individuals are indeed infected, hence $\tilde{\sigma}_i^{\texttt{DD}} = 1 \Rightarrow \boldsymbol{\sigma}_i = 1$. Nevertheless, it might happen that the estimate $\tilde{\sigma}$ does not even *explain* the test-results $\hat{\boldsymbol{\sigma}}$. In this context, we say that an individual $x \in \partial a$ explains test a under $\tilde{\sigma}$ if $\hat{\boldsymbol{\sigma}}_a = 1$ and $\tilde{\sigma}_x = 1$. Conversely, a positive test is called explained by $\tilde{\sigma}$ if there is at least one individual $x \in \partial a$ which explains a. With COMP or DD it might furthermore happen that $\tilde{\sigma}$ contains less than k infected individuals (if DD was applied). In terms of inference, we observe that the estimates of COMP and DD do not necessarily belong to the solution space of the underlying random CSP. If we look a bit closer into the DD-algorithm, we find that the first two steps do not misclassify any individual. We furthermore observe that the estimate was correct if after the first step of DD, there is no individual left that does not belong to at least one (positive) test of degree one. We will formalise this observation in Section 2.1. Suppose that this does not hold and thus, after the second step of DD, we are left with some unexplained tests. As the prevalence is small, it might be a natural idea to declare greedily those individuals as infected that explain the most unexplained tests. This is exactly what the SCOMP-algorithm does.

Positive news about SCOMP is clearly that it produces an estimate $\tilde{\sigma}$ which explains the test-results $\hat{\boldsymbol{\sigma}}$. The flaw is, of course, that it might produce false positive as well as false negative predictions. Nevertheless, it was conjectured based on simulations that SCOMP performs better than DD does [8], thus requires less tests to succeed at inference of $\boldsymbol{\sigma}$. We will see in Section 2.1 that this conjecture turned out to be actually false.

Algorithms 1 – 3 can be applied to any arbitrary (non-adaptive) pooling scheme. Nevertheless, they were studied on the Bernoulli model [8] as well as the random regular model [108]. As it turns out that those algorithms require less tests on the random regular model in order to infer $\boldsymbol{\sigma}$ with high probability, we focus on those results from [108]. More precisely, the authors prove a strict phase transition of the

Input: Pooling scheme $\mathcal{G} = (V \cup F, E)$, test-results $\hat{\boldsymbol{\sigma}} \in \{0,1\}^m$
Output: Estimate $\tilde{\sigma}$ of $\boldsymbol{\sigma}$
1 Mark all individuals occurring in a negative test as uninfected and remove them and the corresponding negative test from the graph.
2 Mark all individuals that are the sole individual in a test in this reduced graph as infected.
3 **while** *there is an unexplained test* a **do**
4 Take the individual of highest degree, breaking ties arbitrarily, and declare it as infected.
5 Remove all adjacent tests from the graph.
6 Declare all left individuals (now, isolated in the reduced graph) as uninfected.

Algorithm 3: The SCOMP algorithm by Aldridge, Baldassini and Johnson [8] can be seen as a greedy extension of DD.

COMP-algorithm on the random regular model at $m_{\mathtt{COMP}}$ as well as an achievability result for DD at $m_{\mathtt{DD}}$ where

$$m_{\mathtt{COMP}} = \frac{1}{(1-\theta)\ln^2 2} k \ln \frac{n}{k} \qquad \text{and} \qquad m_{\mathtt{DD}} \leq \max\left\{\frac{\theta}{(1-\theta)\ln^2 2}, \frac{1}{\ln^2}\right\} k \ln \frac{n}{k}. \tag{1.2.9}$$

Clearly, if DD achieves inference at $m_{\mathtt{DD}}$, so does SCOMP as it performs the two first steps of DD. Comparing the performance of the DD algorithm on the random regular model with its information-theoretic converse (1.2.8), we find that DD is an optimal inference algorithm on the random regular model for $\theta \geq 1/2$ while there remains a gap for smaller θ. This might have been due to a weakness in the achievability proof of DD, a weakness in the information-theoretic converse or because DD does not perform best possible in this regime. As we will see in Section 2.1, the latter is the case.

Let us, at this point, briefly discuss the DD algorithm itself. While it was first stated in its comfortable and easy to digest version in [8], it turns out that the estimate $\tilde{\sigma}$ of $\boldsymbol{\sigma}$ coincides with the estimate computed by the well known Warning Propagation algorithm. Such message passing algorithms were already applied by Mézard, Tarzia and Toninelli [133] to the group testing problem. Indeed it turns out that the estimates of DD and WP coincide. If an individual x is part of a negative test, WP sends the warning to x that it may not take value 1, thus we just need to analyse the messages at positives test. If now there is a positive test a of size Γ_a containing $\Gamma_a - 1$ individuals being part of a (different) negative test, the message sent to the last individual warns this individual not to take value 0 as well, as otherwise the test was unexplained. Finally, if there were two individuals in a not being warned about taking the value 0, the test won't send a warning. The possibility to write down (equivalent) forms of the statistical physics' message passing algorithms seems to be a key feature of group testing and similar problems as we will see in due course more often.

Until now, all presented (non-adaptive) designs and algorithms have one thing in common. If we restrict ourselves to the sublinear regime, each individual takes part in $\Omega(\ln n)$ tests and there are tests containing at least $\Omega(n/k)$ tests. Understanding the group testing problem under given restrictions on the maximum degrees of individuals and tests is not only a challenging problem but it might influence the group testing schemes used in real-world laboratories. Thus, let us introduce *sparsity constrained group testing*.

Sparsity constrained group testing As before, we restrict ourselves to the noiseless case. We distinguish between $\Gamma-$sparse group testing in which each test may contain at most Γ individuals and $\Delta-$divisible group testing whose restriction is that each individual may only be tested at most Δ times. Clearly, these models try to built up real world conditions in the sense that the test's sensitivity might decrease with very large pools ($\Gamma-$sparse) or that it is not possible to duplicate the patient's samples arbitrarily often.

Probably the most influential prior work to this thesis's contributions is the one of Gandikota et al. [86]. They stated (universal) information-theoretic converse results in both restriction models and achievability results using the COMP algorithm on the random regular pooling scheme. Let us denote

by $m_{\inf,\mathrm{G}}(\Gamma)$ and $m_{\inf,\mathrm{G}}(\Delta)$ the information-theoretic converse bounds and by $m_{\mathtt{COMP},\mathrm{G}}(\Gamma)$ and $m_{\mathtt{COMP},\mathrm{G}}(\Delta)$ the achievability bounds of COMP in the $\Gamma-$sparse and $\Delta-$divisible setting respectively obtained in [86]. More precisely, the authors find for $\Gamma = \Theta\left(\left(\frac{n}{k}\right)^{\beta}\right)$ for some $\beta \in [0,1)$ that

$$m_{\inf,\mathrm{G}}(\Gamma) = \frac{1}{1-\beta}\frac{n}{\Gamma} \qquad \text{and} \qquad m_{\mathtt{COMP},\mathrm{G}}(\Gamma) = \left\lceil \frac{1}{(1-\theta)(1-\beta)} \right\rceil \left\lceil \frac{n}{\Gamma} \right\rceil . \tag{1.2.10}$$

Furthermore, with $\Delta = o(\ln n)$ they prove weak converse respectively achievability statements at

$$m_{\inf,\mathrm{G}}(\Delta) = \Delta k \left(\frac{n}{k}\right)^{1/\Delta} \qquad \text{and} \qquad m_{\mathtt{COMP},\mathrm{G}}(\Delta) = \left(\mathrm{e}\Delta k n^{1/\Delta}\right). \tag{1.2.11}$$

Comparing the achievability result with the converse statement, we observe a sizeable gap in the $\Delta-$divisible setting whilst in the $\Gamma-$sparse case the gap is only a multiplicative constant factor. We will improve on the converse statements as well as provide a rigorous analysis of the DD algorithm in a tailor-made pooling scheme which improves the achievability bounds in Section 2.1.

After having presented a prime example of a statistical inference problem in large planted versions of statistical physics' models, let us return to the question of how to express the physics' intuition behind the handling of such large random CSPs in a rigorous way. This might help studying random CSPs as well as their planted versions, hence to study statistical inference problems.

1.3. Large discrete systems and their limits

The purpose of this section is two-fold. First, we will dive deeper into the already defined cut-distance for probability measures and give an overview of recent results prior to this thesis's contributions, for instance we will present a regularity lemma for such measures. We will see that this notion is highly inspired by graph regularity and the cut-distance used in graph limit theory and give a very gentle and short introduction into this field as well. Second, we will slightly change the point of view on large graphs from diluted mean-field models (random graphs) to deterministic graphs which are perturbed slightly with a little bit of randomness. The latter ones are useful structures in order to study the expected behaviour of algorithms or to obtain structural results on real world occurrences of large graphs.

1.3.1. Approaching pure states of spin glass systems: the cut-distance

We already learned about the cut-distance $\Delta_{\boxtimes}$ in Section 1.1.2.4. For the sake of the reading flow, recall that it was defined in (1.1.14) for two probability measures μ, ν on some finite set Ω^n as

$$\Delta_{\boxtimes}(\mu,\nu) = \inf_{\substack{\gamma\in\Gamma(\mu,\nu),\\ \phi\in\mathbb{S}_n}} \sup_{\substack{S\subset\Omega^n\times\Omega^n,\\ X\subset[n],\\ \omega\in\Omega}} \left| \sum_{\substack{(\sigma,\tau)\in S,\\ x\in X}} \gamma(\sigma,\tau)\left(\mathbf{1}\{\sigma_x=\omega\} - \mathbf{1}\{\tau_{\phi(x)}=\omega\}\right)\right|,$$

where $\Gamma(\mu,\nu)$ is the set of couplings of μ and ν and $\mathbb{S}_n$ is the set of permutations on $[n]$. Let us write $\mathscr{P}(\Omega^n)$ for the set of probability measures over Ω^n. It can be easily verified that $\Delta_{\boxtimes}(\cdot,\cdot)$ satisfies the triangle-inequality and is symmetric on $\mathscr{P}(\Omega^n)$ but it might happen that $\Delta_{\boxtimes}(\mu,\nu) = 0$ even if $\mu \neq \nu$. Therefore, let $\mathscr{L}_n(\Omega)$ be the set of equivalence classes over $\mathscr{P}(\Omega^n)$ such that one class consists of those measures with cut-distance zero. With a slight misuse of notation, we say that $\mu \in \mathscr{L}_n(\Omega)$ is a probability measure on Ω^n rather than representing some equivalence class and now, $\Delta_{\boxtimes}(\cdot,\cdot)$ defines a metric on $\mathscr{L}_n(\Omega)$. As n is supposed to grow to ∞ in typical applications, it might be tempting to introduce some kind of limit theory.

1.3.1.1. The cut-distance in the thermodynamic limit

The cut-distance for probability measures was introduced by Coja-Oghlan, Perkins and Skubch [42] and the authors provided an idea of how to get meaningful limit objects of discrete probability measures

by using this cut-distance. Let us introduce some notation in order to grasp this idea. If $\sigma \in \Omega^n$ is a configuration, we can translate this configuration into a measurable function from $[0,1)$ into the set of probability measures over Ω. Denote by Σ_Ω the space of all measurable functions from $[0,1)$ to $\mathscr{P}(\Omega)$ up to equality almost everywhere. Then, express σ as

$$\hat{\sigma} : [0,1) \to \mathscr{P}(\Omega) \qquad \text{s.t.} \qquad x \mapsto \sum_{i=1}^{n} \delta_{\sigma_i} \mathbf{1}\left\{ x \in \left[\frac{i-1}{n}, \frac{i}{n} \right) \right\}.$$

If now $\mu \in \mathscr{P}(\Omega^n)$ is a probability measure on Ω^n, Coja-Oghlan, Perkins and Skubch [42] define

$$\hat{\mu} = \sum_{\sigma \in \Omega^n} \mu(\sigma) \delta_{\hat{\sigma}} \qquad \text{s.t.} \qquad \hat{\mu} \in \mathscr{P}(\Sigma_\Omega). \tag{1.3.1}$$

Thus, μ and $\hat{\mu}$ are in 1-to-1-correspondence. Now it is possible to equip $\mathscr{P}(\Sigma_\Omega)$ with a corresponding continuous version of the cut-distance. To this end, let $\mathbb{S}_{[0,1)}$ denote the set of all measure-preserving bijections on $[0,1)$ whose inverse is measure-preserving as well, then the cut-distance of two measures $\mu, \nu \in \mathscr{P}(\Sigma_\Omega)$ is defined as

$$D_{\boxtimes}(\mu,\nu) = \inf_{\substack{\gamma \in \Gamma(\mu,\nu), \\ \varphi \in \mathbb{S}_{[0,1)}}} \sup_{\substack{B \subset \Sigma_\Omega^2, \\ U \subset [0,1), \\ \omega \in \Omega}} \left| \int_B \int_U \sigma_x(\omega) - \tau_{\varphi(x)}(\omega) \mathrm{d}x \mathrm{d}\gamma(\sigma,\tau) \right|. \tag{1.3.2}$$

Again, we identify two measures μ, ν if their cut-distance is zero and obtain $\mathscr{L}_\Omega$ as the space of all such equivalence classes. It can be proven that $\mathscr{L}_\Omega$ is a compact metric space [42, Corollary 2.5]. The authors prove this by the compactness of (a special version) of the so-called graphon space [127] and the existence of a homeomorphism from $\mathscr{L}_\Omega$ into the latter due to [102, Theorem 7.1]. Nevertheless, prior to this thesis's contributions, there was no rigorous limit theory for such probability measures. Especially the limit objects (we will call them $\Omega-$*laws* later on) were not described explicitly and the connection between the discrete and continuous cut-distance was not understood for embedded measures constructed via (1.3.1). Furthermore, the exact connection to the graph limit theory was not carried out and some important key features of graph limit theory, like reconstruction of a finite object by *sampling* from a limit object were not known. We will discuss those missing pieces in Section 2.3. As we referred to graph limit theory a couple of times, let us briefly describe this beautiful theory on a fairly short and intuitive level. Lovász [127] provides a very detailed introduction into graph limit theory for the interested reader.

1.3.1.2. A spark of graph limit theory

In a series of seminal papers, Borgs et al. and Lovász and Szegedy [31, 32, 128] introduced a very powerful theory which connects large discrete objects - growing sequences of dense graphs - with analytical functions called graphons which are measurable functions from the unit-square into $[0,1]$. We give a little insight into this theory following Lovász [127]. In this context, we call a weighted graph $G = (V = \{1,\dots,n\}, E, \Psi)$ on $|V| = n$ vertices *dense*, if there is a constant $\varepsilon > 0$ such that $|E| \geq \varepsilon n^2$, thus a positive proportion of all edges is present. Furthermore, $\Psi : V \times V \to [0,1]$ describes the weight of each edge (observe that a $\{0,1\}-$valued function Ψ makes G a simple graph). Without going much into detail, one way of obtaining such limit objects is by using their version of the cut-distance for graphs. We begin by introducing a very intuitive version of this distance, namely for graphs on the same vertex set of size n. We identify a graph G with its adjacency matrix and denote by G_{ij} its entry, then the cut-distance reads

$$\delta_{\boxtimes}(G,H) = n^{-2} \min_{\varphi \in \mathbb{S}_n} \max_{S,T \subset V} \left| \sum_{s \in S, t \in T} G_{s,t} - H_{\phi(s),\phi(t)} \right|. \tag{1.3.3}$$

Intuitively speaking, the cut-distance measures the largest deviation of edges on subsets of vertices under a best possible re-labelling of the vertices. In particular, isomorphic graphs are those graphs with

cut-distance zero. One can imagine that this definition gets quite involved if it is defined for graphs on a different number of vertices. Roughly said, in this case the cut-distance is defined as the minimum over all values of the cut-distance between blow-ups of the graphs that have the same number of vertices [127, Section 8.1.4].

Let us now embed such finite graphs into the space of *kernels*, thus measurable functions $W : [0,1]^2 \to [0,1]$. Denote the space of all such kernels as $\mathscr{K}$. In order to embed a graph we can, figuratively speaking, shrink the adjacency matrix of a graph G from size $n \times n$ onto the unit-square $[0,1]^2$ such that the built function takes a constant value on each little square corresponding to a matrix entry. Formally, if $G = (V,E)$ is a graph on n vertices, we construct W_G as the corresponding $G-$kernel as follows. Let $S_1, \dots, S_\ell$ be a partition of $[0,1]$ into n intervals such that $S_i = [(i-1)/n, i/n)$. Then let

$$W_G(x,y) = \sum_{i,j=1}^{n} \mathbf{1}\{x \in S_i, y \in S_j\} G_{i,j}$$

be the step-function which takes value 1 on square $S_i \times S_j$ if and only if edge i,j is present in G. Of course, this definition easily generalises to weighted graphs as well.

Thus, we can embed any finite graph into the space of kernels $\mathscr{K}$. But clearly, there are much more functions in $\mathscr{K}$ as those that correspond to finite graphs. Following [127, Section 8.2.1], we define the continuous version of the cut-distance for two kernels $W, W' \in \mathscr{K}$ as

$$\mathscr{D}_{\boxtimes}(W,W') = \inf_{\varphi \in \mathbb{S}_{[0,1]}} \sup_{S,T \subset [0,1]} \left| \int_{S \times T} W(x,y) - W'(\varphi(x),\varphi(y)) \mathrm{d}y \mathrm{d}x \right|. \tag{1.3.4}$$

As in the previous section, the continuous version seems to be the intuitive generalisation of the discrete cut-distance (1.3.3) by replacing sums with integrals and permutations with measure preserving bijections. On th other hand, we are in a good position that such intuition can be formalised rigorously as for two graphs G, H and their corresponding $G-$kernel and $H-$kernel W_G, W_H, we have [127, Lemma 8.9]

$$\delta_{\boxtimes}(G,H) = \mathscr{D}_{\boxtimes}(W_G, W_H).$$

We will see in Section 2.3 that we are not that lucky in the case of limit objects of probability measures but it is possible to achieve similar but slightly weaker results.

Sampling A very elegant aspect about graph limits is that sampling from a kernel yields a finite graph which is close to the kernel in the cut-distance. This fact is often referred to as a *sampling lemma*. More precisely, given a kernel W, let $k > 1$ be an integer and sample $x_1, \dots, x_k$ uniformly and independently from $[0,1]$. Define a random graph $\boldsymbol{G} = \mathscr{G}(k,W)$ on k vertices such that edge ij is present with probability $W(x_i, x_j)$. Then we have the following sampling lemma [127, Lemma 10.16].

> **Sampling Lemma.** Let W be a kernel and $\boldsymbol{G} = \mathscr{G}(k,W)$ defined as above. Then we have with probability at least $1 - \exp(-k/(2\ln k))$ that
>
> $$\mathscr{D}_{\boxtimes}(W_{\boldsymbol{G}}, W) < \frac{22}{\sqrt{\ln k}}.$$

It will turn out in Section 2.3 that a similar fact holds for $\Omega-$laws as well.

Subgraph counts and homomorphism densities A further fairly important aspect about graph limit theory is that the space of kernels really consists of meaningful limit object for sequences of graphs. To this end, we need to briefly mention that graph convergence is actually defined as the convergence of all homomorphism densities of finite graphs into the graph sequence. Therefore, a series of graphs $(G_n)_n$ converges, by definition, if all series of homomorphism densities of finite graphs $(t(H,G_n))_H$ converge in $\mathbb{R}$. Let us define the homomorphism density. A graph homomorphism f from a graph $F = (V(F), E(F))$

into a graph $G = (V(G), E(G))$ is a function $f : V(F) \to V(G)$ that maps edges on edges, thus $ij \in V(F) \Rightarrow f(i)f(j) \in V(G)$. Now, $\hom(F, G_n)$ counts the number of homomorphisms from a given $k-$vertex graph F into G_n and we define the homomorphism density as

$$t(F, G_n) = \frac{\hom(F, G_n)}{n^k}.$$

With this notation at hand, convergence of graphs should be understand as follows. If two graph sequences produce the same convergent series of homomorphism densities, they should have much in common (as they contain equally many copies of all finite graphs) and therefore converge to the same limit object. Fortunately, it turns out that the kernels are good limit objects as we find that for any convergent sequence $(G_n)_n$ of graphs there is a kernel W such that $t(F, G_n)$ converges to $t(F, W)$ for every finite graph F [127, Theorem 11.22]. Such a kernel is called the limit of the graph sequence $(G_n)_n$ and we say that $(G_n)_n$ converges to W. Now it could in principle happen that multiple kernels satisfy this condition and clearly, kernels representing isomorphic graphs should definitely do so. To this end, denote by $\mathcal{W}$ the space of kernels such that we identify kernels of cut-distance zero. We have that $\mathcal{W}$ is a compact Polish space [127, Theorem 9.23]. It turns out that we are in good shape as [127, Theorem 11.22] proves that a sequence of graphs $(G_n)_n$ with a diverging number of vertices converges to a kernel W if and only if $\mathcal{D}_\boxtimes\left(W_{G_n}, W\right) \to 0$ and as $\mathcal{W}$ is a compact metric space, the uniqueness up to equivalent kernels W' - kernels that satisfy $\mathcal{D}_\boxtimes\left(W, W'\right) = 0$ - follows directly.

After discussing some aspects of the kernel representation of a graph limit, let us shortly describe a second important possibility to describe a limit object.

Aldous-Hoover representation The Aldous-Hoover representation theorem for exchangeable arrays of random variables is closely connected to the graph limit theory [6, 96, 102]. More precisely, with W being a kernel, we recall that $\mathcal{G}(k, W)$ is the random graph obtained as follows.

- Draw $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_k$ uniformly at random and independently from $[0, 1]$.
- Let each edge ij be present in $\mathcal{G}(k, W)$ with probability $W(\boldsymbol{X}_i, \boldsymbol{X}_j)$.

This can be naturally extended to an infinite random graph model $\mathcal{G}(\infty, W)$ by sampling infinitely many points $\boldsymbol{X}_i$. Of course, $\mathcal{G}(\infty, W)$ contains every finite random graph $\mathcal{G}(k, W)$ as an induced subgraph [31]. We observe that $\mathcal{G}(\infty, W)$ is an *exchangeable* random graph. Indeed, its distribution is invariant under permutation of the vertices.

Thus, the Aldous-Hoover representation theorem says that any such infinite random graph can be written as a mixture of random variables $A_{ij} = \tilde{W}_k(\boldsymbol{Y}_i, \boldsymbol{Y}_j)$ where $\{\boldsymbol{Y}_i\}_{i\in\mathbb{N}}$ is a family of mutually independent $[0, 1]-$valued random variables and $\left\{\tilde{W}_k\right\}_k$ is a family of symmetric functions $\tilde{W}_k : [0, 1]^2 \to [0, 1]$. This matrix $A = (A_{ij})_{i,j}$ is clearly an infinite random exchangeable array in $[0, 1]^{\mathbb{N}\times\mathbb{N}}$ and one would suggest that it is in 1-to-1 correspondence with the adjacency matrix of $\mathcal{G}(\infty, W)$. And this intuition is indeed correct. It is possible to proof that the set of exchangeable infinite arrays corresponding to such an infinite graph equals the set of extremal points in the space of all exchangeable random arrays. Therefore, the mixture is no real mixture but there is exactly one function $W : [0, 1]^2 \to [0, 1]$ such that $A_{ij} = W(\boldsymbol{Y}_i, \boldsymbol{Y}_j)$ for all i, j. This is, of course, exactly the kernel W.

After having learned a spark of graph limit theory, let us use it to tackle one of the most influential concepts in graph theory, the graph regularity.

1.3.1.3. Regularity of graphs, graph limits and probability measures

Since the first occurrence of a regularity lemma for graphs in 1975, graph regularity and its generalisation have attracted a lot of attention. We refer to two very detailed overview articles by Rödl and Schacht [151] and Komlos et al. [112] describing and analysing various variants of graph regularity and their applications. In this thesis, we will only sketch two types of regularity, one which is just called *regularity* as defined by Szemerédi [160] and a much weaker version called *weak regularity* by Frieze and Kannan

[84]. Besides belonging clearly to the most studied types of regularity, we will see that those concepts are closely related to the cut-distance for probability distributions.

Throughout this section we suppose that $G = (V, E)$ is a graph on n vertices with at least εn^2 edges. The regularity lemma says, intuitively speaking, that the vertex set of each such graph can be partitioned into finitely many classes such that the edges between (almost all) those classes look random. Let us describe this in more detail.

For two subsets $X, Y \subset V$ of the vertex set we denote by $E(X, Y)$ the set of edges with one endpoint in X and one endpoint in Y. Then the edge density between X and Y is defined as

$$d(X, Y) = \frac{|E(X, Y)|}{|X||Y|}.$$

If the edges between X and Y were randomly chosen, we would expect that the edge density between not too small subsets of X and Y roughly equals the overall edge density. Thus, let us call the pair (X, Y) $\varepsilon-$regular if for all $X' \subset X$, $Y' \subset Y$ with $\left|X'\right| \geq \varepsilon |X|$ and $\left|Y'\right| \geq \varepsilon |Y|$ we have

$$\left|d(X, Y) - d(X', Y')\right| \leq \varepsilon.$$

If we now have a partition $\boldsymbol{S} = (S_0, S_1, \ldots, S_\ell)$ of the vertex set V, we say that $\boldsymbol{S}$ is $\varepsilon-$regular if

- the *exceptional set* S_0 satisfies $|S_0| \leq \varepsilon n$,
- for all $1 \leq i < j \leq \ell$ we have that the pair (S_i, S_j) is $\varepsilon-$regular.

Now, the famous regularity lemma of Szemerédi [160] guarantees that every large enough dense graph has such a partition of its vertex set.

> **Regularity Lemma**. For all $\varepsilon > 0$ and every $t \in \mathbb{N}$ there exists an integer $T = T(\varepsilon, t)$ such that each graph G on at least T vertices has an $\varepsilon-$regular partition $\boldsymbol{S} = (S_0, \ldots, S_\ell)$ of its vertex set where $t \leq \ell \leq T$.

A specific feature of this theorem is that T is independent of the graph G and its size. Nevertheless, it turns out that T is lower bounded by a tower of 2s of height proportional to $\ln(1/\varepsilon)$ [92]. One might ask, do we get smaller partitions if we only want to have the property of being regular on average? Indeed, Frieze and Kannan [84] answer this question positively.

Observe that given a regular partition $\boldsymbol{S} = (S_0, \ldots, S_\ell)$ we find that the number of edges between two disjoint subsets of vertices A, B is within $\pm\varepsilon n^2$ of

$$\sum_{i,j=1}^{\ell} d(S_i, S_j) |A \cap V_i| \left|B \cap V_j\right|,$$

thus the latter expression measures somehow the deviation from being regular. Therefore, we say that a partition $\boldsymbol{S} = (S_1, \ldots, S_k)$ of the vertex set of a graph $G = (V, E)$ is *weakly $\varepsilon-$regular* if we have for all disjoint $A, B \subset V$

$$\left| |E(A, B)| - \sum_{i,j=1}^{\ell} d(S_i, S_j) |A \cap S_i| \left|B \cap S_j\right| \right| < \varepsilon. \tag{1.3.5}$$

Now, the weak regularity lemma [84] guarantees the existence of a weakly regular partition for every graph.

> **Weak Regularity Lemma**. For all $\varepsilon > 0$ and every graph G there is a weakly $\varepsilon-$regular partition of its vertex set into k sets such that $k \leq \exp\left(O\left(\varepsilon^{-2}\right)\right)$.

Thus, the weak regularity lemma provides a partition which is on average regular but consists of considerably less parts. Interestingly, if G_S is the weighted graph on the same vertices as G with edge-weight

$d(S_i, S_j)$ for edge uv with $u \in S_i, v \in S_j$, we have that (1.3.5) implies

$$\delta_{\boxtimes}(G, G_S) < 2\varepsilon. \tag{1.3.6}$$

Therefore, the cut-distance is closely related to the concept of (weak) regularity. Of course, a similar result does hold for kernels as well. If W is a kernel and $\boldsymbol{S} = (S_1, \dots, S_\ell)$ is a partition of $[0,1]$ into measurable sets, we define following [127, Section 9.2.1][2]

$$W_S(x,y) = \frac{1}{\lambda(S_i)\lambda(S_j)} \int_{S_i \times S_j} W(x,y) \mathrm{d}y \mathrm{d}x \qquad \text{for } (x \in S_i, y \in S_j). \tag{1.3.7}$$

Thus, W_S is obtained by averaging over each *step* $S_i \times S_j$. Now the weak regularity lemma states that given a kernel W we find a partition $\boldsymbol{S}$ of the unit interval into $k \leq \exp\left(O\left(\varepsilon^{-2}\right)\right)$ sets such that $\mathscr{D}_{\boxtimes}(W, W_S) < \varepsilon$.

Let us now come back to discrete probability measures on Ω^n for some finite set Ω. More precisely, we look at their continuous embeddings as Ω-laws on Σ_Ω. It is possible to define a similar concept of regularity for those objects inspired by the concept of regularity in graph theory [21, 42]. In order to do so, we need to introduce some notation.

For a set $X \subset [0,1)$ and a configuration $\sigma \in \Sigma_\Omega$ as well as a spin $\omega \in \Omega$ define

$$\sigma[\omega \mid X] = \int_X \sigma_x(\omega) \mathrm{d}x.$$

Thus, $\sigma[\cdot \mid X] : [0,1) \to \mathscr{P}(\Omega)$ is a probability distribution on Ω and more precisely, it can be seen as a continuous valued analogue of the empirical distribution of σ on X.

If we now have a partition $\boldsymbol{V} = (V_0, V_1, \dots, V_\ell)$ of $[0,1)$, and a partition $\boldsymbol{S} = (S_0, S_1, \dots, S_k)$ of Σ_Ω we say that μ is $\varepsilon-$regular with respect to $(\boldsymbol{V}, \boldsymbol{S})$, if

(i) the non-exceptional sets $V_1, \dots, V_\ell$ and $S_1, \dots, S_k$ satisfy

$$\lambda(V_i)\,\mu\left(S_j\right) > 0 \qquad \text{and} \qquad \sum_{i=1}^{\ell} \sum_{j=1}^{k} \lambda(V_i)\mu(S_j) \geq 1 - \varepsilon,$$

(ii) for all $1 \leq i \leq \ell$, $1 \leq j \leq k$ we have

$$\max_{\sigma, \sigma' \in S_i} \left|\left|\sigma[\cdot|V_j] - \sigma'[\cdot|V_j]\right|\right|_1 < \varepsilon,$$

(iii) for all $1 \leq i \leq \ell$ and $1 \leq j \leq k$ we have for $U \subset V_i$ with $\lambda(U) \geq \varepsilon\lambda(V_i)$ and $T \subset S_j$ with $\mu(T) \geq \varepsilon\mu(S_j)$ that

$$\left|\left|\langle \boldsymbol{\sigma}[\cdot \mid U]\rangle_{\mu[\cdot|T]} - \langle \boldsymbol{\sigma}[\cdot \mid V_i]\rangle_{\mu[\cdot|S_j]}\right|\right|_1 < \varepsilon.$$

This definition deviates slightly from the one in [42] with respect to the exceptional sets but is clearly equivalent. As stated in this publication, (ii) guarantees that the averages $\sigma[\cdot \mid V_i]$ and $\sigma'[\cdot \mid V_i]$ over V_i of any two configurations from one cluster S_j are close. More importantly, (iii) requires that the average $\langle \boldsymbol{\sigma}[\cdot \mid U]\rangle_{\mu[\cdot|T]}$ over a large *sub-square* does roughly equal the mean over the square given by the partition $V_i \times S_j$.

As in the case of graph regularity, this can be expressed via the cut-distance. Given an $\varepsilon-$regular partition $(\boldsymbol{V}, \boldsymbol{S})$, we define

$$\sigma_x[\omega|\boldsymbol{V}] = \sum_{i=0}^{\ell} \mathbf{1}\{x \in V_i\}\, \sigma_x[\omega \mid V_i].$$

Furthermore, we let $\mu[\cdot \mid \boldsymbol{V}, \boldsymbol{S}]$ be the conditional expectation of μ with respect to this partition, thus

$$\mu[\cdot \mid \boldsymbol{V}, \boldsymbol{S}] = \sum_{j=0}^{k} \delta_{\int_{S_j} \sigma[\cdot|\boldsymbol{V}] \mathrm{d}\mu(\sigma)}.$$

[2] We denote by $\lambda(\cdot)$ the Lebesgue-measure.

Therefore, $\mu[\cdot \mid \boldsymbol{V}, \boldsymbol{S}] \in \mathscr{L}_\Omega$ is an $\Omega-$law which is supported on a discrete set of configurations $\sigma : [0,1) \to \mathscr{P}(\Omega)$ which themselves are constant on each of the partition classes $\boldsymbol{V}$. Comparing $\mu[\cdot \mid \boldsymbol{V}, \boldsymbol{S}]$ with the step-kernel W_S given via (1.3.7) we see that the two structures express very similar ideas, for graphs on the one hand and probability measures on the other hand.

Thus, it might be no surprise that we find for an Ω-law μ and an $\varepsilon-$regular partition $(\boldsymbol{V}, \boldsymbol{S})$ that [42, Proposition 2.14]

$$D_{\boxtimes}\left(\mu, \mu[\cdot \mid \boldsymbol{V}, \boldsymbol{S}]\right) < 2\varepsilon.$$

Moreover, the authors provide a regularity lemma for such $\Omega-$laws [42, Corollary 2.15].

> **Regularity Lemma for $\Omega-$laws.** For any $\varepsilon > 0$ there is a natural number $N = N(\varepsilon)$ such that for any Ω-law μ there are $\sigma_1, \dots, \sigma_N : [0,1) \to \mathscr{P}(\Omega)$ and $\omega = (\omega_1, \dots, \omega_N) \in \mathscr{P}([N])$ with $D_{\boxtimes}(\mu, \sum_{i=1}^N \omega_i \delta_{\sigma_i}) < \varepsilon$.

A similar statement is known for probability measures μ on Ω^n [21, Theorem 2.1]. Let us bring this notion of regularity together with the idea of $\varepsilon-$symmetry discussed previously. First we observe that any *refinement* of the classes $V_1, \dots, V_\ell$ only increases the cut-distance between μ and $\mu[\cdot \mid \boldsymbol{V}, \boldsymbol{S}]$ by a constant factor. Therefore, we can refine a regular partition $(\boldsymbol{V}, \boldsymbol{S})$ into singletons $V_i = \{i\}$ and observe that in this case $\mu[\cdot \mid \boldsymbol{V}, \boldsymbol{S}]$ becomes a convex combination over (on S_i conditioned) product measures on the marginals of $\mu[\cdot \mid S_i]$. Recall that for a probability measure $\nu \in \mathscr{P}(\Omega^n)$ we denote by $\bar{\nu}$ the corresponding product measure on the same marginals, therefore, on each partition class S_i we find $\mu[\cdot \mid \boldsymbol{V}, \boldsymbol{S}] = \bar{\mu}[\cdot \mid S_i]$.

A direct consequence [21, Corollary 2.2] of the regularity lemma is the following. If $\boldsymbol{S}$ is an ε^3-regular partition of (Ω^n) (we drop the singleton decomposition of $\boldsymbol{V}$ for the sake of readability from now on) w.r.t. $\mu \in \mathscr{P}(\Omega^n)$ then we have:

> **Regularity and symmetry.** For any $\varepsilon > 0$ there is $\eta = \eta(\varepsilon, \Omega) > 0$ such that for all $n \geq \eta^{-1}$ and any probability measure μ on Ω^n we have for $j = 1, \dots, k$ that $\Delta_{\boxtimes}(\mu[\cdot \mid S_j], \bar{\mu}[\cdot \mid S_j]) < O(\varepsilon)$.

Thus, finding an $\varepsilon-$regular partition of the phase space guarantees to express a probability measure as an convex combination of measures conditioned on the partition classes which look like a product measure under the cut-distance. Clearly, it would be interesting to have algorithms that construct such partitions efficiently. Here comes the flaw of all described regularity lemmas. While they guarantee the existence of regular partitions, it is not clear how to generate them. A fairly elegant way of obtaining a partition $\boldsymbol{S}$ of Ω^n for some measure $\mu \in \mathscr{P}(\Omega^n)$ is the so-called *pinning operation* introduced by Coja-Oghlan et al. [49].

1.3.1.4. Pinning

Suppose we have a measure $\mu \in \mathscr{P}(\Omega^n)$ and we want to obtain a fairly related measure which is $\varepsilon-$symmetric. In this case, the *pinning lemma* [49, Lemma 3.5] provides a very simple way of achieving this goal.

> **Pinning Lemma.** For any $\varepsilon > 0$ there is a natural number $T > 0$ such that for every $n > T$ and every $\mu \in \mathscr{P}(\Omega^n)$ the following holds. Construct a (random) probability measure $\boldsymbol{\nu} = \boldsymbol{\nu}(\mu) \in \mathscr{P}(\Omega^n)$ as follows.
>
> - Sample $\tilde{\boldsymbol{\sigma}} \sim \mu$.
> - Choose independently an integer $\boldsymbol{\theta} \in [1, T)$ uniformly at random.
> - Create a random subset $U \subset [n]$ by including each coordinate with probability $\boldsymbol{\theta}/n$ independently.
> - Define
>
> $$\boldsymbol{\nu}(\sigma) = \mu(\sigma) \frac{\mathbf{1}\{\forall i \in U : \sigma_i = \tilde{\boldsymbol{\sigma}}_i\}}{\mu(\{\tau : \forall i \in U : \tau_i = \tilde{\boldsymbol{\sigma}}_i\})}.$$
>
> Then we have $\Delta_{\boxtimes}(\boldsymbol{\nu}, \bar{\boldsymbol{\nu}}) < O\left(\varepsilon^{1/3}\right)$ with probability at least $1 - \varepsilon$.

Therefore, we draw just a single sample from μ and *pin* a relatively small number of coordinates to their spins under this sample. This reweighed measure is now likely to be $\varepsilon-$symmetric. Fairly related versions of such lemmas for probability measures were previously obtained by Montanari [137, Lemma 3.1] and Raghavendra and Tan [148].

It actually turns out that the pinning operation does not only yield such a reweighed measure but that we can actually obtain a partition of the phase space and a corresponding family of reweighed measures, when we just define the partition as given by all $|\Omega|^{|U|}$ possibilities of spins those variables in U can take. While this operation is clearly a mighty tool for producing regular partitions, it was not known prior to this thesis's contributions whether and how similar approaches might work for Ω-laws. We will discuss this in Section 2.3.

Let us for the sake of completeness mention that similar approaches of obtaining regular partitions for graphs have been studied as well by (non-exclusively) Tao [163] and Fischer, Matsliah and Shapira [79] by choosing partitions according to the adjacency of randomly selected vertices. Up to our knowledge, a complete understanding on how the procedure of sampling and reweighing can be explicitly applied in this context is not yet gained.

In this section we discussed the cut-distance and its connection to regularity for probability measures and graphs. We saw that regularity implies that (dense) large graphs cannot *look arbitrarily wild* but do contain a lot of structure. Structural results of (sparse) graphs will be the topic of the next section which introduces the concept of *random perturbation* of graphs.

1.4. Perturbing sparse graphs: when randomness meets determinism

The origin of randomly perturbing deterministic structures can be traced back to a contribution of Spielman and Teng [159] who introduced the *smoothed analysis of algorithms*. The key idea is fairly simple. While it is known for many algorithms that their worst-case running time exceeds the average case running time significantly, it can be observed in real world applications that this worst-case does not usually occur. A prime example might be the simplex algorithm [61] whose worst-case complexity is exponential but nevertheless, the method is used quite frequently in applications. Of course, one could analyse the *average case* running time but this might not give enough performance guarantees in production systems, as it is not unlikely to observe an input deviating from the average case. The smoothed analysis of algorithms overcomes this flaw by analysing a worst-case input on whom random changes have been applied. If the number of changes is fairly high, the instance clearly becomes average case by definition but if the number of manipulations is small, the modified input is fairly close to a worst-case instance which is reasonable to be observed in real applications as hitting the absolute worst-case is very unlikely.

Subsequently, this idea of randomly perturbing deterministic structures became of interest in the study of random graphs. For instance, one of the most famous results of extremal combinatorics might be Dirac's theorem [65] on Hamilton cycles. It states that whenever a graph on n vertices has minimum degree at least $\lceil n/2 \rceil$, this graph contains a Hamilton cycle, thus a closed loop through the graph visiting each vertex exactly once. This result is optimal in the sense that there are graphs with minimum degree $\lfloor n/2 \rfloor$ which do not contain such a cycle, i.e. let $n = 2m-1$ and take two copies of the complete graph on m vertices K_m which share exactly one common vertex. The latter graph has clearly no Hamilton cycle but has minimum degree $\lfloor n/2 \rfloor$. But it is of course very unlikely to observe such an extreme graph in a real world network, thus many graphs with smaller minimum degree contain a Hamilton cycle (clearly, the necessary minimum degree is 2 achieved by the cycle on n vertices C_n). Let us look at the average case, which is the binomial random graph. It is well understood that $\mathcal{G}(n,p)$ undergoes a strict phase transition with respect to containing a Hamilton cycle at $p = n^{-1} \ln n$ [116, 117, 147]. What happens if we combine the probabilistic and deterministic objects? Denote by $G_\alpha = (V,E)$ an arbitrary (probably adversely chosen) graph on $V = [n]$ with minimum degree αn and denote furthermore by $F = G_\alpha \cup \mathcal{G}(n,p)$ the union of this graph and an instance of the binomial random graph with edge probability p. Here we define the union as follows: For each pair of vertices $i, j \in \binom{V}{2}$, we add an edge of $\mathcal{G}(n,p)$ independently

of everything else with probability p. When does F contain a Hamilton cycle with high probability?

Clearly, if $\alpha \geq 1/2$, the existence follows solely from Dirac's theorem applied to G_α. If on the other hand $p \geq (1+\varepsilon)\ln n/n$, we find the cycle inside of the edges of $\mathscr{G}(n,p)$ with high probability. But what happens in between?

This model of randomly perturbed graphs with given minimum degree was introduced by Bohman, Frieze and Martin [28] for $\alpha = \Theta(1)$, thus for dense graphs. In the aforementioned contribution the authors explicitly find the trade-off between α and p. More precisely, they proved that for every constant $0 < \alpha < 1/2$ there are graphs G_α such that $G_\alpha \cup \boldsymbol{G}(n,p)$ does not contain a Hamilton cycle with high probability if $p = o(1/n)$. On the other hand, if $p = \omega(1/n)$, any such union of graphs contains a Hamilton cycle with high probability. It is important to observe that $p = 1/n$ is the phase transition point in (solely) $\mathscr{G}(n,p)$ of containing a cycle on all but εn vertices. Thus the transition point for the spanning structure in the perturbed model equals (in this case) the transition point of the existence of an almost spanning structure in the random graph which is due to the existence of isolated vertices below $p \sim \ln n/n$.

Of course, there is not only interest for Hamilton cycles but also for various different spanning structures. On the side of finding Dirac-like theorems for the existence of specific spanning structures in G_α solely, there are for instance results for spanning trees [115], factors [94] as well as powers of Hamilton cycles [113, 114]. Finally, there are even fairly generic results for the existence of a copy of any bounded degree graph in G_α by Böttcher, Schacht and Taraz [36]. And clearly, the existence of all of those structures is known to undergo phase transitions in the random graph $\mathscr{G}(n,p)$. To briefly name a few important contributions, there are results on the existence of matchings [72], spanning trees [118, 139], factors [107] and powers of Hamilton cycles [125, 141]. Finally, there are also generic results on the phase transitions with respect to general bounded degree graphs [14, 76, 77, 150]. We refer to a recent overview article of Böttcher [33] for a more detailed presentation.

It is not very surprising that since the first discussion of the existence of Hamilton cycles in randomly perturbed graphs various contributions obtained results with respect to the aforementioned spanning structures. Just to name a few, there are results on spanning trees [34, 119], factors [20] as well as powers of Hamilton cycles [26]. Ultimately, there are recent results on the existence of general bounded degree graphs in the perturbed model by Böttcher et al. [35]. Interestingly, in most of the results, the obtained phase transition for p is a multiplicative factor of order $\ln n$ smaller than in $\mathscr{G}(n,p)$.

The knowledge of things becomes completely different if one allows $\alpha = o(1)$, thus one has a deterministic sparse graph G_α and needs more edges from the random graph. We will investigate the existence of perfect matchings, Hamilton cycles and bounded degree trees in this model of sparse perturbed graphs in Section 2.4.

2. Results

This chapter summarises the main results obtained in this thesis's contributions. Those results will be presented and very short proof sketches will be given showing the most important steps in order to achieve those results. We emphasise that those sketches make simplifying assumptions and leave out all technical details because they are just meant to grasp the main idea of how to prove a result. For complete and rigorous proofs we refer the reader to the contributions in the appendix.

We start by discussing results with respect to the group testing problem. Subsequently, we will answer the question of how many satisfying assignments a random 2-SAT formula has and state results in context with a limit theory for discrete probability measures and the role of the cut-distance. Finally, we discuss the existence of spanning structures in randomly perturbed sparse graphs.

2.1. Group Testing

As already discussed in the introduction, all our results are within the framework of Bayes optimal sublinear hypergeometric probabilistic group testing, thus we try to achieve inference with high probability and the ground-truth $\boldsymbol{\sigma}$ is supposed to be chosen uniformly at random from all possible configurations in $\{0,1\}^n$ of Hamming weight $k \sim n^\theta$ for some $\theta \in (0,1)$. The results of this section were obtained in the following contributions which can be found in the appendix:

- *Information-Theoretic and algorithmic thresholds for group testing* [41],
- *Optimal group testing* [46],
- *Near optimal sparsity-constrained group testing: improved bounds and algorithms* [88].

We start by discussing results with respect to non-adaptive group testing schemes. A summary of the obtained results can be found at the end of the section.

2.1.1. Non-adaptive Group Testing

Almost all obtained results make extensive use of the occurrence of different types of individuals in a group testing instance. Following [88], we will shortly describe the combinatorial meaning of those types.

Throughout the section we suppose that we have individuals $V = \{x_1, \dots, x_n\}$ and tests $F = \{a_1, \dots, a_m\}$ and that each individual's infection status is given by the underlying ground-truth $\boldsymbol{\sigma}$ and all test-results are given by $\hat{\boldsymbol{\sigma}}$.

2.1.1.1. Combinatorial properties of individuals

Suppose that some (non-adaptive) pooling scheme is given through the factor graph $\mathcal{G} = (V \cup F, E)$. We abbreviate the set of uninfected individuals to V_0 and the set of infected individuals to V_1, thus

$$V_0(\mathcal{G}) = \{x \in V(\mathcal{G}) : \boldsymbol{\sigma}_x = 0\} \qquad \text{and} \qquad V_1(\mathcal{G}) = \{x \in V(\mathcal{G}) : \boldsymbol{\sigma}_x = 1\}.$$

Those uninfected individuals appearing in a negative test can be classified immediately and play therefore a special role. We define this set of individuals as V_{0-}, formally

$$V_{0-}(\mathcal{G}) = \{x \in V_0(\mathcal{G}) : \exists a \in \partial_{\mathcal{G}} x : \hat{\boldsymbol{\sigma}}_a = 0\}.$$

Furthermore, with respect to the `DD` algorithm, it intuitively makes sense to denote the set of infected individuals which appear in at least one test with only elements of V_{0-}. Thus, upon classifying the latter

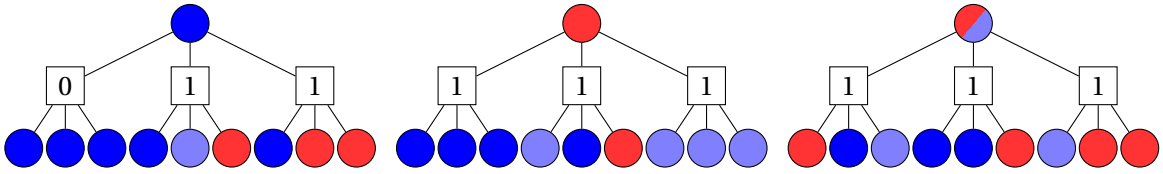

Figure 2.1.: The local structure given by the underlying factor graph in the non-adaptive group testing problem modified after [88]. Blue individuals are uninfected under the ground-truth $\boldsymbol{\sigma}$ while red individuals are infected. More precisely, we suppose that light blue individuals belong to V_{0+} and dark blue individuals to V_{0-}. From left to right, the upper individuals are elements of V_{0-}, V_{1--} and V_+ respectively.

individuals, those infected individuals are easy to identify. More precisely, we define

$$V_{1--}(\mathcal{G}) = \{x \in V_1(\mathcal{G}) : \exists a \in \partial_{\mathcal{G}} x : \partial_{\mathcal{G}} a \setminus \{x\} \subset V_{0-}(\mathcal{G})\}.$$

Moreover, there might be *totally disguised* individuals. Following Aldridge, Johnson and Scarlett [10] and Mézard, Tarzia and Toninelli [133], we say that an individual x is *disguised* in a test a if there is an infected individual $y \in \partial a \setminus \{x\}$. A totally disguised individual is disguised in all of its tests. Clearly, the infection status of those individuals cannot be inferred directly, but using the prior, it is possible to declare any of those individuals as uninfected as long as there are not too many totally disguised individuals [8, 41]. Formally, we let

$$V_+(\mathcal{G}) = \{x \in V(\mathcal{G}) : \forall a \in \partial_{\mathcal{G}} x : (\partial_{\mathcal{G}} a \setminus \{x\}) \cap V_1(\mathcal{G}) \neq \emptyset\}.$$

For the sake of completeness, we furthermore define the set of totally disguised infected and uninfected individuals, thus

$$V_{0+}(\mathcal{G}) = V_+(\mathcal{G}) \cap V_0(\mathcal{G}) \qquad \text{and} \qquad V_{1+}(\mathcal{G}) = V_+(\mathcal{G}) \cap V_1(\mathcal{G}).$$

If it is clear from the context which pooling scheme $\mathcal{G}$ is the matter of discussion, we will write $V_{0+} = V_{0+}(\mathcal{G})$ and equivalently abbreviate the other sets. A graphical visualisation of those types of individuals is given in Figure 2.1.

The sizes of the sets V_{1+}, V_{0+}, V_{1--} have direct impact on the algorithmic and information-theoretic feasibility of a group testing instance. For a configuration $\tau \in \{0,1\}^n$ we denote by $\hat{\tau} = \hat{\tau}(\mathcal{G})$ the corresponding test-results on a pooling scheme $\mathcal{G}$. Now we define

$$S_k = S_k(\mathcal{G}, \boldsymbol{\sigma}) = \left\{\tau \in \{0,1\}^n : ||\tau||_1 = k \quad \text{and} \quad \hat{\tau} = \hat{\boldsymbol{\sigma}}\right\} \qquad \text{as well as} \qquad Z_k = Z_k(\mathcal{G}, \boldsymbol{\sigma}) = |S_k|.$$

This notation enables us to find the following assertion which holds as the infection status of totally disguised individuals can be swapped arbitrarily due to the supposed Bayes optimality ([41, Corollary 2.2] and [88, Claim 2.3]).

Lemma 2.1.1. *Let $Z_k(\mathcal{G}, \boldsymbol{\sigma})$ be defined as above, then the following holds.*

- *If $Z_k(\mathcal{G}, \boldsymbol{\sigma}) = 1$, there exists a (not necessarily efficient) algorithm which infers $\boldsymbol{\sigma}$ from $(\mathcal{G}, \hat{\boldsymbol{\sigma}})$ with high probability.*
- *If $Z_k(\mathcal{G}, \boldsymbol{\sigma}) = \ell$, any algorithm (efficient or not) fails at inference of $\boldsymbol{\sigma}$ from $(\mathcal{G}, \hat{\boldsymbol{\sigma}})$ with probability at least $1 - \ell^{-1}$.*
- *For any test design $\mathcal{G}$, we have $Z_k(\mathcal{G}, \boldsymbol{\sigma}) \geq |V_{1+}(\mathcal{G}) \times V_{0+}(\mathcal{G})|$.*

A similar statement follows for the `DD` algorithm but here we need to take the set V_{1--} into account because it consists of exactly those individuals which will be misclassified by `DD`.

Lemma 2.1.2 (Corollary 2.4 of [88])**.** *The `DD` algorithm recovers $\boldsymbol{\sigma}$ from $(\mathcal{G}, \hat{\boldsymbol{\sigma}})$ if and only if $V_1(\mathcal{G}) = V_{1--}(\mathcal{G})$.*

Let us subsequently state achieved information-theoretic bounds for non-adaptive hypergeometric probabilistic group testing.

2.1.1.2. Information-theoretical results

Recall m_{inf} and $m_{\text{rand-reg}}$ from (1.2.5) and (1.2.8) as

$$m_{\text{inf}} = \frac{1}{\ln 2} k \ln \frac{n}{k} \qquad \text{and} \qquad m_{\text{rand-reg}} = \max\left\{ \frac{\theta}{(1-\theta)\ln^2 2}, \frac{1}{\ln 2} \right\} k \ln \frac{n}{k}.$$

As discussed earlier, the random regular model was known to fail with positive probability below $m_{\text{rand-reg}}$ due to [9]. We strengthen this converse statement and establish an achievability statement at the same value, thus obtaining a strict phase transition in the random regular model. With a slight misuse of notation we refer to a model as the *random (almost) regular model* if each individual chooses Δ tests uniformly at random with or without replacement. While the technical delicacies of the proofs change depending on the exact model formulation the results themselves stay the same.

Theorem 2.1.3 (Theorem 1.1 of [41])**.** *Let $\mathcal{G}$ be the random almost regular pooling scheme with n individuals and m tests, $\varepsilon > 0$ and $k \sim n^\theta$. Then the following holds.*

- *If $m > (1+\varepsilon) m_{\text{rand-reg}}$, there is an (exponential time) algorithm inferring $\boldsymbol{\sigma}$ from $(\mathcal{G}, \hat{\boldsymbol{\sigma}})$ with high probability.*
- *If $m < (1-\varepsilon) m_{\text{rand-reg}}$, no algorithm (efficient or not) outputs $\boldsymbol{\sigma}$ given $(\mathcal{G}, \hat{\boldsymbol{\sigma}})$ with a non-vanishing probability.*

Proof sketch of Theorem 2.1.3. We need to establish two directions of the theorem, let us start with the achievability result. We do this by employing Markov's inequality and an idea from statistical physics. More precisely, we denote by $Z_{k,\ell}$ the number of configurations $\tau \neq \boldsymbol{\sigma}$ satisfying all test-results that have overlap ℓ with $\boldsymbol{\sigma}$ where the overlap is defined as the number of individuals which are infected under τ and $\boldsymbol{\sigma}$. If we can show that the expected number of such individuals (for the sum over $\ell = 0 \dots k-1$) is $o(1)$, Markov's inequality shows that there is, w.h.p., only one satisfying assignment (namely $\boldsymbol{\sigma}$) which can be found by exhaustive search.

Thus, let us bound the expected number of such configurations. First, we suppose $\ell < (1 - 1/\ln n)k$ and calculate the expected number of individuals quite directly as

$$\mathbb{E}[Z_{k,\ell}(\mathcal{G}, \hat{\boldsymbol{\sigma}})] \leq O(1) \binom{k}{\ell} \binom{n-k}{k-\ell} \left(1 - 2(1-k/n)^{\bar{\Gamma}} + 2(1-2k/n+\ell/n)^{\bar{\Gamma}}\right)^m, \tag{2.1.1}$$

where $\bar{\Gamma} = n\Delta/m$ is the average test size. The combinatorial meaning is immediate. While the product of the two binomial coefficients counts the number of possible assignments τ that have overlap ℓ with $\boldsymbol{\sigma}$, the last factor is the probability that an average test renders the same result under $\boldsymbol{\sigma}$ and τ. Of course, we made many simplifying assumptions as, for instance, we supposed that all tests are independent and the single test degrees are sufficiently concentrated. But it turns out that this can be turned into a rigorous argument. Now we find – as long as $\ell < (1 - 1/\ln n)k$ – that an easy calculation provides

$$\sum_{\ell=0}^{(1-1/\ln n)k} \mathbb{E}[Z_{k,\ell}(\mathcal{G}, \hat{\boldsymbol{\sigma}})] = o(1)$$

for the choice of $m = (1+\varepsilon) m_{\text{rand-reg}}$.

Unfortunately, this argument fails for very large overlaps as the r.h.s. of (2.1.1) gets too large, thus the expectation overshoots the value of the random variable dramatically. This is a kind of a *lottery effect* and is well known in the random CSP literature as for big overlap values rare but very solution-rich clusters dominate the expectation [3]. But fortunately, the random CSP literature enables us to cope with such phenomena [1]. In short, we only need to show that the underlying random graph simply does not

allow solutions having an overlap close to k as $\boldsymbol{\sigma}$ is locally rigid. More precisely, we can proof that above $m_{\text{rand-reg}}$, each individual is part of $\Theta(\Delta) = \Theta(\ln n)$ tests such that all other contained individuals belong to V_{0-}. Therefore, upon changing the infection status of one individual from $1 \to 0$, we directly need to change the status of $\sim \ln n$ different individuals from $0 \to 1$ to compensate for those tests. But now, the same applies for those individuals, thus we have to change the status of another $\ln n$ individuals from $1 \to 0$. This argument goes on for a couple of rounds until we proved that ℓ needs to be smaller than $(1 - 1/\ln n)k$.

We observe at this point that the both expressions in the $\max(\cdot,\cdot)$ of $m_{\text{rand-reg}}$ exactly account for those two arguments. While the $\ln^{-1} 2$ from the universal counting bound suffices to guarantee that the expected number of alternative satisfying configurations is small, the non-adaptivity part accounts for the local rigidity. This perfectly fits into the previous discussion.

Let us subsequently sketch how to achieve the converse statement, namely that any inference algorithm fails below $m_{\text{rand-reg}}$ on the random regular model. As $m_{\text{rand-reg}}$ coincides with m_{inf} for $\theta < \frac{\ln 2}{1+\ln 2}$, we only need to consider larger values of θ. The proof idea is fairly simple, as by Lemma 2.1.1, we just need to calculate that we find $\omega(1)$ totally disguised infected and uninfected individuals with high probability. For this proof sketch, we suppose that all tests are stochastically independent but we remark that the actual proof is technically challenging in order to cope with the delicate dependencies in the random regular model. I.e., it is possible to describe the number of infected individuals in each test by a family of independent but conditioned binomial random variables. As a first step, we prove that in the random regular model on $m = ck\ln\frac{n}{k}$ tests the choice $\Delta = c\ln 2 \ln\frac{n}{k}$ is optimal. Intuitively, this maximises the entropy of the system. Having established this, the probability for an individual x to be totally disguised turns out to be roughly

$$(1-(1-k/n)^{\tilde{\Gamma}-1})^{\Delta} \sim 2^{-\Delta} \sim n^{-(1-\theta)c\ln^2 2}.$$

Thus, the expected number of individuals in V_{1+} is $n^{\theta-(1-\theta)c\ln^2 2}$ and for V_{0+} we find $\mathbb{E}[|V_{0+}|] \gg \mathbb{E}[|V_{1+}|]$. Clearly, $\mathbb{E}[|V_{1+}|]$ diverges as fast as a polynomial in n if $m = (1-\varepsilon)m_{\text{rand-reg}}$ and luckily it turns out that Chebyshev's inequality suffices to guarantee enough concentration. □

Actually, it turns out that Theorem 2.1.3 is not the strongest result we can achieve. While it answers all questions regarding information-theoretical inference in the random regular model, it might be the case that there are different pooling schemes facilitating better. This is, indeed, not the case.

Theorem 2.1.4 (Theorem 2 of [46])**.** *Let $\mathcal{G}$ be any arbitrary non-adaptive pooling scheme with n individuals and m tests, $\varepsilon > 0$ and $k \sim n^{\theta}$. If $m < (1-\varepsilon)m_{\text{rand-reg}}$, no algorithm (efficient or not) outputs $\boldsymbol{\sigma}$ given $(\mathcal{G}, \hat{\boldsymbol{\sigma}})$ with a non-vanishing probability.*

Theorem 2.1.4 implies two important facts. First of all, the random regular design is information-theoretically optimal as there cannot be any better designs for inference. Second, as

$$m_{\text{rand-reg}} > m_{\text{inf}} \qquad \text{for} \qquad \theta > \frac{\ln 2}{1+\ln 2},$$

we proved the existence of an *adaptivity-gap*. Thus, performing multiple stages of testing might decrease the number of required tests. Formerly, the question about the existence of such an adaptivity-gap has been raised prominently [10, 108]. In light of the second part of Theorem 2.1.3 and Theorem 2.1.4, we will write

$$m_{\text{non-ada}} = m_{\text{rand-reg}}$$

from now on as this marks a strict phase transition point for non-adaptive group testing. Let us now give a proof sketch of Theorem 2.1.4.

Proof sketch of Theorem 2.1.4. As before, we only need to establish the assertion for $\theta > \frac{\ln 2}{1+\ln 2}$ as the universal counting bound already proves the theorem for smaller values of θ.

This proof comes in three steps. First, we slightly change the model from a hypergeometric group testing ground-truth $\boldsymbol{\sigma}$ to an i.i.d. ground-truth $\tilde{\boldsymbol{\sigma}}$ where each entry is set to one with probability $p \sim$

$\frac{k-\sqrt{k}\ln n}{n}$. Thus, with high probability, we find a coupling of $(\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}})$ such that turning few uninfected individuals in $\tilde{\boldsymbol{\sigma}}$ to infected creates $\boldsymbol{\sigma}$. It turns out that finding totally disguised infected and uninfected individuals in a testing scheme under $\tilde{\boldsymbol{\sigma}}$ implies to find some of them also under $\boldsymbol{\sigma}$. This change of the model simplifies proving converse statements. As the underlying graph is deterministic, this enables us to create the only source of independent randomness within the setup.

Second, we show that the number of totally disguised infected and uninfected individuals below m_{inf} is large for θ really close to one. We first observe that even in an arbitrary test-design no test contains more than $n\ln n/k$ individuals as otherwise a union bound over all tests shows that these tests render a positive result anyways and could also be left out. But this already implies due to the handshaking lemma that there are very few individuals participating in more than $\ln^3 n$ many tests. Thus, the underlying testing scheme is, besides not knowing how it looks like exactly, fairly sparse if θ is large. Next, we strengthen an argument based on the FKG-inequality and the probabilistic method of Aldridge [7] and Mézard, Tarzia and Toninelli [133] to identify one individual y whose probability of being element of V_+ is not too small, say $\geq \exp\left(-m\ln^2(2)/k\right)$. Of course, this probability tends to zero (making it much more difficult to find enough such individuals compared to Aldridge's case). Next, we delete this individual and all tests and individuals in its first, second, third and fourth neighbourhood. Clearly, the left-over individuals are stochastically independent of y with respect to the property of belonging to V_+. Furthermore, upon removal of the individuals, chances of being totally disguised only increase. Let us denote by $\mathcal{G}_1$ the graph after removal of y and its described neighbourhoods. Due to the sparsity of the underlying graph we can show that $\frac{m}{n} \sim \frac{|F(\mathcal{G}_1)|}{|V(\mathcal{G}_1)|}$, thus the ratio between tests and individuals stays roughly the same. We repeat this procedure $n^{1-\delta}$ times where $\delta = \delta(\theta) > 0$ is a small constant depending on the prevalence and identify $n^{1-\delta}$ individuals which all are independently totally disguised with probability at least $q = \exp\left(-m\ln^2(2)/k\right)$. As the infection status is independent of being totally disguised, we find that the number of infected totally disguised individuals is dominated by a $\mathbf{Bin}(n^{1-\delta}, pq)$ random variable while the number of totally disguised uninfected individuals is dominated by a $\mathbf{Bin}(n^{1-\delta},(1-p)q)$ variable. Both binomial random variables turn out to have expectation $n^{\Omega(1)}$ for a suitable choice of δ if $m = (1-\varepsilon)m_{\text{non-ada}}$ and thus the Chernoff bound guarantees the existence of many infected and uninfected totally disguised individuals.

As a third step, we need to show that if we could solve a group testing instance with low prevalence given through θ with $(1-\varepsilon)m_{\text{non-ada}}(\theta)$ tests we were also able to solve an instance of larger prevalence given via θ' with $(1-\eta)m_{\text{non-ada}}(\theta')$ tests. To this end, let $\frac{\ln 2}{1+\ln 2} < \theta < \theta' < 1$ and suppose that $\mathcal{G} = (V \cup F, E)$ is a pooling graph satisfying

$$|V(\mathcal{G})| = n, \qquad \text{and} \qquad |F(\mathcal{G})| = m = (1-\varepsilon)m_{\text{non-ada}}(n,\theta).$$

We construct a pooling graph $\mathcal{G}'$ for $n' \approx n^{\theta/\theta'}$ individuals out of which $k' \sim n'^{\theta'} \sim k$ are infected, thus an instance with the same number of infected individuals but those are found within a much smaller population $n' \ll n$.

- Select n' individuals uniformly at random out of all individuals and define $\mathcal{G}'$ on those individuals but on the same tests as $\mathcal{G}$.
- Select an adjusted ground-truth $\boldsymbol{\sigma}' \in \{0,1\}^{n'}$ u.a.r. with Hamming weight k and denote by $\hat{\boldsymbol{\sigma}}'$ the corresponding test-results.

Now it is easy to proof that the probability of having multiple elements in the solution space of $(\mathcal{G},\hat{\boldsymbol{\sigma}})$ is at least as high as observing that many in the solution space of $(\mathcal{G}',\hat{\boldsymbol{\sigma}}')$. Indeed, by construction there is a coupling of $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}'$ such that all infected individuals coincide. This is true because we first choose n' individuals uniformly at random. But this implies $\hat{\boldsymbol{\sigma}} = \hat{\boldsymbol{\sigma}}'$, thus whenever a configuration τ explains $\hat{\boldsymbol{\sigma}}'$ we can construct a configuration explaining $\hat{\boldsymbol{\sigma}}$ by setting the infection status of the not in $\boldsymbol{\sigma}'$ contained individuals to 0. Finally, the assertion follows from the fact that the choice $n' \sim n^{\theta/\theta'}$ allows to calculate

$$m_{\text{non-ada}}(n',\theta') = \frac{\theta'}{(1-\theta')\ln^2 2}k'\ln\frac{n'}{k'} = \frac{\theta'}{\ln^2 2}k'\ln n' \sim \frac{\theta'\theta}{\theta'\ln^2 2}k\ln n = \frac{\theta}{(1-\theta)\ln^2 2}k\ln\frac{n}{k} = m_{\text{non-ada}}(n,\theta).$$

□

While Theorems 2.1.3 – 2.1.4 answer all questions with respect to information-theoretical phase transitions in non-adaptive group testing, we expect to require significantly more tests if we restrict the design choices. More precisely, as the random regular model has tests of size $\Theta(n/k)$ and each individual takes part in $\Theta(\ln n)$ tests, the number of tests required for inference might increase in sparsity constrained settings. As explained earlier, the work of Gandikota et al. [86] provided some information-theoretical converse statements, i.e. for the maximum test-degree $\Gamma = O\left((n/k)^{\beta}\right)$ and the maximum individual degree $\Delta = o(\ln n)$ they provide information-theoretical converse statements at $m_{\inf,\mathrm{G}}(\Gamma)$ and $m_{\inf,\mathrm{G}}(\Delta)$ respectively given in Equations (1.2.10) – (1.2.11) as

$$m_{\inf,\mathrm{G}}(\Gamma) = \frac{1}{1-\beta}\frac{n}{\Gamma} \qquad \text{and} \qquad m_{\inf,\mathrm{G}}(\Delta) = \Delta k\left(\frac{n}{k}\right)^{1/\Delta}.$$

We need to stress that the converse result with respect to $\Delta-$divisible group testing is of a weak nature, thus they prove that any testing scheme with success probability at least $1-\varepsilon$ requires at least $\Delta k\left(\frac{n}{k}\right)^{\frac{1-5\varepsilon}{\Delta}}$ tests.

Let us begin with the $\Delta-$divisible case. To this end define

$$m_{\text{non-ada}}(\Delta) = \max\left\{\exp(-1)\Delta k^{1+\frac{(1-\theta)}{\Delta\theta}}, \Delta k^{1+\frac{1}{\Delta}}\right\}. \tag{2.1.2}$$

Then we find that no testing-scheme with individual degree at most Δ can be used to infer $\boldsymbol{\sigma}$ from the test-results below $m_{\text{non-ada}}(\Delta)$. We tacitly suppose that

$$\theta/(1-\theta) < \Delta$$

as otherwise $m_{\text{non-ada}}(\Delta)$ would exceed n. Observe that we reduced the exponential dependency on the number of tests provided by [86] to a constant factor of $\exp(-1)$.

Theorem 2.1.5 (Theorem 3.1 and Theorem 3.2 of [88])**.** *Let $\Delta = \ln^{1-\delta} n$ for some $\delta \in (0,1]$ and suppose $k = n^{\theta}$ with $\theta \in (0,1)$. Let furthermore $\mathcal{G}$ be an arbitrary non-adaptive pooling scheme in which each individual gets tested at most Δ times and let $\varepsilon > 0$. Then the following holds.*

- *If $m \leq (1-\varepsilon)\Delta k^{1+1/\Delta}$, any non-adaptive pooling scheme with any algorithm (efficient or not) fails at inferring $\boldsymbol{\sigma}$ from $(\mathcal{G}, \hat{\boldsymbol{\sigma}})$ with probability at least $\max\left\{\Omega(\varepsilon^2), 1 - O\left((1-\varepsilon/2)^{\Delta}\right)\right\}$.*
- *If $m \leq (1-\varepsilon)\exp(-1)\,\Delta k^{1+(1-\theta)/(\theta\Delta)}$, any pooling scheme with any algorithm (efficient or not) fails at inferring $\boldsymbol{\sigma}$ from $(\mathcal{G}, \hat{\boldsymbol{\sigma}})$ with high probability.*

We remark that the second part of Theorem 2.1.5 does not only hold for non-adaptive pooling schemes but for adaptive ones as well. Therefore, we will prove this part later. Let us, at this point, introduce a technique which we will be using for proving the first part of the theorem. We make extensive use of the so-called 2−round exposure technique [103] which will help us to obtain further results later on.

Without going too much into detail, the technique reads as follows. Suppose we want to find a subgraph in $\mathcal{G}(n,p)$. It might be helpful to avoid stochastic dependencies by obtaining one part of the subgraph in $\mathcal{G}(n,p_1)$ (with $p_1 < p$) and afterwards expose the missing edges of $\mathcal{G}(n,p-p_1)$. It is known, that if a subgraph is found in $\mathcal{G}(n,p-p_1)\cup\mathcal{G}(n,p_1)$ with high probability, it is contained in $\mathcal{G}(n,p)$ as well with high probability. We will use this technique in order to expose the infected individuals in two rounds, more precisely, we will find a set of infected individuals with certain properties of size roughly αk and analyse their neighbourhood. Afterwards, we will infect each individual in this neighbourhood with probability $\sim (1-2\alpha)k/n$ independently which will yield totally disguised infected individuals. We will start with a proof sketch of Theorem 2.1.5.

Proof sketch of (the first part of) Theorem 2.1.5. In a first step we argue again that we may employ the i.i.d. model with $p \sim \frac{k+\sqrt{k}\ln n}{n}$. Furthermore, we prove that it suffices to find infected totally disguised individuals in the problem at hand as there will be at least as many uninfected totally disguised individuals with high probability. Before starting the actual proof, we modify the graph such that it does not

contain tests with less than constantly many individuals for a suitable constant and call the resulting pooling graph $\mathcal{G} = (V \cup F, E)$ such that $|V| = n$ and $|F| = (1-\varepsilon)\, m_{\text{non-ada}}(\Delta)$.

A key insight, again provided by the FKG-inequality, is that the probability for an individual x to belong to $V_+(\mathcal{G})$ is at least as high as in a model in which all tests are disjoint, thus

$$\mathbb{P}\left(x \in V_+(\mathcal{G})\right) \geq \prod_{a \in \partial x} \left(1 - \left(1-p\right)^{\Gamma_a - 1}\right)$$

if Γ_a denotes the degree of test a.

Now we employ the 2−round exposure technique and infect every individual with probability αp ending up with a set of infected individuals K_1 of size at least $\alpha k/2$. We observe that the expected number of totally disguised individuals can only increase if those individuals share tests. Therefore, we built an auxiliary model $\mathcal{G}'$ as follows. We include every element of K_1, and for each such individual we include as many disjoint tests of size corresponding to the tests it belongs to in G. Now we mark each individual added with those tests with probability $q = (1-2\alpha)\, p$ as infected.

Let

$$\boldsymbol{X}_u = \prod_{a \in \partial u} \left(1 - \left(1-q\right)^{\Gamma_a - 1}\right)$$

denote the probability that u is totally disguised in this auxiliary model. Jensen's inequality and the inequality of arithmetic and geometric means allow to calculate

$$\mathbb{E}\left[\boldsymbol{X}_u\right] \geq (1-\varepsilon/2)^{-\Delta}\, k^{-1}.$$

Therefore, with $\boldsymbol{X} = \sum_{u \in K_1} \boldsymbol{X}_u$ we have $\mathbb{E}[\boldsymbol{X}] \geq \alpha\,(1-\varepsilon/2)^{-\Delta}/2$ and as $\boldsymbol{X}_u$ and $\boldsymbol{X}_v$ are independent by construction a generalised Chernoff bound yields

$$\mathbb{P}\left(\boldsymbol{X} < \alpha/4\,(1-\varepsilon/2)^{-\Delta}\right) < \exp\left(-\Theta\left(\alpha\,(1-\varepsilon/2)^{-\Delta}\right)\right).$$

Finally, we observe that $\sum_{x \in K_1} \mathbb{P}\left(x \in V_{1+}(\mathcal{G})\right) \geq \boldsymbol{X}$ and Markov's inequality suffices to show that we observe at least one totally disguised infected individual with probability at least $\boldsymbol{X}/(1+\boldsymbol{X})$. Thus with the bound on $\boldsymbol{X}$ and a suitable choice of $\alpha = \alpha(\varepsilon)$ we obtain the result. □

The suspicious reader might ask why the statement (and the end of the proof sketch) require that much detailed calculation and explicit statements of probabilities. This is due to the fact that Δ might be a constant. Indeed, if Δ was diverging, the Chernoff-like bound on $\boldsymbol{X}$ would clearly suffice as $(1-\varepsilon/2)^{-\Delta} \to \infty$ but if Δ is a constant, the calculations become much harder. This is even more challenging under the $\Gamma = \Theta(1)$ restriction as rounding errors need to be taken into account.

In the Γ−sparse case, we strengthen the converse statement at $m_{\text{inf,G}}(\Gamma)$ for the special case of $\Gamma = \Theta(1)$ being a constant independent of the number of individuals. We define

$$m_{\text{non-ada}}(\Gamma) = \max\left\{\left(1 + \left\lfloor \frac{\theta}{1-\theta} \right\rfloor\right) \frac{n}{\Gamma}, 2\frac{n}{\Gamma+1}\right\}. \tag{2.1.3}$$

Again, we tacitly suppose that $\Gamma > \left(1 + \left\lfloor \frac{\theta}{1-\theta} \right\rfloor\right)$ as otherwise individual testing would be superior. Then we find that no non-adaptive pooling scheme can infer the ground-truth using less than $m_{\text{non-ada}}(\Gamma)$ tests.

Theorem 2.1.6 (Theorem 4.1 of [88])**.** *Let $\mathcal{G}$ be any non-adaptive pooling scheme with tests of size at most $\Gamma = \Theta(1)$. Suppose $\mathcal{G}$ contains at most $m = (1-\varepsilon)\, m_{\text{non-ada}}(\Gamma)$ tests for some $\varepsilon > 0$. Then any inference algorithm fails at recovering $\boldsymbol{\sigma}$ from $(\mathcal{G}, \hat{\boldsymbol{\sigma}})$ with high probability if $\theta/(1-\theta)$ is no integer and with probability $\Omega(1)$ if $\theta/(1-\theta)$ is an integer.*

Interestingly, for very few density levels θ, the phase transition could only be proven to be coarse rather than strict. This arises from technical reasons as, for instance, counting the number of nodes with degree at most $\theta/(1-\theta)$ which is tight in the integer-case.

Proof sketch of Theorem 2.1.6. We start by defining

$$d^+ = 1 + \left\lfloor \frac{\theta}{1-\theta} \right\rfloor \qquad \text{and} \qquad d^- = \left\lfloor \frac{\theta}{1-\theta} \right\rfloor$$

and need to distinguish between low prevalence and high prevalence.

Let us begin to sketch the case $1/2 \leq \theta < 1$. Again, we employ the i.i.d. model with $p = \frac{k-\sqrt{k}\ln n}{n}$. A first observation is that whenever $\mathscr{G}'$ is a pooling scheme with maximum test degree Γ on

$$m = (1-2\varepsilon)d^+ n/\Gamma = (1-2\varepsilon)m_{\text{non-ada}}(\Gamma)$$

tests achieving inference, there is a pooling scheme $\mathscr{G}$ on the same set of individuals with

$$m' = (1-\varepsilon)d^+ n/\Gamma$$

tests which achieves inference as well and which has the additional property that each individual gets tested at most $\Delta = \Theta(1)$ times. This follows immediately from a plain counting argument as there can only be n/C individuals of degree $\geq C'$ as m is linear in n. Thus, testing each individual of degree $> C'$ individually causes n/C additional tests. Clearly, inference in $\mathscr{G}'$ implies inference in $\mathscr{G}$ and we get the claim by choosing $C = \Gamma/(\varepsilon d^+)$. By a similar token, one can achieve that we might suppose that the minimum test-degree is at least 2 (if $d^+ \geq 3$).

A second observation is that $m = (1-\varepsilon)d^+ n/\Gamma$ implies that there are at least αn individuals of degree at most d^- by the handshaking lemma. Now we claim that there are also βn individuals of degree at most d^- and distance at least 6 in $\mathscr{G}$ which follows from simple counting as all variables have bounded degree. Let B denote a set of individuals satisfying those two properties.

Then clearly, the property of being disguised is independent for $x, x' \in B$ and the probability turns out to be, as before, at least

$$\mathbb{P}(x \in V_+) \geq \prod_{a \in \partial x} \left(1 - (1-p)^{\Gamma_a - 1}\right) = \Theta\left(p^{d^-}\right).$$

By the independence we directly find that a binomial random variable $\mathbf{Bin}\left(\beta n, \Theta\left(p^{d^-+1}\right)\right)$ dominates $|V_{1+}(\mathscr{G})|$. Therefore, the expected number of totally disguised infected individuals is $\Theta\left(n^{\theta-(1-\theta)d^-}\right)$.

The failure with high probability if $\theta/(1-\theta)$ is no integer and with positive probability if it is follows directly from the Chernoff bound.

Hence, we are only left to prove the result for $0 < \theta < 1/2$. The first fact which follows from double counting the edges of the pooling graph $\mathscr{G} = (V, E)$ is that there are at least εn individuals of degree one. A second observation is that there cannot be many tests containing two (or more) individuals of degree one. Assume there were $n/\sqrt{k}$ such tests, then the Chernoff bound guarantees that we have $\sqrt{k}\ln n$ tests containing two individuals of degree one out of which one is infected. Thus any inference algorithm has to guess the infection status of those individuals and therefore the chance of correct inference is $2^{-\omega(1)} = o(1)$. But those two simple observations already suffice as they imply directly that

$$(2-\varepsilon)n/\Gamma = m \geq \varepsilon n - o(n)$$

needs to hold. Solving for ε implies

$$m \geq 2\frac{n}{\Gamma+1} - o(n).$$

Therefore, the theorem follows from combining the arguments for small and large θ. □

Let us briefly mention that those sparsity constrained results do explicitly not converge to $m_{\text{rand-reg}}$ for $\Delta \to \ln n$ and $\Gamma \to n/k$ what might at the first glance be a surprise. But actually, it is not very surprising. First, the bounds are given only with respect to the first order and lower order terms might get relevant in the limit and second, the proofs extensively make use of the sparsity, thus many partial results do not carry over to the unconstrained setting.

We observed the splitting of the information-theoretic bound into two parts (thus a maximum over

two terms) in the unrestricted group testing in $m_{\text{non-ada}}$ as well as in the $\Delta-$sparse case in $m_{\text{non-ada}}(\Delta)$. This is actually not very surprising. The first part comes in both cases from an information-theoretic argument (for instance, the counting bound) which applies for adaptive as well as non-adaptive pooling schemes, as the amount of information provided by the test-results must exceed certain bounds. On the other hand, the non-adaptivity strikes with respect to the different types of individuals. If the set of totally disguised infected individuals is non-empty, non-adaptive pooling schemes are doomed to fail while adaptive pooling schemes might classify those individuals in subsequent stages of tests. This is why the second term in the maximum actually corresponds to the phase transition point from which on there are totally disguised infected individuals in a pooling scheme.

After having discussed the information-theoretic side, let us now present algorithmic results in the aforementioned models.

2.1.1.3. Algorithms

We will first discuss two major results obtained in [41] and [46] with respect to unrestricted non-adaptive group testing algorithms. The first result shows that DD and SCOMP fail at exactly the same threshold in the random regular model, thus we obtain a fitting converse statement to the achievability result of DD obtained by [108] and settle a strict phase transition. Furthermore, we reject the conjecture of Aldridge, Baldassini and Johnson [8] that SCOMP could outperform DD asymptotically. Recall

$$m_{\text{DD}} = \max\left\{\frac{\theta}{(1-\theta)\ln^2 2}, \frac{1}{\ln^2 2}\right\} k \ln \frac{n}{k}$$

from (1.2.9). Then our result reads as follows.

Theorem 2.1.7 (Theorem 1.2 of [41])**.** *Let $\varepsilon > 0$ be a constant. Then, if $\mathcal{G}$ is an instance of the random regular model with $m \leq (1-\varepsilon) m_{DD}$ tests, both SCOMP and DD fail at inferring $\boldsymbol{\sigma}$ from $(\mathcal{G}, \hat{\boldsymbol{\sigma}})$ with high probability.*

Therefore, Theorem 2.1.7 does not only establish a phase transition but it also shows that there is a gap of a factor of $\ln 2$ between the information-theoretic achievability bound and the best known achieving efficient algorithm on the random regular model. Let us sketch how to prove this converse statement.

Proof sketch of Theorem 2.1.7. In a first step we need to show that DD fails on the random regular model with $m = (1-\varepsilon) m_{\text{DD}}$ tests with high probability. To this end let $m = ck \ln \frac{n}{k}$ and recall that any individual chooses $\Delta = d \ln \frac{n}{k}$ tests uniformly at random. In this contribution we chose the tests with replacement, thus an individual is part of one test twice from time to time. Thus, the average test-degree $\bar{\Gamma}$ turns out to be $\frac{dn}{ck}$. Analogously as before, we suppose for gathering the main idea of the proof that there were no stochastic dependencies (which is clearly false). In this simplified version, the probability for an individual to be disguised is roughly given by

$$\left(1 - \left(1 - \frac{k}{n}\right)^{\bar{\Gamma}}\right)^{\Delta} \sim \left(1 - \exp\left(-\frac{d}{c}\right)\right)^{\Delta}.$$

Therefore, the expected number of totally disguised uninfected individuals is

$$\mathbb{E}\left[|V_{0+}|\right] \sim (n-k)\left(1 - \exp\left(-\frac{d}{c}\right)\right)^{\Delta}.$$

Actually, this argument as well as the fact that the value is concentrated, can be proven rigorously by describing the number of infected and uninfected individuals per test as a family of independent binomials conditioned on a (not too unlikely) event. Furthermore, it is possible to prove that DD achieves its best results for the choice $d = c \ln 2$, therefore $|V_{0+}| \sim n 2^{-\Delta} \sim n^{1-(1-\theta)c\ln^2 2}$.

However, it is not clear at all how to find a rigorous argument which enables us to calculate the size of V_{1--}. Let us change our point of view and observe that DD identifies an infected individual in a positive

test a if and only if, besides x, there are no elements of V_1 nor V_{0+} in a. Therefore, let

$$\boldsymbol{W} = \text{number of positive tests containing exactly one element of } V_1 \cup V_{0+}.$$

We are left with calculating the expectation of $\boldsymbol{W}$ which turns out to be a mildly delicate calculation involving the description of the number of uninfected disguised, uninfected non-disguised and infected individuals per test as a family of independent conditioned multinomial-variables. Nevertheless, it can be shown that

$$\mathbb{E}[\boldsymbol{W}] \sim \frac{k\Delta}{2}\exp\left(-\ln 2 n^{(1-\theta)(1-c\ln^2 2)}\right)$$

which tends to zero below the DD threshold. Thus, by Markov's inequality we already established the converse statement. Let us furthermore show how to calculate the size of V_{1--} from that point. It is possible to show that $\boldsymbol{W}$ is actually tightly concentrated around its mean, thus

$$\boldsymbol{W} \sim \frac{k\Delta}{2}\exp\left(-\ln 2 n^{(1-\theta)(1-c\ln^2 2)}\right).$$

Now, we can rigorously calculate the probability that a given infected individual x does not belong to V_{1--}, as

$$\mathbb{P}(x \notin V_{1--} \mid x \in V_1) \sim \binom{k\Delta - \boldsymbol{W}}{\Delta}\binom{k\Delta}{\Delta}^{-1}.$$

Indeed, such an infected individual would need to choose all its Δ edges out of the $k\Delta - \boldsymbol{W}$ edges belonging to tests containing either a totally disguised individual or a second infected one. By plugging in $m = (1-\varepsilon)m_{\mathtt{DD}}$ Markov's inequality shows that DD fails with high probability. Nevertheless, we are still left to prove that the greedy extension SCOMP fails as well.

To this end, let us observe that there are $\Omega(|V_{0+}|) \gg k$ totally disguised uninfected individuals which are contained in exactly Δ tests. Indeed, with high probability, each individual is contained in at least $\Delta - \ell$ ($\ell = O(1)$) different tests which can be easily verified. Furthermore, the probability of being in exactly $\Delta - \ell'$ tests for $0 \le \ell' \le \ell$ is $\Omega(1)$, thus a positive fraction of all totally disguised uninfected individuals is in Δ different tests. Therefore, already the first individual taken by SCOMP is only infected with probability $\frac{k}{\Omega(|V_{0+}|)+k} = o(1)$. Thus, SCOMP fails with high probability. □

While the previous result is a converse statement about the performance of specific algorithms, we could not answer the question whether there is an efficient algorithm achieving at $m_{\text{non-ada}}$ on the random regular model. Fortunately, we could prove that this gap between information-theoretic and algorithmic achievability is not due to the group testing problem itself. More precisely, we can introduce a different (random) pooling scheme coming with an efficient decoding algorithm called *Spatial Inference Vertex Cover*-algorithm (SPIV) which succeeds with $m = (1+\varepsilon)m_{\text{non-ada}}$ many tests at inference of $\boldsymbol{\sigma}$ with high probability.

Theorem 2.1.8 (Theorem 1.2 of [46])**.** *Let $\varepsilon > 0$ be a constant. Then there is a pooling scheme called* spatially coupled random regular model $\mathcal{G}_{sc}$ *coming with an efficient inference algorithm* SPIV *which succeeds at inferring $\boldsymbol{\sigma}$ from $(\mathcal{G}_{sc}, \hat{\boldsymbol{\sigma}})$ with $m = (1+\varepsilon)m_{\text{non-ada}}$ many tests.*

In the specific case of Theorem 2.1.8 we will not give an explicit proof sketch as the explanation of the pooling scheme and the inference algorithm allow presenting the proof idea on the fly whereby we follow the contribution [46].

We start by defining the pooling scheme. The main idea of the *spatial coupling* has its origins in coding theory [15, 123, 124]. It was applied, for instance, to LDPC-codes. The key idea is to add some kind of geometry to the random regular graph such that, for a specific individual, the neighbourhood looks fairly similar in both models (as the random regular model is known to be information-theoretic optimal). But this geometry constraint allows a trick at inference. Let us partition the set of individuals into $\ell \sim \sqrt{\ln n}$ *compartments* $V[1], \dots, V[\ell]$ of size n/ℓ and the m tests into compartments $F[1], \dots, F[\ell]$ as well. Let $s \sim \ln\ln n$ be the size of the *sliding window*. Then we furthermore add $10ks\ln n/\ell$ additional tests into a larger compartment $F[0]$ whose purpose will become clear in due course.

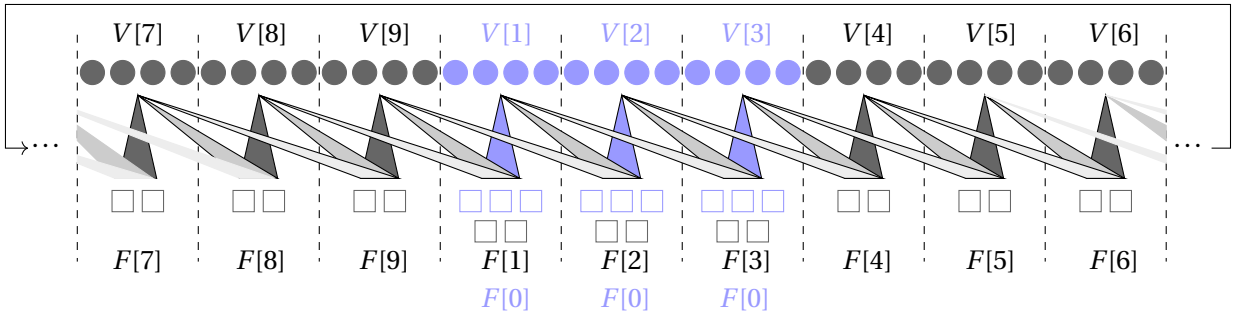

Figure 2.2.: The spatially coupled test design with $n = 36, \ell = 9, s = 3$. The graphic is modified after [41]. The individuals in the seed (blue) join additional tests from $F[0]$.

As in the random regular graph, we let $\Delta = c\ln 2 \ln \frac{n}{k}$ denote the individual degree if $m = ck\ln\frac{n}{k}$ and let the individuals choose their tests as follows.

- For $i = 1,\dots,\ell$ and $j = 1,\dots,s$ every individual $x \in V[i]$ chooses independently from all other randomness Δ/s tests from $F[i+j-1]$ without replacement.
- Those individuals from $V[1] \cup \dots \cup V[s]$ independently choose $10\ln(2)\ln n$ additional tests from $F[0]$ uniformly at random without replacement.

All indices of compartments need to be read such that $V[\ell+i] = V[i]$ and $F[i+\ell] = F[i]$ for $i = 1\dots\ell$. Thus the pooling graph has a ring structure such that, locally seen, all edges from an individual are going to its *right* whereas all incoming edges in a test come from the *left*. A visualisation can be found in Figure 2.2.

Why should such a graph perform better? Suppose we already classified all individuals in compartments $V[1],\dots,V[h]$ successfully. If we now are about to infer the infection status of those individuals in $V[h+1]$ we can use a lot of information. Indeed, in the tests in $F[h+1]$ there are just the individuals of $V[h+1]$ unclassified whilst in the tests in $F[h+1+j]$ at least proportion $(s-j)/s$ is already known. Therefore, if an *early* test emphasises that an individual was infected or uninfected, this information might be more confidential than an information gained in a test far to the right. This is exactly the idea behind the `SPIV` algorithm.

The main challenge of inference is to distinguish the k infected individuals from those $\omega(k)$ totally disguised uninfected individuals. Therefore, we introduce a random variable $\boldsymbol{W}_{x,j}$ whose distribution differs for real infected individuals x and uninfected but disguised individuals. By choice of Δ, the number of infected individuals per test a is (approximately) $\mathbf{Po}(\ln 2)$ distributed. But it turns out that this number's distribution changes when conditioning on specific events.

We define for an individual $x \in V[i+1]$ and a compartment $F[i+j]$ for $j = 1\dots s$

$$\begin{aligned}\boldsymbol{W}_{x,j} = {}& \text{number of tests in compartment } F[i+j] \text{ containing } x \\ & \text{and no infected individual from the preceding compartments.}\end{aligned}$$

Because the number of infected individuals in a test from a specific compartment is $\mathbf{Po}(\ln 2/s)$ distributed, the probability that a specific test contains no infected individual from any of the compartments $V[i+j-s+1],\dots,V[i]$ turns out to be $2^{-(s-j)/s}$, therefore

$$\mathbb{E}\left[\boldsymbol{W}_{x,j} \mid x \in V[i+1] \cap V_1\right] \sim \frac{\Delta}{s} 2^{j/s-1}.$$

Equivalently, we can calculate

$$\mathbb{E}\left[\boldsymbol{W}_{x,j} \mid x \in V[i+1] \cap V_{0+}\right] \sim \frac{\Delta}{s}\left(2^{j/s}-1\right)$$

as the sole difference is that a specific test a needs to contain at least one infected individual which is not x. Therefore, the number of infected individuals in a, conditioned on containing an element of V_{0+}, turns out to be a conditioned Poisson distributed variable $\mathbf{Po}_{\geq 1}(\ln 2)$. We clearly find

$$\mathbb{E}\left[\boldsymbol{W}_{x,j} \mid x \in V[i+1] \cap V_{0+}\right] < \mathbb{E}\left[\boldsymbol{W}_{x,j} \mid x \in V[i+1] \cap V_1\right].$$

Unfortunately, it turns out that trying to distinguish the infection status solely on the sum of those random variables does not suffice as there will be too much fluctuation [46, Section 4.3]. Therefore, we should try to incorporate the intuition that tests in a compartment close to the individual might contain more valuable information. We define

$$\boldsymbol{W}^*_x = \sum_{i=1}^{s-1} w_j \boldsymbol{W}_{x,j} \qquad \text{where} \qquad w_j \sim -\ln\left(1-2^{-j/s-1}\right). \tag{2.1.4}$$

Thus, W^*_x weighs information from closer tests higher. The optimal value of the weighs w_j was obtained by a delicate large deviation analysis using a Lagrange optimisation [46, Sections 4.7 and 4.8].

If we call the individuals in $V[1] \cup \ldots \cup V[s] = V_{seed}$ the *seed* (those individuals which are connected to tests in $F[0]$) the idea is now to classify the individuals subsequently starting at the seed. As we clearly cannot compute $\boldsymbol{W}^*_x$ as this would require knowledge about $\boldsymbol{\sigma}$, we have to rely on the current estimate of the ground-truth, thus let

$$W_{x,j}(\tau) = \left|\left\{a \in \partial x \cap F[i+j-1] : \max_{y \in \partial a \cap (V[1] \cup \ldots \cup V[i])} \tau_y = 0\right\}\right|$$

and analogously

$$W^*_x(\tau) = \sum_{i=1}^{s-1} w_j W_{x,j}(\tau).$$

We apply `DD` to the seed individuals as the graph on the seed individuals and the tests of $F[0]$ is an instance of the random regular model which is dense enough to allow classification by `DD`. We let τ be the current estimate of $\boldsymbol{\sigma}$ which we initialise to the all zero vector outside of the seed. Then we proceed with the individuals in the next compartment and

- declare an individual as uninfected if it is in at least one negative test,
- declare an individual x as uninfected, if $W^*_x(\tau)$ is smaller than the expected value for infected individuals under $\boldsymbol{\sigma}$,
- declare an individual as infected otherwise.

We iterate with this procedure through the graph and stop when all individuals are classified. We call the estimate of the ground-truth produced by this algorithm τ. Unfortunately, it turns out that τ does probably not coincide with $\boldsymbol{\sigma}$. But fortunately, if $|F[1] \cup \ldots F[\ell]| = (1+\varepsilon)m_{\inf}$, we have at least [46, Proposition 4.6]

$$|\{x : \tau_x \neq \boldsymbol{\sigma}_x\}| = kn^{-\Omega(1)}.$$

Thus, using only as many tests as given by the universal counting bound, we achieve already *partial recovery* of $\boldsymbol{\sigma}$ with high probability. To this end observe that $|F[0]| = o(m)$ is of lower order.

Corollary 2.1.9. *The `SPIV` algorithm on $\mathcal{G}_{sc}$ with $(1+\varepsilon)m_{\inf}$ tests succeeds at $(1-o(1))-$partial recovery of $\boldsymbol{\sigma}$ from $(\mathcal{G}_{sc}, \hat{\boldsymbol{\sigma}})$ with high probability.*

But of course we want to achieve *exact* recovery. It turns out that a rigidity argument helps in establishing an exact recovery statement. If $m = (1+\varepsilon)m_{\text{non-ada}}$, each infected individual is contained in at least $\Theta(\ln n)$ tests in which no second infected individual appears. Let $S_x(\tau)$ denote the number of (positive) tests in which an individual would be the only infected individual under τ if it was infected. Then a combinatorial clean-up step could read as follows.

For $\ln n$ steps repeat thresholding $S_x(\tau)$ with respect to the current estimate τ in order to classify individual x as infected or uninfected. By the expansion properties of the random graph it is possible to prove that this reduces the number of misclassified individuals by a factor of at least 3 each step. Thus, the final estimate τ coincides with $\boldsymbol{\sigma}$ w.h.p.. This is again an example of the splitting of the phase transition point into two parts. The non-adaptive tail of the phase transition point corresponds to the combinatorial observation of having some kind of local rigidity, similarly as we already saw in the proof sketch of Theorem 2.1.3.

We can now state the SPIV algorithm completely and formally.

Input: Spatially coupled pooling scheme $\mathcal{G}_{sc} = (V \cup F, E)$, test-results $\hat{\boldsymbol{\sigma}} \in \{0,1\}^m$
Output: Estimate $\tilde{\sigma}$ of $\boldsymbol{\sigma}$

Infer the infection status for all $x \in V_{seed}$ by DD and obtain $\tilde{\sigma}_{V_{seed}}$.
Initialise $\tilde{\sigma}_x = 0$ for all $x \notin V_{seed}$.
for $i = s, \dots, \ell - 1$ **do**
  for $x \in V[i+1]$ **do**
    if *x is in at least one negative test* **then**
      $\tilde{\sigma}_x = 0$ // set infection status to uninfected
    else if $W_x^*(\tilde{\sigma}) < (1-\zeta)\sum_{j=1}^{s-1} \frac{\Delta}{s} w_j 2^{j/s-1}$ **then**
      $\tilde{\sigma}_x = 0$ // set infection status to uninfected
    else
      $\tilde{\sigma}_x = 1$ // classify as infected
Let $\tilde{\sigma}^{(1)} = \tilde{\sigma}$.
for $i = 1, \dots, \ln n$ **do**
  For all $x \in V[s+1] \cup \dots \cup V[\ell]$ calculate

$$S_x\left(\tilde{\sigma}^{(i)}\right) = \sum_{a \in \partial x : \hat{\boldsymbol{\sigma}}_a = 1} \mathbf{1}\left\{\forall y \in \partial a \setminus \{x\} : \tilde{\sigma}_y^{(i)} = 0\right\}$$

  Let $\tilde{\sigma}_x^{(i+1)} = \begin{cases} \tilde{\sigma}_x^{(i)} & \text{if } x \in V[1] \cup \dots \cup V[s], \\ \mathbf{1}\left\{S_x\left(\tilde{\sigma}^{(i)}\right) > \ln^{1/4} n\right\} & \text{otherwise} \end{cases}$
return $\tilde{\sigma}^{(\lceil \ln n \rceil)}$

Algorithm 4: The SPIV algorithm by Coja-Oghlan et al. [46] using a spatially coupled pooling scheme.

We should emphasise that the decoding algorithms in prior applications on spatially coupled inference graphs like coding or compressed sensing were approximate message passing or BP like algorithms [106, 123, 124]. On the first glance, those algorithms are clearly much more sophisticated than the combinatorial and easy to digest SPIV algorithm. But based on the discussion about WP and DD, we should not be too surprised that actually the decisions based on (a normalised version of) W_x^* correspond to the estimate after one round of BP.

Therefore, we have proven that the group testing problem is very special with respect to its solvability. Either, it is completely impossible or it is easy (there is an efficient algorithm for inference) but there is no hard phase. Such phenomena are known as *impossible-easy* phase transitions. Furthermore, we discussed that there is indeed a gap between adaptive algorithms and non-adaptive algorithms, thus multiple stages of testing can decrease the number of tests required in total, at least if many individuals are supposed to be infected.

As we will discuss next, we observe similar phenomena even if we restrict the maximum capacity of a test or the number of tests an individual can be part of.

Let us first present the results we obtained with respect to the $\Delta-$divisible restricted group testing problem. We recall that we already provided a result showing that each non-adaptive group testing scheme which tests each individual at most Δ times fails with high probability if it contains less than $m_{\text{non-ada}}(\Delta)$ tests where

$$m_{\text{non-ada}}(\Delta) = \max\left\{\exp(-1)\Delta k^{1+\frac{(1-\theta)}{\Delta\theta}}, \Delta k^{1+\frac{1}{\Delta}}\right\}.$$

We will show that the DD algorithm on the random regular model solves the group testing problem almost at this bound, thus let $m_{\mathtt{DD}}(\Delta)$ denote the phase transition point of DD in the random (almost) regular model, then we have

$$m_{\mathtt{DD}}(\Delta) = \max\left\{\Delta k^{1+\frac{(1-\theta)}{\Delta\theta}}, \Delta k^{1+\frac{1}{\Delta}}\right\} \tag{2.1.5}$$

which is a factor of e away from the converse statement for $\theta < 1/2$ and matches it for $\theta \geq 1/2$. We define the random almost regular model in a way that each individual chooses Δ tests uniformly at random with replacement, thus the test-degree sequence is random (but sufficiently concentrated). Then we obtain the following theorem.

Theorem 2.1.10 (Theorem 3.3 [88])**.** *Let $\mathcal{G}_\Delta$ be the random regular model where each individual joins $\Delta = O(\ln^{1-\delta} n)$ $(0 < \delta \leq 1)$ tests on m tests. Then , if $m \geq (1+\varepsilon)m_{DD}(\Delta)$, DD succeeds at inference of $\boldsymbol{\sigma}$ from $(\mathcal{G}_\Delta, \hat{\boldsymbol{\sigma}})$ with probability $\Omega(1)$ if $\Delta = O(1)$ and with probability $(1-o(1))$ if $\Delta = \omega(1)$.*

Thus, Theorem 2.1.10 shows in combination with the information-theoretic converse at $m_{\text{non-ada}}(\Delta)$ that the random-regular model is information-theoretic optimal for large θ and that DD performs optimally in this regime.

Proof sketch of Theorem 2.1.10. Analogously as in the DD analysis in the unrestricted problem (Theorem 2.1.7), we bound the expected number of positive tests containing exactly one infected and no uninfected totally disguised individual $\mathbb{E}[\boldsymbol{W}]$ with the only difference that some probabilities have to be calculated slightly more carefully if Δ does not diverge. Again we require to describe some local properties of tests by a family of independent conditioned multinomial random variables. Of course, the resulting formulas differ such that

$$\mathbb{E}[\boldsymbol{W}] \sim \Delta k \cdot \left(1 - (1+\varepsilon)^{-1} k^{-1/\Delta}\right).$$

Nevertheless, we will omit any details here.

But in this case, as we want to proof that DD actually succeeds, we require a stronger result as Markov's inequality applied on the expectation of $\boldsymbol{W}$ does not suffice. Luckily, it turns out that $\boldsymbol{W}$ is fairly concentrated around its expectation. Therefore, with high enough probability, we find

$$\boldsymbol{W} \sim \Delta k \cdot \left(1 - (1+\varepsilon)^{-1} k^{-1/\Delta}\right).$$

Now, we can rigorously calculate the probability that a given infected individual x is classified falsely by DD, thus does not belong to V_{1--}, as

$$\mathbb{P}(x \notin V_{1--} \mid x \in V_1) \sim \binom{k\Delta - \boldsymbol{W}}{\Delta}\binom{k\Delta}{\Delta}^{-1} \sim \left((1+\varepsilon)^{-1} k^{-1/\Delta}\right)^{\Delta}.$$

Indeed, such an infected individual would need to choose all its Δ edges out of the $k\Delta - \boldsymbol{W}$ edges belonging to tests containing either a totally disguised individual or a second infected one. A standard calculation involving a case distinction between small and large θ suffices in order to obtain the expected number of individuals $x \in V_1 \setminus V_{1--}$ and to observe that this number is at most $(1+\varepsilon)^{-\Delta}$ if $m = (1+\varepsilon)m_{\mathtt{DD}}(\Delta)$ tests are carried out and the theorem follows by Markov's inequality. □

Of course, above's theorem does only give an achievability result for DD on a specific design. We can actually prove that DD is failing below $m_{\mathtt{DD}}(\Delta)$ on the random almost regular model. Clearly, we only need to show this for small θ as the assertion for large θ follows from the universal converse bound.

Theorem 2.1.11 (Theorem 3.4 of [88])**.** *Let $0 < \theta < 1/2$ and let $\mathcal{G}_\Delta$ be the random almost regular model on m tests. If $m \leq (1-\varepsilon)\Delta k^{1+(1-\theta)/(\Delta\theta)}$, DD fails at inference of $\boldsymbol{\sigma}$ from $(\mathcal{G}_\Delta, \hat{\boldsymbol{\sigma}})$ with probability $1-o(1)$ of $\Delta = \omega(1)$ and with probability $\Omega(1)$ if $\Delta = \Theta(1)$.*

Proof sketch of Theorem 2.1.11. Similarily as in the achievability proof, we find by analysing the number of positive tests containing exactly one infected and no uninfected totally disguised individual that

$$\mathbb{E}\left[|V_{1--}(\mathcal{G}_\Delta)|\right] \sim k\left(1-\left(1-\exp\left(-(1-\varepsilon)^{-\Delta}(1-1/\Delta)\right)\right)^{\Delta}\right).$$

But then Markov's inequality readily yields that the probability of having at least γk infected individuals outside of $V_{1--}(\mathcal{G})$ is at least

$$1-\frac{1-\left(1-\exp\left(-(1-\varepsilon)^{-\Delta}(1-1/\Delta)\right)\right)^{\Delta}}{1-\gamma}$$

and therefore, the assertion of the theorem follows. Indeed, this probability is $1-o(1)$ for $\Delta\to\infty$ for any $0<\gamma<1$ and if Δ is a constant, a suitable choice of γ makes the probability $\Omega(1)$. □

Thus, we understand DD on the random regular model completely.

We are left to discuss the $\Gamma-$sparse case. This is an especially interesting case as the underlying (almost regular) model has constant degrees on both sides, thus the construction of the graph requires a lot more attention. Ultimately, we will see that DD actually succeeds at the information-theoretic universal converse $m_{\text{non-ada}}(\Gamma)$ on a suitable chosen almost regular graph for all θ outside of a set of Lebesgue-measure zero. Recall that

$$m_{\text{non-ada}}(\Gamma)=\max\left\{\left(1+\left\lfloor\frac{\theta}{1-\theta}\right\rfloor\right)\frac{n}{\Gamma},2\frac{n}{\Gamma+1}\right\}$$

and define

$$m_{\mathtt{DD}}(\Gamma)=\begin{cases}\max\left\{\left(1+\left\lfloor\frac{\theta}{1-\theta}\right\rfloor\right)\frac{n}{\Gamma},2\frac{n}{\Gamma+1}\right\}, & \frac{\theta}{1-\theta}\notin\mathbb{Z}\\ \max\left\{\left(2+\left\lfloor\frac{\theta}{1-\theta}\right\rfloor\right)\frac{n}{\Gamma},2\frac{n}{\Gamma+1}\right\}, & \frac{\theta}{1-\theta}\in\mathbb{Z}.\end{cases}$$

Therefore, $m_{\mathtt{DD}}$ coincides with the universal converse if $\theta/(1-\theta)$ is no integer.

Let us first describe how to obtain the almost regular pooling scheme. If $\theta\geq 1/2$, we create the random pooling scheme by the configuration model as a random regular multi-graph $\mathcal{G}(\Gamma,\Delta)$. Thus, with $\Delta=m\Gamma/n$, each individual node gets Δ clones and each test node gets Γ clones and a perfect matching is chosen uniformly at random. On the other hand, if $\theta<1/2$, it turns out that this model is not optimal. In this case, we select $\gamma\leq\frac{2n}{\Gamma+1}$ individuals $X\subset\{x_1,\dots,x_n\}$ uniformly at random and put them apart. The exact value of γ is chosen such that the remaining individuals can be pooled by $\mathcal{G}(\Gamma-1,2)$. Now we select a uniform matching between the tests $F(\mathcal{G}(\Gamma-1,2))$ and the remaining individuals X. For a brief check of sanity, observe that this pooling is only possible for $m\geq 2\frac{n}{\Gamma+1}$ by comparing degrees. Furthermore observe that in the final graph any test has size $\Gamma-1$ or Γ. We will call this graph model $\mathcal{G}^*(\Gamma)$. Therefore, let us define

$$\mathcal{G}(\Gamma)=\begin{cases}\mathcal{G}(\Gamma,m\Gamma/n), & \theta\geq 1/2\\ \mathcal{G}^*(\Gamma), & \theta<1/2.\end{cases}.$$

With this model at hand, we can state the achievability result.

Theorem 2.1.12 (Theorems 4.10 and 4.18 of [88])**.** *If DD is applied on an instance of $\mathcal{G}(\Gamma)$ with $m\geq m_{DD}(\Delta)$, it succeeds at inference of $\boldsymbol{\sigma}$ from $(\mathcal{G}(\Gamma),\hat{\boldsymbol{\sigma}})$ with high probability.*

Observe that interestingly, the achievability bound is tight, thus we do not even need a multiplicative factor of $(1+\varepsilon)$.

Proof sketch of Theorem 2.1.12. The proof comes in two major steps. First, we give an achievability result of DD on the random regular model $\mathcal{G}(\Gamma,\Delta)$ for any choice of θ. More precisely, let

$$\Delta_{\mathtt{DD}}=\max\left\{2,1+\left\lfloor\frac{\theta}{1-\theta}\right\rfloor\right\}$$

denote the individual degree, then we have the following.

Achievability on the random regular model. DD recovers $\boldsymbol{\sigma}$ from $(\mathcal{G}_\Gamma(\Gamma,\Delta_{\mathtt{DD}}),\hat{\boldsymbol{\sigma}})$ correctly w.h.p..

Observe that by definition of the random regular graph we have $m \geq \Delta_{\mathtt{DD}}\frac{n}{\Gamma}$. It turns out that this part of the proof follows completely analogous ideas as the achievability proof of DD in the $\Delta-$divisible case and will thus be omitted. Of course, a technical challenge is to deal with the (higher) number of multi-edges. Fortunately, the heavy stochastic dependencies caused by the configuration model vanish as we can describe the important local properties of tests by a family of independent conditioned multinomial random variables as before. The sole technical challenge is to deal with the (higher) number of multi-edges.

As we have proven the success of DD with $m = \Delta_{\mathtt{DD}}\frac{n}{\Gamma}$ tests, the part of the theorem for $\theta \geq 1/2$ follows. But it turns out that we can indeed perform better if $\theta < 1/2$. The main idea is the following. We let $\mathcal{G}_\Gamma^{*,r}$ be the subgraph created in the first step of generating $\mathcal{G}(\Gamma)$, thus an instance of $\mathcal{G}(\Gamma-1,2)$ on $n' = n-\gamma$ individuals where $\gamma \leq \frac{2}{\Gamma+1}n$. Without loss of generality we suppose that $\gamma = \frac{2}{\Gamma+1}n$ as otherwise we could fill the instance with dummy-individuals which are known to be uninfected.

We furthermore let $\boldsymbol{\sigma}[\mathcal{G}_\Gamma^{*,r}]$ and $\hat{\boldsymbol{\sigma}}[\mathcal{G}_\Gamma^{*,r}]$ be the infection status vector and test-result vector on this induced subgraph. Observe that for $\theta < 1/2$ we have $\Delta_{\mathtt{DD}} = 2$. Now we obtain the result as follows.

(i) Let $\theta' = \theta'(\theta)$ describe the prevalence on $\boldsymbol{\sigma}[\mathcal{G}_\Gamma^{*,r}]$, then we have $\theta' \sim \theta$ with high probability.

(ii) We already know that $m = 2\frac{n-\gamma}{\Gamma-1} = 2\frac{n}{\Gamma+1}$ tests suffice for DD to infer $\boldsymbol{\sigma}[\mathcal{G}_\Gamma^{*,r}]$ from $\mathcal{G}(\Gamma-1,2)$ and $\hat{\boldsymbol{\sigma}}[\mathcal{G}_\Gamma^{*,r}]$.

(iii) We prove that adding the matching edges in order to generate $\mathcal{G}(\Gamma)$ from $\mathcal{G}(\Gamma-1,2)$ does, with high probability, enable DD to infer $\boldsymbol{\sigma}$ from $\mathcal{G}(\Gamma)$.

It is easy to see that (i) follows from the Chernoff bound as the number of infected individuals in $\boldsymbol{\sigma}[\mathcal{G}_\Gamma^{*,r}]$ is a hypergeometrically distributed random variable $\boldsymbol{k}' \sim H\left(n,k,n'\right)$ and thus concentrated around its mean $\boldsymbol{k}' \sim \frac{\Gamma-1}{\Gamma+1}k$. Furthermore, (ii) is a direct consequence of above's result with respect to achievability on the random regular model. Therefore, we only need to discuss (iii).

The key property of the proof is that for $k = o(\sqrt{n})$ we do not find two infected individuals in any bounded part of the graph. Suppose that an individual x gets connected to a negative test. If x is uninfected itself, the test stays uninfected an DD recovers x (and the other individuals) correctly. If x is infected, each uninfected individual in the test joins a second test which is negative as well with high probability due to the fact that there are no two infected individuals within a finite range within the random graph with high probability.

If on the other hand x connects to a positive test and is uninfected, a similar argument shows that the causing already contained positive individual can be identified by DD through both its tests with high probability. Thus the previously infected individual in $\boldsymbol{\sigma}[\mathcal{G}_\Gamma^{*,r}]$ can be still inferred by DDand x will be declared uninfected.

Finally, if x is infected, it will connect to a negative test with high probability. Indeed, as it is infected with probability $\sim k/n$ and its test is chosen uniformly at random, this test will contain a second infected individual with probability $\sim k^2/n^2$ and a union bound shows that therefore, with high probability, all infected individuals connect to negative tests in the matching process.

Therefore, DD infers $\boldsymbol{\sigma}$ from $\mathcal{G}(\Gamma)$ and $\hat{\boldsymbol{\sigma}}$ w.h.p., if it infers $\boldsymbol{\sigma}[\mathcal{G}_\Gamma^{*,r}]$ from $\mathcal{G}^{*,r}$ and $\hat{\boldsymbol{\sigma}}[\mathcal{G}_\Gamma^{*,r}]$ w.h.p..

Above's bounding of the probability of bad events can be done rigorously and hinges on precise but elementary calculations which make extensive use of the fact that a uniformly at random chosen subset of individuals was put apart before creating the regular part of the graph. □

Next, we will present the results we obtained with respect to adaptive group testing. Afterwards, we summarise the results on non-adaptive group testing as well as adaptive group testing shortly in Section 2.1.3.

2.1.2. Adaptive Group Testing

As in the section about non-adaptive group testing, we split our results into a section about information-theoretic aspects as well as algorithmic aspects respectively.

2.1.2.1. Information-theoretic results

With respect to the information-theoretic phase transitions, we achieved primarily results in the context of sparsity constrained group testing. More precisely, while the contribution of Gandikota et al. [86] only provided non-adaptive converse statements, we give information-theoretic converse results for all adaptive pooling schemes in the $\Delta-$divisible and the $\Gamma-$sparse case. To this end, let $\Delta = \ln^{1-\delta} n$ and $\Gamma = \left(\frac{n}{k}\right)^{\beta}$ with $0 < \delta \leq 1$ and $0 \leq \beta < 1$. Then define

$$m_{\text{inf}}(\Delta) = \exp(-1)\Delta k^{1+\frac{1-\theta}{\Delta\theta}} \qquad \text{and} \qquad m_{\text{inf}}(\Gamma) = \frac{n}{\Gamma}$$

as the information-theoretic threshold for any test-design in which individuals get tested at most Δ times and, respectively, each test has size at most Γ. We let $\boldsymbol{\sigma}$ denote the ground-truth and for an $\ell-$stage testing procedure $\mathcal{G} = (\mathcal{G}_1, \dots, \mathcal{G}_\ell)$ we let $\hat{\boldsymbol{\sigma}} = (\hat{\boldsymbol{\sigma}}^1, \dots, \hat{\boldsymbol{\sigma}}^\ell)$ denote the sequence of test-results.

Theorem 2.1.13 (Theorems 3.1 and 6.2 of [88])**.** *Suppose $\mathcal{G}(\Delta) = (\mathcal{G}_1(\Delta), \dots, \mathcal{G}_\ell(\Delta))$ is an $\ell-$stage testing procedure such that any individual gets tested at most Δ times in total and respectively that $\mathcal{G}(\Gamma) = (\mathcal{G}_1(\Gamma), \dots, \mathcal{G}_\ell(\Gamma))$ is an $\ell-$stage testing procedure such that any test contains at most Γ individuals. Then the following holds.*

- *If $\mathcal{G}(\Delta)$ contains at most $(1-\varepsilon)m_{\text{inf}}(\Delta)$ tests, inference of $\boldsymbol{\sigma}$ from $(\mathcal{G}(\Delta), \hat{\boldsymbol{\sigma}})$ fails with high probability.*
- *If $\mathcal{G}(\Gamma)$ uses at most $(1-\varepsilon)m_{\text{inf}}(\Gamma)$ tests, inference of $\boldsymbol{\sigma}$ from $(\mathcal{G}(\Gamma), \hat{\boldsymbol{\sigma}})$ fails with probability $\Omega(1)$.*

Proof sketch of Theorem 2.1.13. The proof of the first part of the theorem resembles the counting based proof of the universal counting bound in the unrestricted group testing problem.

We first show that any adaptive strategy with m tests succeeds with probability at most

$$\frac{\sum_{i=0}^{\Delta k}\binom{m}{i}}{\binom{n}{k}} \sim \frac{\exp\left(H(\Delta k/m)\right)}{\binom{n}{k}}$$

which is given by a short calculation using the Nishimori property that guarantees that choosing one possible solution is the best an inference algorithm can do. We directly find that there can be at most Δk positive tests in total, as each infected individual can be tested at most Δ times. Therefore, the summation accounts for all possible choices of positive tests. Plugging in $m = (1-\varepsilon)m_{\text{inf}}(\Delta)$ yields the assertion of the theorem.

The second part, namely the $\Gamma-$sparse case, might be actually called a folklore argument. Indeed, $\frac{n}{\Gamma}(1-o(1))$ tests of size at most Γ are clearly required to test $n-o(n)$ of all individuals at least once, which is a necessary requirement for inference. □

Overall it turns out that under both kinds of restrictions the adaptive converse statements are strictly below the non-adaptive converse results and non-adaptive achievability results. In the case of $\Delta-$ divisible group testing, this is only true for $\theta > 1/2$. This alone can clearly not answer the question whether there is an adaptivity-gap in restricted group testing or not. But the next section will show that such a gap really exists.

2.1.2.2. Algorithms

We introduce two new algorithms for the two restricted group testing models. Let us start with the $\Delta-$divisible case which can be solved by Algorithm 5.

Clearly, this algorithm strongly resembles the binary splitting approach of Hwang [98] and Allemann [13]. The only major difference is that we do not split groups into halves in order to guarantee the

Input: n, k, Δ
Output: Estimate $\tilde{K} \subset \{x_1, \dots, x_n\}$ of the infected individuals.
Set $\tilde{n} = \left(\frac{n}{k}\right)^{\frac{\Delta-1}{\Delta}}$.
Set $\tilde{K} = \emptyset$.
Arbitrarily divide the n individuals into $n/\tilde{n}$ groups of size $\tilde{n}$.
Test each of these groups and discard those with a negative result.
Denote the remaining groups by $A_j^{(0)}$.
for $i = 1$ *to* $\Delta - 1$ **do**
  for *each group* $A_j^{(i-1)}$ *from the previous stage* **do**
    Arbitrarily divide all individuals in $A_j^{(i-1)}$ into $\tilde{n}^{1/(\Delta-1)}$ sub-groups of size $\tilde{n}^{1-i/(\Delta-1)}$.
    Test each sub-group and discard any that returns a negative outcome.
    Label the remaining sub-groups as $A_j^{(i)}$.
Add the individuals from all of the remaining singleton groups $A_j^{(\Delta-1)}$ to $\tilde{K}$.

Algorithm 5: Splitting algorithm for the $\Delta-$divisible restricted group testing as posed in [88].

$\Delta-$restriction. As we will see, it does not perform tightly at the information-theoretic converse bound proven before. Let

$$m_{\text{adap-alg}}(\Delta) = \Delta k^{1+\frac{1-\theta}{\theta\Delta}}$$

be an algorithmic threshold. Then we find the following theorem.

Theorem 2.1.14 (Theorem 5.1 of [88])**.** *There is a choice of* $\tilde{n} = \tilde{n}(n, k, \Delta)$ *such that Algorithm 5 succeeds with at most* $(1+\varepsilon)m_{\text{adap-alg}}(\Delta)$ *tests.*

Therefore, the algorithm performs for all θ at exactly one of the bounds of non-adaptive group testing ($m_{\text{non-ada}}(\Delta)$) establishing an adaptivity-gap as in the unrestricted group testing problem.

Proof sketch of Theorem 2.1.14. It is clear from the definition of the algorithm that it recovers $\boldsymbol{\sigma}$ correctly as all infected individuals will be tested individually. The core idea is to optimise the choice of $\tilde{n}$ in such a way that tests do not get too large as otherwise they will be very likely positive. Starting initially with groups of size $\tilde{n}$, the size is continuously decreasing whenever the test-outcome is still positive.

Clearly, in the first stage we conduct $\tilde{n}$ tests and as there are k infected individuals, each subsequent stage of testing can produce at most $k\tilde{n}^{\frac{1}{\Delta-1}}$ additional (smaller) tests, therefore

$$m \leq \frac{n}{\tilde{n}} + (\Delta - 1)\, k\tilde{n}^{\frac{1}{\Delta-1}}.$$

The assertion of the theorem follows with $\tilde{n} = n^{(1-\theta)(\Delta-1)/\Delta}$. □

After having understood a splitting approach towards $\Delta-$divisible constrained instances of group testing, we will subsequently present an algorithm for $\Gamma-$sparse group testing. In this case, we will use a standard binary splitting approach as a sub-routine, thus let us shortly describe how it works in detail. Given a group of individuals, test the whole group and if the test is positive, split the group into two equal parts and tests each part. If the test is negative, we know that all individuals are uninfected and do not need to test further. Iterate this process until all positive tests contain exactly one individual. Now we can state Algorithm 6. We stress that the algorithm basically reduces to applying the first round of the Dorfman-algorithm followed by a binary splitting algorithm for all positive tests.

How many tests does this algorithm require? It turns out that at least its first order complexity coincides (up to rounding) with the information-theoretic converse if $\Gamma = O\left(\frac{n}{k}\right)^{\beta}$ for some $\beta \in [0, 1)$. Thus, let

$$m_{\text{adap-alg}}(\Gamma) = \left\lceil \frac{n}{\Gamma} \right\rceil + \frac{\ln \Gamma}{\ln 2} k$$

denote the algorithmic achievability bound.

Input: n, k, Γ
Output: Estimate $\tilde{K} \subset \{x_1, \dots, x_n\}$ of the infected individuals.
Set $T = \lceil n/\Gamma \rceil$ and $\tilde{K} = \emptyset$.
Choose one partition of all individuals into groups $G_1, \dots, G_T$ of size Γ.
for $i = 1 \dots T$ **do**
    Test group G_i.
    if *the outcome is positive* **then**
        Infer the infection status of all individuals in G_i by binary splitting.
    else
        Declare all individuals in G_i as uninfected.
Add the individuals from all of the remaining singleton groups $A_j^{(\Delta-1)}$ to $\tilde{K}$.

Algorithm 6: Splitting algorithm in the $\Gamma-$sparse restricted group testing. A similar version was introduced in [88].

Theorem 2.1.15. *Algorithm 6 succeeds at inference of $\boldsymbol{\sigma}$ requiring at most $(1+\varepsilon)m_{\mathrm{adap-alg}}(\Gamma)$ tests.*

Proof sketch of Theorem 2.1.15. Again, it is clear that the algorithm succeeds at inference. Thus, we only need to bound the number of tests required. At the first stage, we clearly conduct $T = \lceil \frac{n}{\Gamma} \rceil$ tests. Subsequently, we apply the binary splitting algorithm to groups of at most Γ individuals each. But due to Hwang [98] it is known that binary splitting on a group of Γ individuals out of which k_i are infected requires not more than

$$m_{\mathrm{Hwang}} \sim \frac{k_i \ln \frac{\Gamma}{k_i}}{\ln 2}$$

tests. As $\Gamma \geq 2$, we find $\frac{\ln \Gamma}{\ln 2} \geq 1$ and therefore, testing all those groups of size at most Γ requires at most

$$\sum_{i=1}^{T} \frac{k_i \ln \Gamma - k_i \ln k_i}{\ln 2} \leq k \frac{\ln \Gamma}{\ln 2}$$

tests. Therefore, we clearly find

$$m \leq \left\lceil \frac{n}{\Gamma} \right\rceil + k \frac{\ln \Gamma}{\ln 2}$$

yielding the assertion of the theorem. □

Finally, with respect to the unrestricted group testing problem, we find as a direct consequence of the previous discussion on `SPIV` that we can use `SPIV` to infer all individuals with m_{inf} tests within two rounds.

Corollary 2.1.16 (Theorem 1.3 of [46])**.** *There is a two-stage inference algorithm which achieves inference of $\boldsymbol{\sigma}$ with high probability requiring at most m_{inf} tests.*

Indeed, it turns out that applying `SPIV` with the spatially coupled testing strategy using m_{inf} tests does render an estimate $\tilde{\sigma}$ of $\boldsymbol{\sigma}$ where all but $kn^{-\Omega(1)} = o(k/\ln n)$ individuals are identified correctly. Instead of applying the previously described clean-up step based on a local rigidity argument (which requires $m_{\mathrm{non-ada}}$ tests), we proceed as follows.

(i) Test any individual which is infected under $\tilde{\sigma}$ individually.

(ii) Test all individuals which are uninfected under $\tilde{\sigma}$ with the random regular model using `DD`.

Clearly, (i) requires at most $(1+o(1))k = o(m_{\mathrm{inf}})$ tests and the prevalence in (ii) is $o(k/\ln n)$, thus `DD` requires $o(m_{\mathrm{inf}})$ tests in order to succeed.

2.1.3. Summary of phase transitions in group testing

Let us begin by drawing a picture of unrestricted group testing. In Figure 2.3 we present a phase diagram based on the results obtained in [41, 46]. While exact recovery in the red areas (below $m_{\text{non-ada}}$) is not possible within one round of testing, in the light red area adaptive algorithms (e.g. Alleman's algorithm) succeed. The existence of this adaptivity-gap was unknown prior to this thesis's contributions. Furthermore, recovery in the blue area is information-theoretically possible on the random-regular model which was previously only known for $\theta > \frac{\ln 2}{1+\ln 2}$ while the efficient DD algorithm was proven to fail in the light blue area. Finally, we introduced the spatially coupled test-design and the efficient SPIV algorithm which already succeeds in this light blue area. Observe that we also managed to proof that inference of all but $o(k)$ individuals is possible within one round and inference of all individuals within two rounds by SPIV outside of the dark-red area.

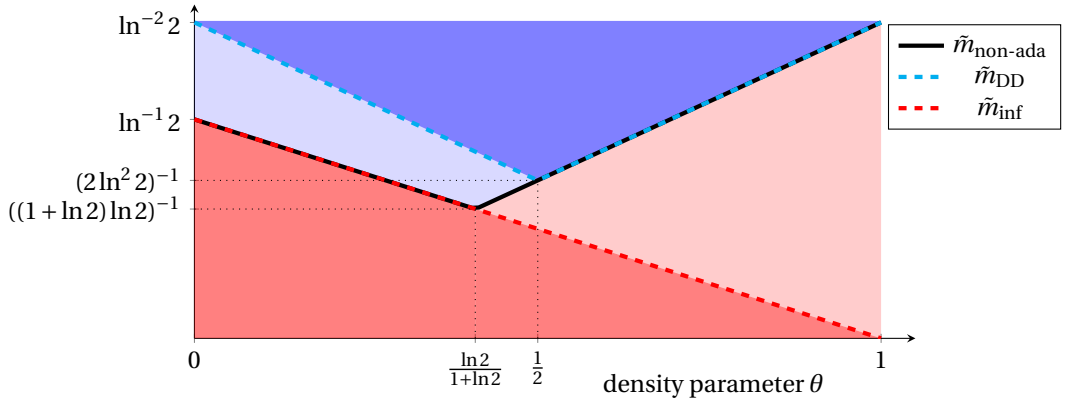

Figure 2.3.: The phase transitions in non-adaptive unrestricted hypergeometric group testing. The graphic is modified after [46, Figure 1]. We write $\tilde{m} = m \cdot (k \ln(n))^{-1}$.

With respect to the $\Delta-$divisible group testing problem, we present the (not completely understood) phase diagram in Figure 2.4. In the dark blue area, the simple COMP algorithm was known to perform on the random-regular model non-adaptively [86] while we proved that DD performs on the same model already in the light blue area. Furthermore, below the black line (red area), every testing scheme (even adaptive schemes) fail while below the red-dotted line all non-adaptive schemes do not succeed. Thus, the yellow area shows a regime where there might be non-adaptive schemes coming with efficient algorithms (e.g. SPIV-like ideas) but they are currently not known. Further, in the yellow and orange area there might be adaptive algorithms for inference of $\boldsymbol{\sigma}$ but they are also currently not known. Finally, the light red area marks a regime where we proved the existence of an adaptivity-gap, thus we found an efficient adaptive algorithm performing in this regime but there cannot be a non-adaptive pooling scheme facilitating inference.

Finally, the discussed phase transitions in the $\Gamma-$sparse group testing problem for $\Gamma = \Theta(1)$ are given in Figure 2.5. Above the blue line, the COMP algorithm studied by Gandikota et al. [86] succeeds at inference on the random regular model. In contrast, we can observe that DD (succeeding at $m_{\text{non-ada}}(\Gamma)$ on all θ outside of a set of measure zero) requires n/Γ tests less almost everywhere for $\theta > 1/2$ on the same model. Furthermore, DD performs slightly better in sparse instances on the matching model. Moreover, no algorithm (efficient or not) can succeed at inference below the red line on any non-adaptive pooling scheme. Finally, any adaptive testing scheme fails below the black line while we presented an algorithm who succeeds at this point (up to rounding). We stress that the points on which the achievability bound of DD satisfies $\tilde{m}_{\text{DD}}(\Gamma) = \tilde{m}_{\text{non-ada}}(\Gamma) + 1$ rather than being equal correspond to the jumps in the phase diagram.

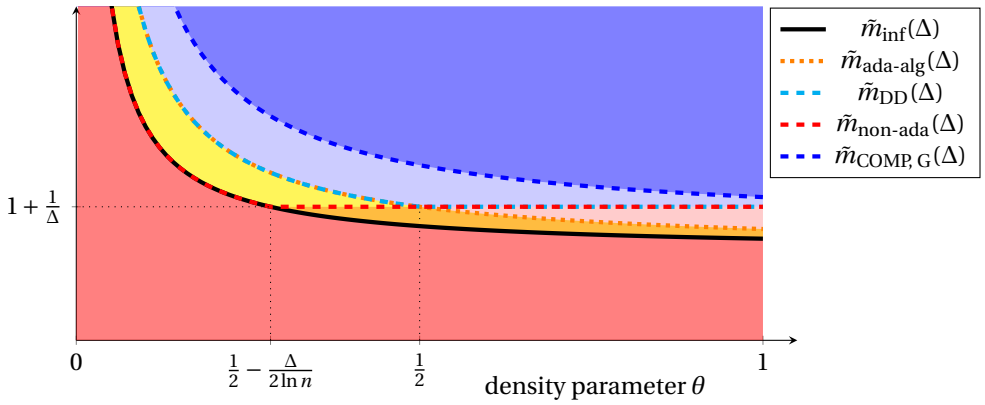

Figure 2.4.: Important phase transitions in $\Delta-$divisible hypergeometric group testing with $\Delta = O\left(\ln^{1-\delta} n\right)$ for some $\delta \in (0,1]$. The plot is with respect to the choice of parameters $\Delta = 5$, $n = 10^5$. The phase-transition lines correspond to the exponents of the actual phase transition points, thus we have $m = \Delta k^{\tilde{m}}$.

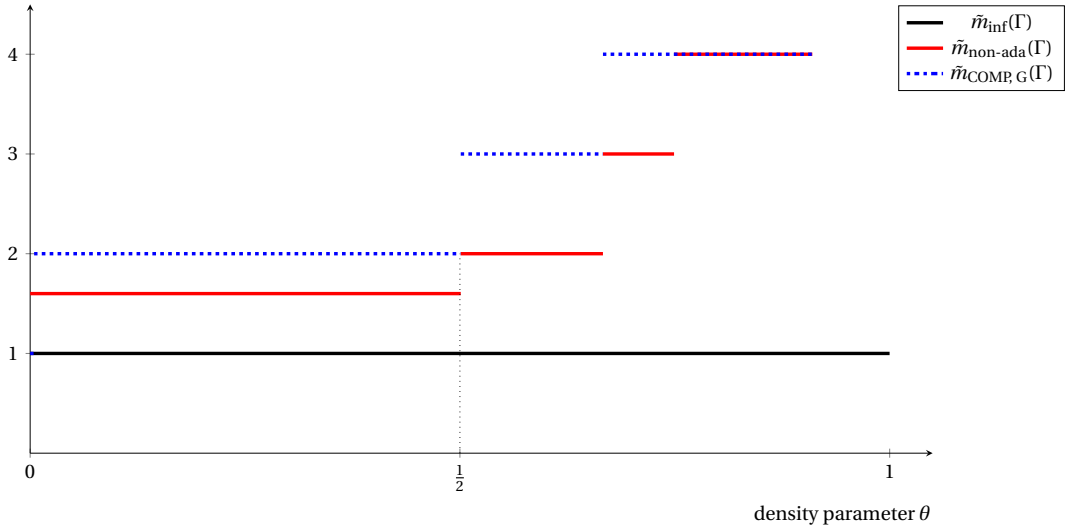

Figure 2.5.: The phase transitions in non-adaptive hypergeometric group testing under the $\Gamma-$sparse restriction for $\Gamma = \Theta(1)$. We define $m = \tilde{m}\frac{n}{\Gamma}$ and tacitly assume that $n/\Gamma \in \mathbb{Z}$. Furthermore, the plot is with respect to the choice $\Gamma = 4$.

In the next section, we will present how to achieve a formula that counts the number of solutions for a random $2-$SAT formula.

2.2. Counting solutions of a random 2-SAT formula

As already discussed in the introduction, it was a prominently posed open question, how many satisfying assignments a random $2-$SAT formula typically possesses [78]. Fortunately, the marginals obtained through Belief Propagation plugged into the Bethe functional yield a precise prediction. In the contribution

The number of satisfying assignments of random 2-SAT formulas [2]

we prove that this non-rigorous prediction is indeed correct. Before stating the main theorem, we require a bit of additional notation. Suppose we have n variables $x_1, \ldots, x_n$ taking spins in $\Omega = \{\pm 1\}$ and

a parameter $0 < d < 2$. Then we create $\boldsymbol{m} \sim \mathbf{Po}(dn/2)$ clauses and obtain a formula $\boldsymbol{\Phi}$ by choosing one uniformly at random from all possible formulas that have two distinct variables per clause. As before, $Z(\boldsymbol{\Phi})$ denotes the corresponding partition function. Let us introduce a Belief Propagation operator $\mathtt{BP}_d : \mathscr{P}((0,1)) \to \mathscr{P}((0,1))$ that maps a probability measure π onto π' as follows. Let $\boldsymbol{d}^+, \boldsymbol{d}^- \sim \mathbf{Po}(d/2)$ and $\left\{\boldsymbol{\mu}_{\pi,j}\right\}_j$ be a family of independent samples from π. Thus, $\boldsymbol{\mu}_{\pi,j}$ is a random probability measure on $(0,1)$. Then we define $\pi' = \mathtt{BP}_d(\pi)$ as the distribution of

$$\frac{\prod_{i=1}^{\boldsymbol{d}^-} \boldsymbol{\mu}_{\pi,j}}{\prod_{i=1}^{\boldsymbol{d}^-} \boldsymbol{\mu}_{\pi,j} + \prod_{i=1}^{\boldsymbol{d}^+} \boldsymbol{\mu}_{\pi,j+\boldsymbol{d}^-}}.$$

We will see in due course that this operator basically coincides with the Belief Propagation estimate of the marginals (1.1.20). We furthermore denote by $\mathtt{BP}_d^\ell$ the $\ell-$fold iteration of the $\mathtt{BP}_d$ operator. Finally, we let $\delta_x \in \mathscr{P}((0,1))$ denote the atom on x. Now we can finally state our main theorem.

Theorem 2.2.1 (Theorem 1.1 of [2])**.** *For any $0 < d < 2$ the limit $\pi_d = \lim_{\ell\to\infty} BP^\ell(\delta_{0.5})$ exists and*

$$\lim_{n\to\infty} \frac{1}{n} \ln Z(\boldsymbol{\Phi}) = \mathbb{E}\left[\ln\left(\prod_{i=1}^{\boldsymbol{d}^-} \boldsymbol{\mu}_{\pi_d,i} + \prod_{i=1}^{\boldsymbol{d}^+} \boldsymbol{\mu}_{\pi,i+\boldsymbol{d}^-}\right) - \frac{d}{2}\ln\left(1 - \boldsymbol{\mu}_{\pi_d,1}\boldsymbol{\mu}_{\pi_d,2}\right)\right]$$

in probability.

Theorem 2.2.1 implies that the free entropy density is really given through the Bethe functional (1.1.22). Let us stress that the paper's main contribution is actually a lower bound on $\ln Z(\boldsymbol{\Phi})$ as a tight upper bound could be computed by the so-called interpolation method [142]. Furthermore we stress that, of course, the result of the theorem might be hard to digest as it is far from obvious how to calculate a closed form. Nevertheless, it turns out that numerical approximations are easy to get.

How can we proof the statement of Theorem 2.2.1? To this end, let $\mu_{\boldsymbol{\Phi}} \in \mathscr{P}(\{\pm 1\}^n)$ be the Boltzmann distribution of the physical system given through the random factor graph corresponding to the random formula $\boldsymbol{\Phi}$. As we are in the zero-temperature limit, $\mu_{\boldsymbol{\Phi}}$ corresponds to the uniform distribution over all satisfying assignments (see (1.1.7)).

Supposing that the BP prediction is correct, it is not very surprising that π_d can actually be written in terms of the marginals of the Boltzmann distribution, more precisely as the random probability measure which is the $\mathbb{P}-$weak limit of

$$\pi_{\boldsymbol{\Phi}} = \frac{1}{n}\sum_{i=1}^n \delta_{\mu_{\boldsymbol{\Phi}}(\boldsymbol{\sigma}_{x_i}=1)}.$$

A core argument of the proof is that Belief Propagation is able to not only find the correct marginals but even the correct marginals for any boundary condition. Let $\boldsymbol{\sigma} \sim \mu_{\boldsymbol{\Phi}}$ be a sample from the Boltzmann distribution and τ be a second satisfying assignment. We define $\partial^{2\ell} x$ as the variables of distance exactly 2ℓ from variable x and ν^ℓ as the Belief Propagation estimation of the marginals. For the sake of the reading flow we recall the Belief Propagation messages and its estimation of the marginals from the introduction and plug in the specific setup at hand. A more detailed derivation can be found in [2]. We let $r \in \{\pm 1\}$ indicate whether variable x appears positively or negatively in clause a and let s be the corresponding sign of variable y. Then we let

$$\nu^{(\ell)}_{\boldsymbol{\Phi},a\to x}(t) = \frac{1 - \mathbf{1}\{r \neq t\}\nu^{(\ell-1)}_{\boldsymbol{\Phi},y\to a}(-s)}{1 + \nu^{(\ell-1)}_{\boldsymbol{\Phi},y\to a}(s)}, \qquad \nu^{(\ell)}_{\boldsymbol{\Phi},x\to a}(t) = \frac{\prod_{b\in\partial x\setminus\{a\}} \nu^{(\ell)}_{\boldsymbol{\Phi},b\to x}(t)}{\prod_{b\in\partial x\setminus\{a\}} \nu^{(\ell)}_{\boldsymbol{\Phi},b\to x}(1) + \prod_{b\in\partial x\setminus\{a\}} \nu^{(\ell)}_{\boldsymbol{\Phi},b\to x}(-1)}$$

be the Belief Propagation messages and define

$$\nu^{(\ell)}_{\boldsymbol{\Phi},x}(t) = \frac{\prod_{a\in\partial x} \nu^{(\ell)}_{\boldsymbol{\Phi},a\to x}(t)}{\prod_{a\in\partial x} \nu^{(\ell)}_{\boldsymbol{\Phi},a\to x}(1) + \prod_{a\in\partial x} \nu^{(\ell)}_{\boldsymbol{\Phi},a\to x}(-1)},$$

as the BP estimate of the marginals. Finally, we let

$$\mu_{\mathbf{\Phi}}\left(\cdot \mid \boldsymbol{\sigma}_{\partial^{2\ell} x_1} = \tau_{\partial^{2\ell} x_1}\right) = \mu_{\mathbf{\Phi}}\left(\cdot \mid \forall y \in \partial^{2\ell} x_1 : \boldsymbol{\sigma}_y = \tau_y\right)$$

be the Boltzmann distribution conditioned on having a specific boundary condition on variables of distance 2ℓ from x_1.

Then, if $\mathbf{\Phi}$ is satisfiable with high probability, we have [2, Theorem 1.2]

$$\lim_{\ell\to\infty} \limsup_{n\to\infty} \mathbb{E}\left[\left|\mu_{\mathbf{\Phi}}(\boldsymbol{\sigma}_{x_1} = \pm 1) - \nu^{(\ell)}_{\mathbf{\Phi},x_1}(\pm 1)\right| \mid Z(\mathbf{\Phi}) > 0\right] = 0$$

and

$$\lim_{\ell\to\infty} \limsup_{n\to\infty} \mathbb{E}\left[\max_{\tau}\left|\mu_{\mathbf{\Phi}}(\boldsymbol{\sigma}_{x_1} = 1 \mid \boldsymbol{\sigma}_{\partial^{2\ell} x_1} = \tau_{\partial^{2\ell} x_1}) - \nu^{(\ell)}_{\mathbf{\Phi},x_1}(1)\right| \mid Z(\mathbf{\Phi}) > 0\right] = 0.$$

We emphasise again that this observation is really strong and it is clearly one of the most important features of the contribution to prove this assertion. Indeed, it shows that Belief Propagation does render the correct marginals given any boundary condition. Why is this so important? It formally justifies core assumptions of the statistical physics' 1-RSB Ansatz (see Section 1.1.5.1) in the 2−SAT problem. Having the (non-rigorous) discussion of Section 1.1.5.1 in mind, let $S(\mathbf{\Phi})$ be the set of all satisfying assignments of $\mathbf{\Phi}$. Then we can observe by the triangle inequality that the Boltzmann distribution itself is a Bethe state as

$$\lim_{\ell\to\infty} \limsup_{n\to\infty} \mathbb{E}\left[\max_{\tau\in S(\mathbf{\Phi})}\left|\mu_{\mathbf{\Phi}}(\boldsymbol{\sigma}_{x_1} = 1 \mid \boldsymbol{\sigma}_{\partial^{2\ell} x_1} = \tau_{\partial^{2\ell} x_1}) - \mu_{\mathbf{\Phi}}(\boldsymbol{\sigma}_{x_1} = 1)\right| \mid Z(\mathbf{\Phi}) > 0\right] = 0.$$

Thus, as in Sections 1.3.1 and 1.1.5.1, we established being in a replica symmetric phase, hence the Boltzmann distribution does not exhibit long-range correlations but is $o(1)$−symmetric.

We recall, that $o(1)$−symmetry is defined as

$$\sum_{s,t\in\{\pm 1\}} \mathbb{E}\left[\left|\mu_{\mathbf{\Phi}}(\boldsymbol{\sigma}_{x_1} = s, \boldsymbol{\sigma}_{x_2} = t) - \mu_{\mathbf{\Phi}}(\boldsymbol{\sigma}_{x_1} = s)\cdot\mu_{\mathbf{\Phi}}(\boldsymbol{\sigma}_{x_2} = t)\right| \mid Z(\mathbf{\Phi}) > 0\right] = o(1).$$

Let us now sketch in four steps how to obtain the results, of course, as in the previous sections, this sketch only carries the main ideas of the proof and is not meant to be complete nor rigorous.

Existence of the limit. First, we need to verify that the limit π_d actually exists. This can be done rigorously by showing that the Belief Propagation operator BP is a contraction and actually converges quite fast towards a unique fixed-point. Furthermore, we can prove that the limit π_d does not only exist but satisfies a tail-bound of the form $\mathbb{E}\left[\ln^2\left(\boldsymbol{\mu}_{\pi_d}/\left(1-\boldsymbol{\mu}_{\pi_d}\right)\right)\right] < \infty$.

The Boltzmann distribution is a Bethe state. Second, we prove that BP renders the correct conditional marginals and that the marginals do not depend on the boundary conditions. We stress that the formula locally looks like a Galton-Watson tree, thus we suppose working on a tree. We make use of a feature which might be exclusive in the 2−SAT problem compared to general k−SAT, namely that we can actually construct the worst-case boundary conditions. Indeed, suppose we start at a root vertex x_0 and we want BP to output an as-high-as-possible marginal of the truth value +1 for x_0. Look at the clauses x_0 is part of, more precisely, partition them into those clauses A^- in which x_0 is negated and those clauses in which x_0 comes positively (A^+). As we want to nudge x_0 towards taking the value +1, we set all the variables being its partner in a clause of A^+ to a value that does not satisfy the clause, thus x_0 needs to be set to +1 in order to satisfy all clauses in A^+. Analogously, we set all its partner-variables in the tests of A^- to the value that satisfies the clauses already, thus x_0 does not need to satisfy them. This procedure can now be iterated until depth 2ℓ and a visualisation is given in Figure 2.6.

Of course, there are two different worst-case boundary conditions, one that nudges x_0 to +1 and one which nudges it to −1. Due to symmetry, it actually suffices to prove that the marginals of the +1-nudging configuration, call it $\boldsymbol{\sigma}^+$, coincide with the unconditional marginals. As the construction of $\boldsymbol{\sigma}^+$ clearly depends on the formula $\mathbf{\Phi}$, respectively on the Galton-Watson tree with root x_0, we have to

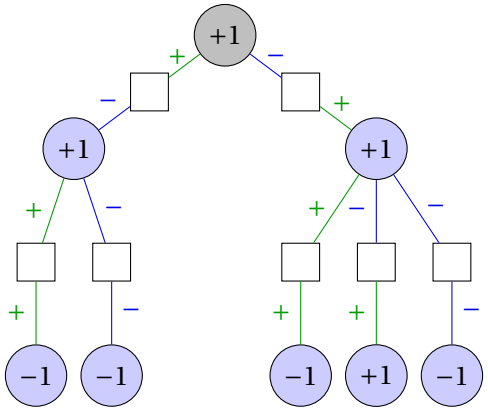

Figure 2.6.: Construction of a worst-case boundary condition. The value of the variables are chosen in the (unique) way that nudges the parent variables in the direction provided by setting x_0 (grey vertex) to +1. The graphic is modified after [2, Figure 2].

deal with delicate dependencies. More precisely, if we *really* went down the tree starting at x_0 to create $\boldsymbol{\sigma}^+$, we already would have revealed all randomness and it is far from clear how to work up this tree afterwards in order to compute the BP marginals. Fortunately, we can associate a random variable $\boldsymbol{\eta}_x \in \mathbb{R} \cup \{\pm\infty\}$ with each variable x of this tree which expresses how much x can nudge its grand-parent to the value it should take under $\boldsymbol{\sigma}^+$. These random variables $\boldsymbol{\eta}_x$ have a very comfortable (Markov-like) property which basically says that the value at vertices of distance $k \gg k'$ from x_0 does not depend on the vertices of distance at most k' from x_0.

How could such a random variable look like? To this end let $Z(\boldsymbol{T}_{x_0}, \boldsymbol{\sigma}^+, \boldsymbol{\sigma}^+_x)$ denote the number of satisfying assignments with boundary $\boldsymbol{\sigma}^+$ on the tree $\boldsymbol{T}_{x_0}$ of depth 2ℓ rooted at x_0 and truth value $\boldsymbol{\sigma}^+_x$ at x. Analogously, let $Z(\boldsymbol{T}_{x_0}, \boldsymbol{\sigma}^+, -\boldsymbol{\sigma}^+_x)$ count exactly those such assignments that have truth value $-\boldsymbol{\sigma}^+_x$ at x. Now we define $\boldsymbol{\eta}_x$ as the log-likelihood ratio between those quantities

$$\boldsymbol{\eta}_x = \ln \frac{Z(\boldsymbol{T}_{x_0}, \boldsymbol{\sigma}^+, \boldsymbol{\sigma}^+_x)}{Z(\boldsymbol{T}_{x_0}, \boldsymbol{\sigma}^+, -\boldsymbol{\sigma}^+_x)}.$$

Intuitively, as BP should calculate the marginal of x_0 with respect to a uniformly chosen satisfying assignment, the fraction between $Z(\boldsymbol{T}_{x_0}, \boldsymbol{\sigma}^+, \boldsymbol{\sigma}^+_x)$ and $Z(\boldsymbol{T}_{x_0}, \boldsymbol{\sigma}^+, -\boldsymbol{\sigma}^+_x)$ exactly expresses the extend to which x can be used to nudge its grand-parent into the correct direction. It turns out that the distribution of $\boldsymbol{\eta}_{x_0}$ can be expressed as the iterative application of a suitable operator which itself turns out to be a W_1-contraction. Analysing this operator is technically challenging, i.e. due to studying the problem at zero temperature which allows Z to decrease from an exponentially large number to zero by setting just a single variable to a certain truth value. But at least it is possible to analyse it and therefore prove that BP renders the correct marginals. The details can be found in [2, Section 5].

The Aizenman-Sims-Starr scheme. The third step is to prove that

$$\frac{1}{n}\mathbb{E}[\ln(Z(\boldsymbol{\Phi}) \vee 1)] \to \mathbb{E}\left[\ln\left(\prod_{i=1}^{\boldsymbol{d}^-} \boldsymbol{\mu}_{\pi_d,i} + \prod_{i=1}^{\boldsymbol{d}^+} \boldsymbol{\mu}_{\pi,i+\boldsymbol{d}^-}\right) - \frac{d}{2}\ln\left(1 - \boldsymbol{\mu}_{\pi_d,1}\boldsymbol{\mu}_{\pi_d,2}\right)\right] \tag{2.2.1}$$

in probability where we suppose that $\boldsymbol{d}^+, \boldsymbol{d}^- \sim \mathbf{Po}(d/2)$ denote the number of clauses in which a variable appears positively and negated respectively and where $\vee$ abbreviates the maximum. This truncation inside of the mean is actually necessary to deal with the case that, with very little probability, the formula could be unsatisfiable. The proof is done by the so-called *Aizenman-Sims-Starr* scheme [5]. To explain the main-idea, let Z_n denote the partition function of a particle system with n variables. Then we clearly find

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}[\ln Z_n] = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n-1} (\mathbb{E}[\ln Z_{i+1}] - \mathbb{E}[\ln Z_i]) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n-1} \mathbb{E}\left[\ln \frac{Z_{i+1}}{Z_i}\right].$$

Therefore, in order to calculate the partition function, we only need to understand its development when going from a system of size n to a system of size $n+1$. How can this be calculated? We need to couple a system with n particles and a corresponding system with $n+1$ particles by adding a particle and a few random clauses such that both systems follow the correct distribution. Let us discuss this in the setting of the 2−SAT problem. Suppose $\mathbf{\Phi}_n$ is a random formula with n variables and $m_n \sim \mathbf{Po}(dn/2)$ clauses while $\mathbf{\Phi}_{n+1}$ has $n+1$ variables and $m_{n+1} \sim \mathbf{Po}(d(n+1)/2)$ clauses. How could we possibly couple the formulas?

We start by a formula $\mathbf{\Phi}'$ on n variables that obtains some cavities, thus a few clauses less than it actually requires. More precisely, it has $m' \sim \mathbf{Po}(dn/2 - d/2)$ clauses. Now we obtain $\mathbf{\Phi}_n$ from $\mathbf{\Phi}'$ by adding $\mathbf{Po}(d/2)$ uniformly at random chosen clauses and $\mathbf{\Phi}_{n+1}$ by adding one variable x_{n+1} coming along with $\mathbf{Po}(d)$ clauses to $\mathbf{\Phi}'$, each of them taking a uniformly at random chosen second variable. Suppose we add from $\mathbf{\Phi}'$ to $\mathbf{\Phi}_{n+1}$ the clauses $a_1, \dots, a_{\boldsymbol{d}}$ such that the sign of x_{n+1} in clause a_i is randomly sampled as $\boldsymbol{s}_i$ from $\{-1,+1\}$. A visualisation is given in Figure 2.7.

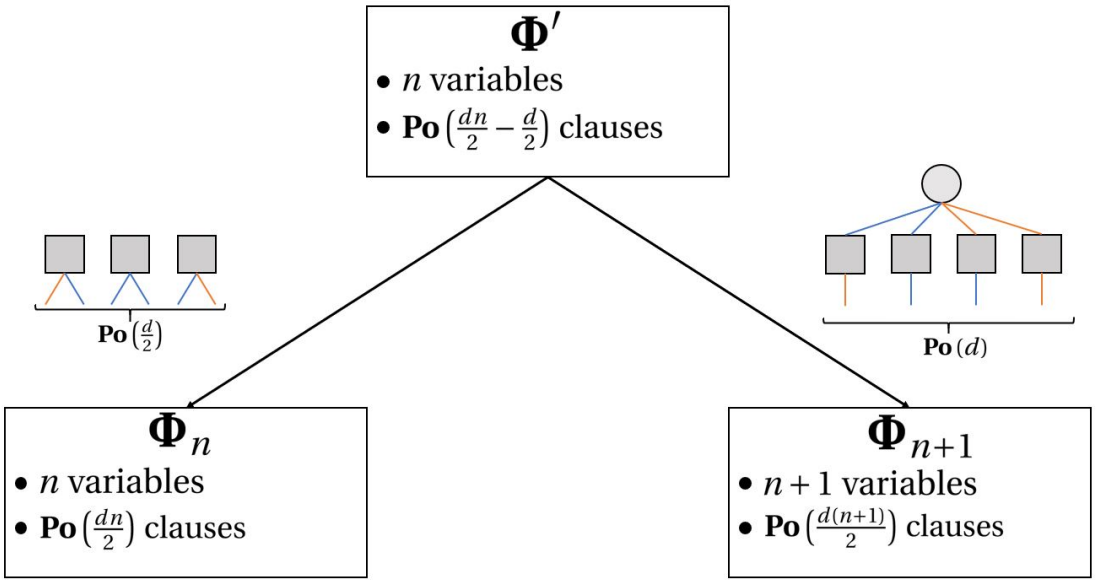

Figure 2.7.: A sketch of the Aizenmann-Sims-Starr scheme in the random 2−SAT problem. The formula $\mathbf{\Phi}'$ contains too few clauses and can be used to couple $\mathbf{\Phi}_n$ and $\mathbf{\Phi}_{n+1}$. The signs of the literals (red and blue edges) are drawn uniformly at random.

Then it suffices to prove that

$$\mathbb{E}\left[\ln \frac{Z(\mathbf{\Phi}_n)}{Z(\mathbf{\Phi}')}\right] \sim \frac{d}{2}\mathbb{E}\left[\ln\left(1-\boldsymbol{\mu}_{\pi_d,1}\boldsymbol{\mu}_{\pi_d,2}\right)\right] \quad \text{and} \quad \mathbb{E}\left[\ln \frac{Z(\mathbf{\Phi}_{n+1})}{Z(\mathbf{\Phi}')}\right] \sim \mathbb{E}\left[\ln\left(\sum_{\omega=\pm 1}\prod_{i=1}^{\boldsymbol{d}}\left(1-\mathbf{1}\{\omega \neq s_i\}\,\boldsymbol{\mu}_{\pi_d,i}\right)\right)\right]. \tag{2.2.2}$$

Indeed, suppose this is true. Then the second summand of (2.2.1) is already found as

$$\mathbb{E}\left[\ln \frac{Z(\mathbf{\Phi}_{n+1})}{Z(\mathbf{\Phi}_n)}\right] = \mathbb{E}\left[\ln \frac{Z(\mathbf{\Phi}_{n+1})}{Z(\mathbf{\Phi}')}\right] - \mathbb{E}\left[\ln \frac{Z(\mathbf{\Phi}_n)}{Z(\mathbf{\Phi}')}\right].$$

Furthermore, if we partition the newly added clauses with respect to the sign of x_{n+1} in the clause, we find that the last term can be written as

$$\mathbb{E}\left[\ln \frac{Z(\mathbf{\Phi}_n)}{Z(\mathbf{\Phi}')}\right] \sim \mathbb{E}\left[\ln\left(\sum_{\omega=\pm 1}\prod_{i=1}^{\boldsymbol{d}}\left(1-\mathbf{1}\{\omega \neq s_i\}\,\boldsymbol{\mu}_{\pi_d,i}\right)\right)\right] \sim \mathbb{E}\left[\ln\left(\prod_{i=1}^{\boldsymbol{d}^-}\left(1-\boldsymbol{\mu}_{\pi_d,i}\right) + \prod_{i=1}^{\boldsymbol{d}^+}\left(1-\boldsymbol{\mu}_{\pi_d,i+\boldsymbol{d}^-}\right)\right)\right].$$

This equals the first summand of (2.2.1) as $1-\boldsymbol{\mu}_{\pi,i}$ and $\boldsymbol{\mu}_{\pi,i}$ are equally distributed due to the symmetry of the signs of the single clauses. Thus, let as argue why (2.2.2) should intuitively hold.

Suppose we equipped $\mathbf{\Phi}'$ with a randomly chosen satisfying assignment $\boldsymbol{\tau}$, thus with a random sample from the Boltzmann distribution. Upon adding x_{n+1} with its connected clauses, there are two possibilities. Either, the truth value of x_{n+1} already satisfies a clause. In this case, the probability of extending $\boldsymbol{\tau}$ is one. If x_{n+1} does not satisfy clause a_i, then the second (randomly chosen) variable needs to satisfy it. This does happen with probability $(1-\boldsymbol{\mu}_{\pi_d,i})$. Altogether, this yields

$$\mathbb{E}\left[\ln\frac{Z(\mathbf{\Phi}_{n+1})}{Z(\mathbf{\Phi}')}\right] \sim \mathbb{E}\left[\ln\left(\sum_{\omega=\pm 1}\prod_{i=1}^{\boldsymbol{d}}\left(1-\mathbf{1}\left\{\omega\neq s_i\right\}\boldsymbol{\mu}_{\pi_d,i}\right)\right)\right].$$

On the other hand, if we add $\mathbf{Po}(d/2)$ clauses and connect them randomly to two variables and equip them with random signs, the probability that those two chosen variables both do not satisfy the clause under $\boldsymbol{\tau}$ is given by $\boldsymbol{\mu}_{\pi_d,1}\boldsymbol{\mu}_{\pi_d,2}$. Therefore, the proportion of satisfying assignments of $\mathbf{\Phi}'$ that are still satisfying for $\mathbf{\Phi}_n$ is expected to be $\left(1-\boldsymbol{\mu}_{\pi_d,1}\boldsymbol{\mu}_{\pi_d,2}\right)^{\frac{d}{2}}$ yielding

$$\mathbb{E}\left[\ln\frac{Z(\mathbf{\Phi}_n)}{Z(\mathbf{\Phi}')}\right] \sim \frac{d}{2}\mathbb{E}\left[\ln\left(1-\boldsymbol{\mu}_{\pi_d,1}\boldsymbol{\mu}_{\pi_d,2}\right)\right].$$

Of course, in a rigorous proof we need to deal with a lot of technical challenges. First, one needs to explicitly find a valid coupling $(\mathbf{\Phi}_n,\mathbf{\Phi}_{n+1})$. Second, randomly chosen variables might be almost and on average independent due to replica symmetry, but they are not completely independent. Third and most importantly, we are in the low-temperature limit. Therefore, adding one single clause could erase all satisfying assignments. A detailed proof how to handle those delicate technicalities can be found in [2, Section 6].

Concentration of the free entropy. Finally, as a last step, we need to show that $\ln(Z(\mathbf{\Phi})\vee 1)$ is concentrated around its mean. Proving this assertion seems on the first glance very challenging due to the huge fluctuations occurring in the low-temperature limit. Indeed, standard tools like the Azuma-Hoeffding inequality are doomed to fail as with little probability there might be exponentially large changes in the partition function. But fortunately, we are in the good shape that Panchenko and Talagrand [142] already proved that the partition function of the 2−SAT problem at positive temperature is concentrated applying the interpolation method of statistical physics. Let $Z_\beta(\mathbf{\Phi})$ be the partition function with respect to formula $\mathbf{\Phi}$ of the 2−SAT problem at inverse temperature β. It is clear by the definition that $Z_\beta(\mathbf{\Phi})\geq Z(\mathbf{\Phi})$ and it is known from [142] that $n^{-1}\ln Z_\beta(\mathbf{\Phi})$ does only exceed the value of the corresponding Bethe functional $\mathscr{B}_\beta$ by a factor of $(1+o(1))$ with high probability. Even more importantly, we find that the Bethe functional has a natural limit $\mathscr{B}_\infty(\pi_d)<\infty$ which coincides with the expectation of (2.2.1) and therefore, it is possible to show that $\ln Z(\mathbf{\Phi})$ does not exceed its expectation by more than $\pm\varepsilon n$ (for any $\varepsilon>0$). The details dealing with the exact functions, calculations and requirements for taking the limit can be found in [2, Section 7].

Therefore, the four described steps suffice to calculate the number of satisfying assignments of a random 2−SAT formula with high probability. In the next section we will discuss results with respect to a limit theory for discrete probability measures akin to the graph limit theory.

2.3. Limits of discrete probability measures and the cut-distance

All results of this section were obtained in

The cut metric for probability distributions [47]

and establish a consistent limit theory for discrete probability measures.

We recall from the introduction that the cut-distance of two probability measures $\mu,\nu\in\mathscr{P}(\Omega^n)$ is

defined as

$$\Delta_{\boxtimes}(\mu,\nu) = \inf_{\substack{\gamma\in\Gamma(\mu,\nu),\\ \phi\in\mathbb{S}_n}} \sup_{\substack{S\subset\Omega^n\times\Omega^n,\\ X\subset[n],\\ \omega\in\Omega}} \left| \sum_{\substack{(\sigma,\tau)\in S,\\ x\in X}} \gamma(\sigma,\tau)\left(\mathbf{1}\{\sigma_x=\omega\} - \mathbf{1}\left\{\tau_{\phi(x)}=\omega\right\}\right)\right|,$$

where $\Gamma(\mu,\nu)$ is the set of couplings of μ and ν and $\mathbb{S}_n$ is the set of permutations on $[n]$. Further, $\mathscr{L}_n(\Omega)$ is the set of equivalence classes over $\mathscr{P}(\Omega^n)$ identifying those measures with cut-distance zero.

Embedding discrete measures Moreover, we already learned about some kind of continuous embedding of configurations $\sigma\in\Omega^n$ into the space Σ_Ω of measurable functions from $[0,1)\to\mathscr{P}(\Omega)$ by

$$\hat{\sigma}:[0,1)\to\mathscr{P}(\Omega) \qquad \text{s.t.} \qquad x\mapsto\sum_{i=1}^{n}\delta_{\sigma_i}\mathbf{1}\left\{x\in\left[\frac{i-1}{n},\frac{i}{n}\right)\right\}$$

such that an associated probability measure $\mu\in\mathscr{P}(\Omega^n)$ can be expressed as

$$\hat{\mu}=\sum_{\sigma\in\Omega^n}\mu(\sigma)\delta_{\hat{\sigma}} \qquad \text{s.t.} \qquad \hat{\mu}\in\mathscr{P}(\Sigma_\Omega).$$

Finally, we recall that the cut-distance of two such measures $\mu,\nu\in\mathscr{P}(\Sigma_\Omega)$ is defined as

$$D_{\boxtimes}(\mu,\nu)=\inf_{\substack{\gamma\in\Gamma(\mu,\nu),\\ \varphi\in\mathbb{S}_{[0,1)}}}\sup_{\substack{B\subset\Sigma_\Omega^2,\\ U\subset[0,1),\\ \omega\in\Omega}}\left|\int_B\int_U\sigma_x(\omega)-\tau_{\varphi(x)}(\omega)\mathrm{d}x\mathrm{d}\gamma(\sigma,\tau)\right|$$

and identify measures with cut-distance zero. The resulting space of $\Omega-$laws is denoted by $\mathscr{L}=\mathscr{L}_\Omega$. From the definition it is immediate that $D_{\boxtimes}(\hat{\mu},\hat{\nu})\le\Delta_{\boxtimes}(\mu,\nu)$ but it is less clear if the other direction holds. We provide an argument that the embedding is, nevertheless, consistent as we have the following.

Theorem 2.3.1 (Theorem 1.2 of [47])**.** *There is a function $f:[0,1]\to[0,1]$ with $f^{-1}(0)=\{0\}$ such that for all $n\ge1$ and all $\mu,\nu\in\mathscr{P}(\Omega^n)$ we have*

$$f\left(\Delta_{\boxtimes}(\mu,\nu)\right)\le D_{\boxtimes}(\hat{\mu},\hat{\nu})\le\Delta_{\boxtimes}(\mu,\nu).$$

This assertion is far from being obvious. Indeed, equality would hold if the measure preserving bijection in the definition of $D_{\boxtimes}(\cdot,\cdot)$ was restricted to map intervals $I_i=[(i-1)/n,i/n)$ onto intervals I_j completely. But one can proof that the mass of one such interval does at least not split too much, more precisely, any reasonable such bijection maps at least mass n^{-3} from one interval into another one. Therefore, we get instantly $\Delta_{\boxtimes}(\mu,\nu)\le n^3D_{\boxtimes}(\hat{\mu},\hat{\nu})$. If n is small, this observation suffices clearly. If, on the other hand, n is sufficiently large, we partition the phase space Ω^n as well as the coordinates $[n]$ by the regularity lemma in such a way that on each little part of the partition, we can argue that the induced step functions $\hat{\mu},\hat{\nu}$ have to be very close to μ,ν under any permutation and obtain $\Delta_{\boxtimes}(\mu,\nu)\le D_{\boxtimes}(\hat{\mu},\hat{\nu})+o(1)$, establishing the theorem.

We will first show that there is a different, very elegant possibility to describe $\Omega-$laws. More precisely, such a description was already introduced in [42].

The kernel representation The kernel representation of $\Omega-$laws says briefly that it is possible to describe such a measure as something very similar to a graph limit if we define an appropriate distance. We will start discussion the other way round, thus we first define the space of kernels $\mathscr{K}$ by identifying all measurable functions $\kappa,\kappa':[0,1]^2\to\mathscr{P}(\Omega)$ with cut-distance zero where the cut-distance is defined as

$$D_{\boxtimes}(\kappa,\kappa')=\inf_{\phi,\phi'\in\mathbb{S}_{[0,1]}}\sup_{\substack{S,X\subset[0,1],\\ \omega\in\Omega}}\left|\int_S\int_X\kappa_{s,x}(\omega)-\kappa'_{\phi(s),\phi'(x)}(\omega)\mathrm{d}x\mathrm{d}s\right|.$$

Comparing $D_{\boxtimes}(\cdot,\cdot)$ with the usual cut-distance for kernels from the graph limit theory $\mathscr{D}_{\boxtimes}(\cdot,\cdot)$, we observe that (for a given ω) the major difference is that we may permute the entries on the two axis independently which clearly accounts for the fact that in the graph limit case both axes correspond to vertices of a graph while in the setting at hand, the $x-$axis represents coordinates while the $s-$axis represents configurations. Furthermore, we allow maximisation with respect to the spin ω while in the graph case we already have $\kappa_{s,x} \in \mathbb{R}$.

Now, suppose we have a kernel κ, then it induces a function $\kappa_s : [0,1] \to \mathscr{P}(\Omega) \in \Sigma_\Omega$ mapping $x \to \kappa_{s,x}$. We define $\mu^\kappa \in \mathscr{L}$ as the distribution of $\kappa_{\boldsymbol{s}}$ for an uniformly at random chosen $\boldsymbol{s} \in [0,1]$. Of course, each $\Omega-$law μ can be seen as the distribution of some $\kappa^\mu_{\boldsymbol{s}}$ as well, i.e. κ^μ coincides with $s : [0,1] \to \mathscr{P}(\Omega)$ on a set of measure $\mu(s)$.

With this picture in mind, it is actually not very surprising that those two objects are basically the same.

Theorem 2.3.2 (Theorem 1.4 of [47])**.** *The map $\mathscr{K} \to \mathscr{L}$ induced by $\kappa \mapsto \mu^\kappa$ is an isometric bijection.*

The actual proof which can be found in [47, Section 3] is quite technical but its main idea is to start with an arbitrary map $f : [0,1] \to \mathscr{P}(\Omega)$ that maps $s \mapsto f_s$ and to associate a kernel $\kappa^f : [0,1]^2 \to \mathscr{P}(\Omega)$ (by $\kappa^f_{s,x} = f_{s,x}$) with it. Analogously, the same map f can be used to transform the Lebesgue measure into a $\Omega-$law μ^f. All we need to show is that for two such functions f, g we find

$$D_{\boxtimes}(\kappa^f, \kappa^g) = D_{\boxtimes}\left(\mu^f, \mu^g\right).$$

Let us shortly come back to the connection between graph limits and our kernels. For the sake of clarity, we will refer to the graph limit kernels as graphons from now on. First, a graphon is a symmetric mapping and it turns out that we can make our kernels symmetric as follows. To this end, we define the *transpose* $\kappa^\dagger : (s,x) \mapsto \kappa_{x,s}$ of a kernel κ and observe that κ is symmetric if $\kappa = \kappa^\dagger$. For $\kappa \in \mathscr{K}$, define a family $\{\kappa^{(\omega)}\}_{\omega\in\Omega}$ of symmetric functions by

$$\kappa^{(\omega)}_{s/2,(1+x)/2} = \kappa_{s,x}(\omega), \qquad \kappa^{(\omega)}_{(1+s)/2,x/2} = \kappa_{x,s}(\omega), \qquad \kappa^{(\omega)}_{s/2,x/2} = \kappa^{(\omega)}_{(1+s)/2,(1+x)/2} = 0. \tag{2.3.1}$$

Intuitively, we squeeze the kernel from $[0,1]^2$ to $[0,\frac{1}{2}]^2$ and put it in the bottom left part of a unit square and add its transpose as the upper right part. The upper left as well as the upper right part will just equal zero. Clearly, each $\kappa^{(\omega)}$ is a symmetric function from $[0,1]^2 \to [0,1]$, thus a (bipartite) graphon where $\kappa^{(\omega)}_{s,x}$ is the corresponding edge weight.

Therefore, it is not surprising that various of the properties and results of graph limit theory carry over to the limit theory for discrete probability measures. We start by presenting one very important feature of graph limits, namely that it is possible to find a representation as an *exchangeable array*.

Exchangeable arrays We recall from the introduction that the graph limits were originally defined by the convergence of all series of homomorphism densities which basically expresses the density of how often which subgraph is present in the sequence of graphs. We will see that a similar equivalence is correct in the case of convergence of a sequence of probability measures $(\mu_n)_n$ to an $\Omega-$law μ with specific $\Omega^{n\times n}$-matrices replacing subgraphs.

To this end, we call a probability distribution $\Xi \in \mathscr{P}\left(\Omega^{\mathbb{N}\times\mathbb{N}}\right)$ *exchangeable* if the distribution of $\boldsymbol{X}^\Xi(i,j)$ coincides with the distribution of $\boldsymbol{X}^\Xi(\varphi(i),\psi(j))$ for any $\varphi,\psi \in \mathbb{S}_n$ where $i,j \in [n]$ and $\boldsymbol{X}^\Xi$ is a two-dimensional infinite array over Ω sampled from Ξ. The space of such exchangeable arrays has nice properties, for instance, it is compact and separable if equipped with the weak topology (Tychonoff's theorem). It is easy to generate such an infinite array from a kernel. Indeed, sample $\boldsymbol{s}_1, \boldsymbol{x}_1, \boldsymbol{s}_2, \boldsymbol{x}_2, \ldots \in [0,1]$ uniformly at random and independently and create an array $\boldsymbol{X}^\kappa$ such that each entry $\boldsymbol{X}^\kappa(i,j)$ is just an independent sample from $\kappa_{\boldsymbol{s}_i,\boldsymbol{x}_j} \in \mathscr{P}(\Omega)$. For the sake of brevity, we denote by $\boldsymbol{X}^\mu$ the infinite array obtained from κ^μ if κ^μ is the corresponding kernel to the $\Omega-$law μ.

Of course, if $\pi \in \mathscr{P}(\mathscr{K})$ is a distribution on such kernels, the same procedure induces a distribution Ξ^π on infinite arrays by first drawing $\boldsymbol{\kappa}$ from π and then creating $\boldsymbol{X}^{\boldsymbol{\kappa}}$. It turns out that this operation

is indeed a homeomorphism. If both probability spaces, $\mathscr{P}\left(\Omega^{\mathbb{N}\times\mathbb{N}}\right)$ and $\mathscr{P}(\mathscr{K})$ are equipped with the weak topology[1], we have the following theorem.

Theorem 2.3.3 (Theorem 1.8 of [47]). *The map $\pi \mapsto \Xi^\pi$ is a homeomorphism.*

This theorem is kind of an extension of a result in graph limit theory. Indeed, in the special case $\Omega = \{0,1\}$ it boils down to the the directed graph version of [63, Theorem 5.3]. Now we can see that a similar principle as the subgraph count is really an important feature in the theory of convergence of probability measures. Suppose that $(\mu_N)_{N\geq 1}$ is a sequence of Ω-laws that converges to $\mu \in \mathscr{L}$. Then we find an exchangeable array $\boldsymbol{X}^{\mu_N}$ such that for all $n \geq 1$ and all matrices $A \in \Omega^{n\times n}$ we have that

$$\lim_{N\to\infty} \mathbb{P}\left[\forall i,j \in [n] : \boldsymbol{X}^{\mu_N}(i,j) = A_{i,j}\right] = \mathbb{P}\left[\forall i,j \in [n] : \boldsymbol{X}^{\mu}(i,j) = A_{i,j}\right]. \tag{2.3.2}$$

If on the other hand (2.3.2) holds for all $n \geq 1$ and all $A \in \Omega^{n\times n}$, then the theorem implies

$$\lim_{N\to\infty} D_{\boxtimes}(\mu_N, \mu) = 0.$$

The existence of such a representation as an exchangeable array for an $\Omega-$law enables us to obtain a very basic result of graph limit theory in the context of $\Omega-$laws very elegantly: the sampling lemma.

Sampling from an $\Omega-$law A further feature of graph limit theory is the sampling operation. Given a large enough random graph obtained from a graphon by sampling, this finite graph will be very close to the original graphon under the cut-distance. We will use the possibility to express $\Omega-$laws as exchangeable infinite arrays to obtain a similar result for discrete probability measures. More precisely, given an $\Omega-$law μ and its representation as an exchangeable array $\boldsymbol{X}^\mu$, we define μ_n as the empirical distribution of the rows of the upper $n\times n-$submatrix of $(\boldsymbol{X}^\mu)$, thus formally

$$\mu_n(\sigma) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\left\{\forall j \in [n] : \boldsymbol{X}^\mu(i,j) = \sigma_j\right\} \qquad (\sigma \in \Omega^n).$$

As $\boldsymbol{X}^\mu$ is clearly dependent on the random coordinates $(\boldsymbol{s}_i, \boldsymbol{x}_j)_{i,j\geq 1}$, the obtained measure μ_n is a random probability distribution. We obtain the following *sampling lemma*.

Theorem 2.3.4 (Theorem 1.9 of [47]). *There is $c > 0$ such that for all $n > 1$ and all $\mu \in \mathscr{L}$ we have $\mathbb{E}\left[D_{\boxtimes}(\mu,\mu_n)\right] \leq c/\sqrt{\ln n}$.*

It turns out that this dependence on n is actually best possible (besides a possibly optimised choice of the constant c).

Theorem 2.3.5 (Theorem 1.10 of [47]). *There is a constant $d > 0$ such that for any $\varepsilon > 0$ we find some $\Omega-$law μ such that $D_{\boxtimes}(\mu,\nu) \geq \varepsilon$ for every $\nu \in \mathscr{L}$ which is supported on at most $\exp(d/\varepsilon^2)$ configurations.*

The proofs of Theorems 2.3.4 and 2.3.5 make extensive use of the close connection to graph limit theory. The sampling lemma itself is proven analogously as its corresponding form in the case of graph limits (see, for instance, [127, Section 10]) and can be found in [47, Section 3]. While the proof idea is similar, technical challenges appear due to the missing symmetry of the kernels. Let κ be a kernel, then we begin by obtaining a (finite) kernel κ_n as the matrix $(\kappa_{\boldsymbol{s}_i,\boldsymbol{x}_j})_{i,j}$ which is obtained by sampling $\boldsymbol{x}_1, \boldsymbol{s}_1, \ldots, \boldsymbol{x}_n, \boldsymbol{s}_n \in [0,1]$ u.a.r.. Furthermore, $\hat{\kappa}_n$ is the $n\times n$ upper left sub-matrix of $\boldsymbol{X}^\kappa$. We first show that, with sufficiently high probability, the cut-distance between two kernels κ, κ' is sufficiently well approximated by the cut-distance between the corresponding finite kernels κ_n, κ'_n which follows from a carefully applied result from the graph limit theory. Afterwards, we use a standard large deviation bound to prove that $\mathbb{E}[D_{\boxtimes}(\kappa_n, \hat{\kappa}_n)] = o(1)$. To obtain the result itself, we require the the regularity lemma. More precisely, we find that the step-kernel $\hat{\kappa}$ coincides with the step-kernel guaranteed by the regularity

[1] We define the weak topology on a set with respect to a family of functions as the coarsest topology under which all those functions are continuous.

lemma, thus it is close to κ under the cut-distance and therefore κ_n and $\tilde{\kappa}_n$ are close as well. Finally, it suffices to see that the two step-kernels $\hat{\kappa}$ and $\hat{\kappa}_n$ are close to each other which can be done by comparing them on each step.

We furthermore want to show that the dependency on n in the sampling lemma is roughly optimal. In this case apply a result from graph theory directly. It is known that there are graphons W_G such that each partition P of the vertex set which induces a step-graphon satisfying $\mathscr{D}_{\boxtimes}(W_G, W_{G^P}) < \varepsilon$ (c.f. Section 1.3.1.2) has to consist of at least $\exp\left(\Theta(\varepsilon^{-2})\right)$ parts [56, Theorem 7.1]. Now, given such a graphon W_G, we create a kernel κ^G by (2.3.1) and apply the sampling lemma. If the assertion of Theorem 2.3.5 was false, we would obtain a step-kernel κ_n^G on fewer steps satisfying $D_{\boxtimes}(\kappa^G, \kappa_n^G) < \varepsilon/2$. But as $\Omega = \{0,1\}$, we have

$$\mathscr{D}_{\boxtimes}(\kappa^{G,1}, \kappa_n^{G,1}) \leq 2D_{\boxtimes}(\kappa^G, \kappa_n^G) < \varepsilon$$

which is a contradiction to the aforementioned result from graph limit theory [56, Theorem 7.1].

In the next paragraph we will discuss the pinning operation (c.f. Section 1.3.1.4) in the general context of $\Omega-$laws.

Pinning Recall from the introduction that the pinning operation says roughly that conditioned on pinning a few coordinates to specific values each probability measure over Ω^n becomes extremal. For the convenience of the reader, we recall that for $\mu \in \mathscr{P}(\Omega^n)$ we denote by

$$\bar{\mu}(\sigma) = \prod_{i=1}^{n} \mu_i(\sigma_i)$$

the corresponding product measure on the same marginals. Furthermore, we call a measure $\varepsilon-$*extremal* if we have $\Delta_{\boxtimes}(\mu, \bar{\mu}) < \varepsilon$.

It is possible to generalise this notion to $\Omega-$laws. Indeed, if $\mu \in \mathscr{L}$ is an $\Omega-$law, we define $\bar{\mu} \in \mathscr{L}$ as the generalised product measure on the same marginals, thus it is the atom concentrated on

$$[0,1] \to \mathscr{P}(\Omega), \qquad x \mapsto \int_{\Sigma_\Omega} \sigma_x \mathrm{d}\mu(\sigma). \tag{2.3.3}$$

We find clearly that $D_{\boxtimes}(\bar{\mu}, \bar{\nu}) = 0$ whenever $D_{\boxtimes}(\mu, \nu) = 0$, therefore, (2.3.3) induces a mapping from the space of $\Omega-$laws into itself by $\mu \mapsto \bar{\mu}$. It turns out that those generalised product measures carry a lot of information about the actual $\Omega-$laws.

Theorem 2.3.6 (Theorem 1.11 of [47])**.** *Let μ, ν be $\Omega-$laws and $\bar{\mu}, \bar{\nu}$ the corresponding generalised product measures given via* (2.3.3)*. Then we have*

$$D_{\boxtimes}(\bar{\mu}, \bar{\nu}) \leq D_{\boxtimes}(\mu, \nu) \quad \textit{and} \quad D_{\boxtimes}(\bar{\mu}, \bar{\nu}) \leq \max_{\omega \in \Omega} \int_0^1 \left| \int_{\Sigma_\Omega} \sigma_x \mathrm{d}\mu(\sigma) - \int_{\Sigma_\Omega} \sigma_x \mathrm{d}\nu(\sigma) \right| \mathrm{d}x \leq 2D_{\boxtimes}(\bar{\mu}, \bar{\nu}). \tag{2.3.4}$$

Furthermore, the set of extremal laws is a closed subset of all $\Omega-$laws.

The proof of this lemma follows from a fairly short technical calculation whose most important idea it is to partition the space of coordinates $[0,1]$ into $X^+(\omega)$ and $X^-(\omega)$ such that

$$X^+(\omega) = \left\{ x \in [0,1] : \int_{\Sigma_\Omega} \sigma_x(\omega) \mathrm{d}\mu(\sigma) - \int_{\Sigma_\Omega} \sigma_x(\omega) \mathrm{d}\nu(\sigma) \geq 0 \right\}.$$

This helps to cope with the possible cancelling out of contributions with different signs in the definition of the cut-distance.

Once we introduced $\varepsilon-$extremity also in the limit case of $\Omega-$laws, it is a natural question whether the pinning operation of Coja-Oghlan et al. [49] generalises and yields $\varepsilon-$extremal $\Omega-$laws. To this end, let us define the pinning operation for an $\Omega-$law μ. Given some integer $\theta \geq 1$ and θ coordinates $x_1, \ldots, x_\theta \in$

$[0,1]$ as well as a configuration $\tau \in \Omega^\theta$, we define a normalising constant

$$Z = Z_\mu(\tau, x_1, \dots, x_\theta) = \int_{\Sigma_\Omega} \prod_{i=1}^{\theta} \sigma_{x_i}(\tau_i) \mathrm{d}\mu(\sigma).$$

Furthermore, if $Z > 0$, let $\mu_{\tau\downarrow x_1,\dots,x_\theta}$ be defined as

$$\mathrm{d}\mu_{\tau\downarrow x_1,\dots,x_\theta}(\sigma) = \frac{1}{Z} \prod_{i=1}^{\theta} \sigma_{x_i}(\tau_i) \mathrm{d}\mu(\sigma) \tag{2.3.5}$$

and $\mu_{\tau\downarrow x_1,\dots,x_\theta} = \mu$ if $Z = 0$.

This pinning operation looks very similar to the discrete version and indeed, the discrete version is contained as a special case in which each of the factors on the r.h.s. of (2.3.5) is either one or zero.

As in the discrete case, it becomes interesting if we chose $\boldsymbol{x}_1, \dots, \boldsymbol{x}_\theta$ randomly. More specifically, for a given $\theta \geq 1$,

(i) let $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots \in [0,1]$ be u.a.r. and mutually independent,

(ii) draw $\boldsymbol{\tau} \in \Sigma_\Omega$ from the distribution μ,

(iii) pick a reference configuration $\hat{\boldsymbol{\tau}}$ from $\boldsymbol{\tau}_{\boldsymbol{x}_1} \otimes \cdots \otimes \boldsymbol{\tau}_{\boldsymbol{x}_\theta} \in \mathscr{P}\left(\Omega^\theta\right)$,

(iv) obtain $\mu_{\hat{\boldsymbol{\tau}}\downarrow\theta} = \mu_{\hat{\boldsymbol{\tau}}\downarrow\hat{\boldsymbol{x}}_1,\dots,\hat{\boldsymbol{x}}_\theta}$ via (2.3.5).

By the choice of $\boldsymbol{\tau}$ being sampled from μ, we clearly find that $\boldsymbol{Z}_\mu(\hat{\boldsymbol{\tau}}) > 0$ almost surely. Finally, we let

$$\mu_{\downarrow\theta} = \mathbb{E}\left[\overline{\mu_{\hat{\boldsymbol{\tau}}\downarrow\theta}} \mid \boldsymbol{x}_1, \dots, \boldsymbol{x}_\theta\right] \in \mathscr{L}.$$

Intuitively spoken, $\mu_{\downarrow\theta}$ is a weighted probability measure such that each configuration's probability is weighted according to the probability of its reference configuration.

The pinning lemma itself in the continuous case does guarantee that we find with high probability an approximation of an Ω−law μ supported on few ε−extremal Ω−laws. More precisely, it reads as follows.

Theorem 2.3.7 (Theorem 1.12 of [47])**.** *Given $\varepsilon \in (0,1)$ and an Ω−law μ, draw $0 \leq \boldsymbol{\theta} = \boldsymbol{\theta}(\varepsilon) \leq 64\varepsilon^{-8}\ln|\Omega|$ uniformly and independently of everything else. Then we find with probability at least $1-\varepsilon$ that $\mu_{\hat{\boldsymbol{\tau}}\downarrow\boldsymbol{\theta}}$ is ε−extremal and $\mathbb{E}[D_\boxtimes(\mu, \mu_{\downarrow\boldsymbol{\theta}})] < \varepsilon$.*

While the proof of the pinning lemma is the technical main contribution of [47] and relies on delicate and technically challenging results, it is surprisingly easy to give a high level sketch. In the following sketch, let $\delta, \delta', \dots$ be small and suitable chosen constants.

First, we need to verify that the pinning operation is continuous with respect to the cut-distance. Second, given an Ω−law μ, we use the sampling lemma to obtain a discrete probability measure $\nu \in \mathscr{P}(\Omega^n)$ such that $D_\boxtimes(\mu, \nu) < \delta$. Now we apply the pinning operation to μ as well as to ν and obtain (written a bit shortly) $\mu_{\downarrow n}$ and $\nu_{\downarrow n}$. By the continuity of the pinning operation we have

$$D_\boxtimes(\mu_{\downarrow n}, \nu_{\downarrow n}) < \delta'$$

and as the pinning operation reduces to the discrete pinning regarding ν, the discrete pinning lemma guarantees that

$$\mathbb{E}\left[\Delta_\boxtimes\left(\overline{\nu_{\downarrow\boldsymbol{\theta}}}, \nu_{\downarrow\boldsymbol{\theta}}\right)\right] < \delta''.$$

Now, as the embedding of discrete probability measures into the Ω−laws respects the cut-distance (Theorem 2.3.1), we directly get

$$\mathbb{E}\left[D_\boxtimes\left(\overline{\nu_{\downarrow\boldsymbol{\theta}}}, \nu_{\downarrow\boldsymbol{\theta}}\right)\right] < \delta'''.$$

Now the results on how far the generalised product measures are from the actual measures in the cut-distance (Theorem 2.3.6) and the triangle inequality give

$$D_\boxtimes\left(\overline{\mu_{\downarrow\boldsymbol{\theta}}}, \mu_{\downarrow\boldsymbol{\theta}}\right) < \delta'''' + D_\boxtimes\left(\overline{\nu_{\downarrow\boldsymbol{\theta}}}, \nu_{\downarrow\boldsymbol{\theta}}\right).$$

Therefore, it is possible to reduce the continuous pinning to the discrete version and the assertion follows from Markov's inequality.

While the pinning operation stands at the heart of the contribution we obtained two more results. The first of those two is with respect to *multi-overlaps*.

Overlaps We will first describe two very basic operations on probability measures that turn out to be continuous with respect to the cut-distance. For the sake of simplicity, we will describe their meaning for discrete measures over Ω^n and will just shortly state the corresponding operation on $\Omega-$laws.

The first such operation is the construction of a product measure. More precisely, suppose we have two probability measures $\mu, \nu \in \mathscr{P}(\Omega^n)$. Then their product $\mu \otimes \nu$ is a probability distribution on $(\Omega \times \Omega)^n$ such that $(\mu \otimes \nu)(\sigma, \tau) = \mu(\sigma)\nu(\tau)$.

Analogously, we could define a tensor $\mu \otimes \nu$ such that for $\sigma_1, \tau_1, \ldots, \sigma_n, \tau_n \in \Omega$ we obtain

$$\mu \otimes \nu \left(\begin{pmatrix} \sigma_1 \\ \tau_1 \end{pmatrix}, \ldots, \begin{pmatrix} \sigma_n \\ \tau_n \end{pmatrix} \right) = \mu(\sigma_1, \ldots, \sigma_n)\nu(\tau_1, \ldots, \tau_n).$$

Clearly, both constructions are equivalent but the – perhaps less intuitive – tensor variant naturally extends to $\Omega-$laws [49]. More precisely, it is convenient to identify the $\Omega-$laws with their corresponding kernel. To be more precise, suppose that $\Lambda : [0,1] \to [0,1] \times [0,1]$, $x \mapsto (\Lambda_1(x), \Lambda_2(x))$ is any measurable bijection mapping the Lebesgue measure λ_1 on $[0,1]$ onto the Lebesgue measure λ_2 on $[0,1]^2$. Furthermore, Λ needs to satisfy that Λ^{-1} maps the Lebesgue measure on $[0,1]^2$ to the Lebesgue measure on $[0,1]$. Such transformations of the Lebesgue measure clearly exist, for instance, a proof is given by [102, Theorem A.7]. If now κ and κ' are the kernel representations of two $\Omega-$laws, we define their generalised product as

$$\kappa \otimes \kappa' : [0,1]^2 \to \mathscr{P}(\Omega^2), \qquad (s,x) \in [0,1] \times [0,1] \mapsto \kappa_{\Lambda_1(s),x} \otimes \kappa'_{\Lambda_2(s),x} \in \mathscr{P}(\Omega^2).$$

Thus, by re-translating the kernels into $\Omega-$laws we immediately find that, given Ω-laws μ, ν, above's procedure yields an Ω^2-law $\mu \otimes \nu$. As already pointed out, this operation is continuous.

Theorem 2.3.8 (Theorem 1.15 of [47])**.** *The map $(\mu, \nu) \in \mathscr{L}(\Omega) \mapsto \mu \otimes \nu \in \mathscr{L}(\Omega^2)$ is continuous with respect to the cut-distance.*

The second basic operation resembles the procedure which is used to generate an $n \times n$ (rank one) matrix over Ω from two vectors $\sigma, \tau \in \Omega^n$. More precisely, if $\sigma, \tau \in \Omega^n$ are two vectors we define $\sigma \oplus \tau \in (\Omega^2)^{n \times n}$ as the $n \times n$-matrix with entries $(\sigma \oplus \tau)_{ij} = (\sigma_i, \tau_j)$ for all $i, j \in [n]$. Given two probability distributions $\mu, \nu \in \mathscr{P}(\Omega^n)$, we define $\mu \oplus \nu$ as follows. First, sample $\boldsymbol{\sigma} \sim \mu$ and $\boldsymbol{\tau} \sim \nu$. Second, obtain $\mu \oplus \nu$ as $\boldsymbol{\sigma} \oplus \boldsymbol{\tau}$.

Of course, this operation can be generalised to $\Omega-$laws directly and interestingly, it can be expressed by the generalised product measure $\otimes$. Suppose that κ, κ' are kernel representations of $\Omega-$laws and define

$$\kappa \oplus \kappa' : [0,1]^2 \to \mathscr{P}(\Omega^2), \qquad (s,x) \mapsto \kappa_{s,\Lambda_1(x)} \otimes \kappa_{s,\Lambda_2(x)}.$$

As the $\oplus-$operation boils down to the $\otimes-$operation, it is not hard to guess that it is continuous as well.

Theorem 2.3.9. *The map $\mathscr{L}(\Omega) \to \mathscr{L}(\Omega^2)$, $(\mu, \nu) \mapsto \mu \oplus \nu$ is continuous with respect to the cut-distance.*

This two (very fundamental) operations are everything we need to express the quantity of multi-overlaps quite elegantly. Recall from the introduction that the overlap matrix of two configurations σ, τ was denoted by $\langle \sigma, \tau \rangle$ and expresses on how many particles the spins of σ and τ coincide. This can be naturally extended with regard to two aspects. First, instead of comparing discrete configurations we can calculate the overlap of two generalised configurations $\sigma, \tau \in \Sigma_\Omega$. Second, we can compute the overlap of more than two (generalised) configurations. Thus, let $\sigma_1, \ldots, \sigma_n \in \Sigma_\Omega$ and $\omega_1, \ldots, \omega_n \in \Omega$ and

define

$$R_{\omega_1,\dots,\omega_n}(\sigma_1,\dots,\sigma_n) = \int_0^1 \prod_{i=1}^n \sigma_{i,x}(\omega_i)\mathrm{d}x.$$

Given an $\Omega-$law μ and an integer $\ell \geq 1$ we calculate a similar quantity by averaging over the choice of the configuration through μ, thus

$$R_{\ell,\omega_1,\dots,\omega_n}(\mu) = \int_{\Sigma_\Omega} \cdots \int_{\Sigma_\Omega} R_{\omega_1,\dots,\omega_n}(\sigma_1,\dots,\sigma_n)^\ell \,\mathrm{d}\mu(\sigma_1)\cdots\mathrm{d}\mu(\sigma_n).$$

This finally enables us to define the *multi-overlap* of an $\Omega-$law μ as the array

$$R_{\ell,n}(\mu) = (R_{\ell,\omega_1,\dots,\omega_n}(\mu))_{\omega_1,\dots,\omega_n\in\Omega}.$$

As those multi-overlaps are built by concatenations of $\otimes$ and $\oplus$, it is no surprise that the following holds.

Corollary 2.3.10 (Corollary 1.15 of [47])**.** *The functions $\mu \in \mathscr{L} \mapsto R_{\ell,n}(\mu)$ with $\ell, n \geq 1$ are continuous.*

Let us now come to the last result obtained in [47] which can be seen as one of the most fundamental properties of a consistent limit theory.

Compactness of the space of $\Omega-$laws While [49] already provided an argument that the space $(\mathscr{L}, D_{\boxtimes})$ is a compact metric space by comparing it to the space of decorated graph limits, we give a self-contained argument which is based on previous results within the theory of limits of discrete probability measures.

We will first analyse the space of kernels $\mathscr{K}$. To this end, we define three different variants of the cut-distance. More precisely, we define

$$D_{\boxtimes}(\kappa,\kappa') = \inf_{\phi,\psi\in\mathbb{S}_{[0,1]}} \sup_{\substack{S,X\subset[0,1]\\ \omega\in\Omega}} \left|\int_S\int_X \left(\kappa_{s,x}(\omega) - \kappa'_{\phi(s),\psi(x)}(\omega)\right)\mathrm{d}x\mathrm{d}s\right|,$$

$$D_{\boxslash}(\kappa,\kappa') = \inf_{\phi\in\mathbb{S}_{[0,1]}} \sup_{\substack{S,X\subset[0,1]\\ \omega\in\Omega}} \left|\int_S\int_X \left(\kappa_{s,x}(\omega) - \kappa'_{\phi(s),x}(\omega)\right)\mathrm{d}x\mathrm{d}s\right|,$$

$$D_{\square}(\kappa,\kappa') = \sup_{\substack{S,X\subset[0,1]\\ \omega\in\Omega}} \left|\int_S\int_X \left(\kappa_{s,x}(\omega) - \kappa'_{s,x}(\omega)\right)\mathrm{d}x\mathrm{d}s\right|.$$

Therefore, the strongest version of the cut-distance is $D_{\square}(\cdot,\cdot)$ which does not allow for any measure preserving transformations while the slightly weaker variant $D_{\boxslash}(\cdot,\cdot)$ allows to permute the coordinates corresponding to the generalised configurations. Finally, $D_{\boxtimes}(\cdot,\cdot)$ which we previously studied, is the weakest form of the cut-distance which allows measure preserving transformations in both dimensions.

In a first step we verify that the set of kernels is complete with respect to $D_{\square}(\cdot,\cdot)$ which requires a delicate technical analysis employing the Riesz representation theorem as well as the Radon-Nikodym theorem. Then, the second and third step adapt the well known fact that the quotient space with respect to a linear subspace of each complete metric space is complete itself. Indeed, $(\mathscr{K}, D_{\boxslash})$ is a quotient space of $(\mathscr{K}, D_{\square})$ by identifying kernels κ,κ' with $D_{\boxslash}(\kappa,\kappa') = 0$. Analogously, $(\mathscr{K}, D_{\boxtimes})$ is a quotient space of $(\mathscr{K}, D_{\boxslash})$. By the kernel representation (Theorem 2.3.2) this immediately implies that the space of $\Omega-$laws is complete with respect to $D_{\boxtimes}(\cdot,\cdot)$ as well.

We are left to show that the space of kernels is separable. Fortunately, by the regularity lemma we know that the $\Omega-$laws with finite support are actually a dense subset of $\mathscr{L}$. Therefore, it suffices to prove that Σ_Ω is separable. But the latter is clear as the set of all finite linear combinations of indicator functions $x \mapsto \mathbf{1}\{a < x < b\}$ with $a, b \in \mathbb{Q}$ is a dense subset of all measurable continuous functions from $[0,1] \to \mathbb{R}$. Again employing the kernel representation, we find that this separability carries over to the space of kernels.

Up to now we found that $\mathscr{L}$ is complete and separable but of course this does not necessarily imply compactness. We use the representation of $\pi \in \mathscr{P}(\mathscr{L})$ as a distribution over exchangeable infinite arrays in $\mathscr{E} \subset \mathscr{P}\left(\Omega^{\mathbb{N}\times\mathbb{N}}\right)$ to prove compactness. As the latter space is known to be compact it suffices to find a continuous mapping from $\mathscr{E}$ to $\mathscr{P}(\mathscr{L})$.

Let us sketch how such a mapping could look like. Let $\xi \in \mathscr{E}$ be a distribution over infinite arrays, then we define a mapping $\xi \mapsto \rho^{\xi,n}$ as follows. First, we draw an array $\boldsymbol{X}^\xi$ from ξ. Subsequently, we define $\boldsymbol{\mu}^{\xi,n} \in \mathscr{P}(\Omega^n)$ as the empirical distribution of the rows of the top-left $n \times n$ submatrix of $\boldsymbol{X}^\xi$ and identify it with its embedding into $\mathscr{L}$. Finally, $\rho^{\xi,n}$ is the distribution of $\boldsymbol{\mu}^{\xi,n}$ w.r.t. the choice of $\boldsymbol{X}^\xi$. It turns out that we have for every $\xi \in \mathscr{E}$ that $\rho^\xi = \lim_{n\to\infty} \rho^{\xi,n}$ exists and $\xi \mapsto \rho^\xi$ is continuous [47, Lemma 3.13].

Into the other direction, it is also possible to associate a distribution $\xi^\mu \in \mathscr{E}$ with an $\Omega-$law μ. Indeed, as discussed in the paragraph about the representation as exchangeable arrays, we can simply define ξ^μ as the distribution of $\boldsymbol{X}^\mu$. This mapping actually shows that the space of $\Omega-$laws can be embedded into $\mathscr{E}$ as ρ^{ξ^μ} turns out to be the atom on μ (e.g. $\delta_\mu \in \mathscr{P}(\mathscr{L})$). Again, the corresponding mapping $\mu \mapsto \xi^\mu$ is continuous which directly implies that the mapping $\mathscr{E} \to \mathscr{P}(\mathscr{L})$, $\xi \mapsto \rho^\xi$ is surjective [47, Lemma 3.14, Corollary 3.15]. Altogether, this implies the compactness of the space of $\Omega-$laws.

2.3.1. Summary: the cut-distance for probability measures

We established a self-contained and consistent limit theory for discrete probability measures on Ω^n akin to the graph limit theory. We showed that the limit space $\mathscr{L} = \mathscr{L}(\Omega)$ of $\Omega-$laws is a compact space closely related to the space of graphons.

We furthermore analysed different representations for an $\Omega-$law μ. First, it is possible to define a corresponding kernel, thus a function from the unit square into the probability measures over Ω. Second, each μ is in correspondence with a (random) exchangeable two-dimensional infinite array $\Xi^\mu \in \Omega^{\mathbb{N}\times\mathbb{N}}$.

The latter representation enabled us to prove a sampling lemma comparable to the sampling lemma of graph limit theory. Moreover, we extended the pinning operation of [49] to the more general case of $\Omega-$laws. Finally, we proved that the operation of obtaining multi-overlaps is continuous with respect to the cut-distance which might be an important step towards rigorising some statistical physics' predictions as the (multi-)overlap is a frequently studied observable.

Let us in the next section leave the world of statistical physics once more and discuss results with respect to perturbed random graphs.

2.4. Spanning structures in randomly perturbed sparse graphs

Recall that in the setting of randomly perturbed graphs some arbitrary (possibly deterministic) graph $\mathscr{G}_\alpha = (V_\alpha, E_\alpha)$ with minimum degree αn is given. Furthermore, we take the edges from an instance $\boldsymbol{G}$ of $\mathscr{G}(n,p)$ and examine whether there are certain spanning structures present in $\mathscr{G}_\alpha \cup \boldsymbol{G}$ with high probability. We start by stating the obtained results.

Theorem 2.4.1 (Theorems 1.1 and 1.2 of [93])**.** *Let $\mathscr{G}_\alpha$ be a graph with minimum degree αn and $\boldsymbol{G}$ an instance of $\mathscr{G}(n,\beta/n)$.*

- *If $\beta \geq -(6+o(1))\ln\alpha$, $\mathscr{G}_\alpha \cup \boldsymbol{G}$ contains a Hamilton cycle with high probability.*
- *If $\beta \geq -(4+o(1))\ln\alpha$, $\mathscr{G}_\alpha \cup \boldsymbol{G}$ contains a perfect matching with high probability.*

Let us first observe that the theorem is tight (up to a constant factor). Indeed, if $\mathscr{G}_\alpha$ is the complete bipartite graph on classes $V_\alpha, V_{1-\alpha}$ of size αn and $(1-\alpha)n$, we cannot find a perfect matching if there is an independent set of size larger than αn in $V_{1-\alpha}$. It is known that the number of isolated vertices in the random graph is $\sim n\exp\left(-\beta\right) \gg \alpha n$ if $\beta = o(-\ln\alpha)$. Therefore, one requires $\beta = \Omega(-\ln\alpha)$, and of course, if there is no perfect matching, we cannot find a Hamilton cycle either.

Furthermore, we obtained results with respect to the existence of spanning bounded degree trees. More precisely, we state some kind of a *meta-theorem* which allows – given an almost spanning structure of sufficient size in $\mathscr{G}(n,\beta/n)$ – to find the spanning structure by the edges in the deterministic graph.

Theorem 2.4.2 (Theorem 1.6 of [93]). *Let $\Delta \geq 2$ be an integer and suppose that $\alpha, \beta, \varepsilon \colon \mathbb{N} \mapsto [0,1]$ are such that $4(\Delta+1)\varepsilon < \alpha^{\Delta+1}$. Furthermore suppose that $\mathcal{G}\left(n, \beta/n\right)$ contains a given tree with maximum degree Δ on $(1-\varepsilon)n$ vertices w.h.p. and that $\mathcal{G}_\alpha$ is an arbitrary graph with minimum degree αn.*

Then any tree with maximum degree Δ on n vertices can be found in $\mathcal{G}_\alpha \cup \mathcal{G}\left(n, \beta/n\right)$ with high probability.

The theorem solely does of course not answer the question whether we find spanning trees or not. But with a recent result of Balogh et al. [19], we find the following.

Corollary 2.4.3 (Corollary 1.8 of [93]). *For $\Delta \geq 2$ there exists $C > 0$ such that for $\alpha = \alpha(n) : \mathbb{N} \mapsto (0,1)$ and $\beta = \beta(\alpha) = -C\alpha^{-(\Delta+1)} \ln \alpha$ the following holds. Any n-vertex tree T with maximum degree Δ is a subgraph of $\mathcal{G}_\alpha \cup \mathcal{G}\left(n, \beta/n\right)$ with high probability.*

How can we proof such statements? We will only give an idea of the proof with respect to the Hamilton cycle as the other theorems follow fairly similar ideas. We need two major ingredients. First, we require a long cycle in the random graph. Fortunately, such a result already exists.

Lemma 2.4.4 (Frieze [83]). *Let $0 < \beta = \beta(n) \leq \ln n$. Then the random graph $\mathcal{G}(n, \beta/n)$ contains a cycle of length at least $\left(1 - (1-o(1))\,\beta \exp\left(-\beta\right)\right) n$ with high probability.*

Second, we will use the previously explained multiple round exposure technique with the difference that we in this case reveal edges of the random graph in multiple rounds instead of infected individuals like in the group testing problem. Now we already have everything at hand to sketch the proof of (the first part of) Theorem 2.4.1.

Proof sketch of the first part of Theorem 2.4.1. Observe that for very small $\alpha = O(n^{-1/6})$ our choice of β already guarantees that the random graph is known to contain a Hamilton cycle w.h.p..

Thus, suppose that $\alpha = \omega(n^{-1/6})$. We start by revealing almost all edges of the random graph. More precisely, we reveal the edges of $\mathcal{G}(n, (\beta-1)/n)$.

Within this random graph we find by the previous lemma a path P on all but at most $\beta \exp\left(-\beta\right) n$ vertices. Say that those left-over vertices are denoted by V'. We will subsequently *absorb* all but two such vertices onto the path P using edges from the deterministic graph $\mathcal{G}_\alpha$.

To this end, let $\boldsymbol{B}(u,v)$ denote the set of vertices x that lie on P and are a neighbour of u in $\mathcal{G}_\alpha$ such that the neighbours of u on the path P are also connected to v via edges in $\mathcal{G}_\alpha$. Formally,

$$\boldsymbol{B}(u,v) = \left\{x \in \partial_{\mathcal{G}_\alpha}(u) \cap P \mid \partial_P(x) \subset \partial_{\mathcal{G}_\alpha}(v)\right\}.$$

A visualisation is given in Figure 2.8. Suppose $P = p_1 \dots p_\ell$ is the path. Then clearly, if for a vertex

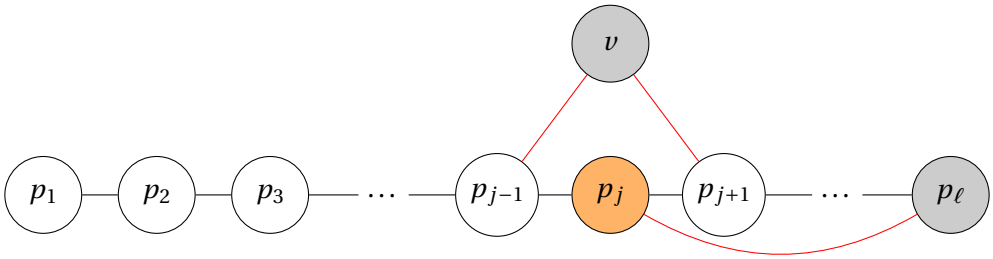

Figure 2.8.: Absorbing structure for vertex v onto the path $p_1 \dots p_\ell$ using edges from $\mathcal{G}_\alpha$ (red). The graphic is modified after [93].

$v \in V'$ there is some $p_j \in \boldsymbol{B}(p_\ell, v)$, we can create a longer path $p_1 \dots p_{j-1} v p_{j+1} \dots p_\ell p_j$ which contains v. It turns out that, up to some technicalities which guarantee that those absorbing structures do not overlap too much, we can follow this approach greedily until all but 2 vertices are absorbed. Indeed, as $\boldsymbol{B}(u,v)$ is a uniformly at random chosen set of vertices due to the randomness of $\mathcal{G}(n, (\beta-1)/n)$, we find

$$|\boldsymbol{B}(u,v)| \geq \frac{2}{5}\alpha^3 \beta \exp\left(-\beta\right) \geq \frac{\alpha^3 n}{4}$$

with high probability for the choice of β. While absorbing the left-over vertices greedily for

$$\left|V'\right| - 2 \leq \beta \exp\left(-\beta\right) n$$

rounds, we remove in every step all used vertices from all sets $\boldsymbol{B}(u,v)$ and the absorbed vertex from V'. As after $\beta \exp\left(-\beta\right) n$ rounds the size of $\boldsymbol{B}(u,v)$ only decreased by at most $\beta \exp\left(-\beta\right) n$ (for any pair of vertices u, v), we have after absorbing that

$$|\boldsymbol{B}(u,v)| \geq \frac{\alpha^3 n}{4} - \beta \exp\left(-\beta\right) n \geq \frac{\alpha^3 n}{8}$$

if $\alpha = \omega\left(n^{-1/6}\right)$. Thus, absorbing all but two vertices is indeed possible to be done greedily.

Therefore, all we need to do at this point is to close the cycle. Recall that we obtained a path $P = p_1 \dots p_{n-2}$ and are left with two additional vertices a, b. If there is an edge $p_i p_j$ between $\boldsymbol{B}(p_1, a)$ and $\boldsymbol{B}(p_{n-2}, b)$, we already found a Hamilton cycle $C = p_i p_1 \dots p_{i-1} a p_{i+1} \dots p_{j-1} p_{j+1} \dots p_{n-2} p_j p_i$. And indeed, by revealing the missing edges from $\mathcal{G}(n, 1/n)$, we find that such an edge is present with high probability. □

3. Outlook

This chapter will state open problems related with this thesis's contributions and aims for giving rise to possibly interesting further research directions within the respective fields. We start by stating such research questions with respect to the group testing problem.

3.1. Group testing

Non-adaptive group testing in the sublinear regime In the sublinear regime where the number of infected individuals scales as $k = n^\theta$, this thesis's contributions draw a somehow complete picture of non-adaptive noiseless hypergeometric group testing. While the analysis of SPIV and the proof of the universal information theoretic converse does not require exact knowledge about k but its order of magnitude suffices, it seems very likely that those results might carry over to i.i.d. models as well. Nevertheless, while SPIV is an efficient algorithm from a theoretical point of view, requiring $\ln\ln n \to \infty$ shows that the result is only of theoretical nature. Furthermore, the spatially coupled design is clearly much more complicated than the random regular model. Therefore, an intriguing question would be the following.

Question 3.1.1. *Is there an efficient algorithm succeeding at $m_{\text{non-ada}}$ on the random regular model? Furthermore, is there a deterministic graph G with $(1+\varepsilon)\, m_{\text{non-ada}}$ tests coming with an efficient algorithm for high probability recovery of $\boldsymbol{\sigma}$ from $(G, \hat{\boldsymbol{\sigma}})$?*

Besides searching for more practical algorithms, one could ask what happens when we leave the Bayes optimal setting. Suppose the student does only receive an estimate $\tilde{k}$ of k (thus, not the complete teacher's prior). If $\tilde{k} > k$, inference will still be possible by SPIV as we conduct only too many tests and each test is more likely to be negative. If, on the other hand, k exceeds the guess $\tilde{k}$, complete recovery is not possible under the studied designs. Therefore, the following question might arise.

Question 3.1.2. *Given an estimate $\tilde{k} < k$ of the number of infected individuals, how many tests suffice to infer the ground-truth completely or at least up to small errors depending on $\tilde{k}$ and k? How would an optimal design look like?*

Similarly, while we proved that $(1-o(1))-$recovery is possible at m_{inf} using the spatially coupled design with SPIV, it is also known that, information-theoretically, it is possible to achieve $(1-\gamma)-$recovery on a Bernoulli test-design with no more than m_{inf} tests. Up to our knowledge, there are only non-rigorous contributions which provide evidence that it might be possible algorithmically [99]. Further, there cannot be any design on less than $(1-\gamma)\, m_{\text{inf}}$ tests achieving $(1-\gamma)-$ recovery [152].

Question 3.1.3. *Is partial recovery of all but γk infected individuals possible algorithmically under Bernoulli group testing at the information-theoretic threshold?*

But nevertheless, this part of group testing is fairly well understood. Things are completely different for sparsity constrained group testing.

Sparsity constrained group testing As already seen in the previous chapter, the sparsity constrained group testing problem is not as well understood as the unrestricted case. A fairly natural question is if the phase diagram in the $\Delta-$divisible case does actually look similar to the one in unrestricted group testing.

Question 3.1.4. *Is there an adaptive algorithm testing each individual at most Δ times succeeding at $m_{\text{inf}}(\Delta)$ or can we proof that $m_{\text{inf}}(\Delta)$ is not tight? Furthermore, is there a spatially coupled design coming with an efficient algorithm performing as well as the binary splitting approach (or even better)?*

In the $\Gamma-$sparse case we completely understood the problem if Γ is a constant.

Question 3.1.5. *Does the analysis of the* $\Gamma = \Theta(1)$ *case extend to larger values of* Γ*? And, if so, does it strengthen existing bounds?*

Finally, one might be slightly irritated by the fact that the achievability and converse bounds in the $\Delta-$divisible case do not converge to their unrestricted counterpart.

Question 3.1.6. *How does the group testing problem behave in the critical regime* $\omega(\ln^{1-\delta} n) = \Delta = o(\ln n)$ *and what happens at the phase transition* $\Delta \to \ln n$*?*

Furthermore, there is already extensive work on noisy group testing.

Noisy group testing Without going too much in detail, the state of the play is comparable to the state of the art in noiseless non-adaptive group testing prior to this thesis's contributions. While simple algorithms are well understood [89, 108, 155], it is even not known under simplistic noise models like the binary symmetric channel if recovery of the ground-truth is possible efficiently at the Shannon capacity bound. It might be tempting to analyse whether a spatially coupled design could improve bounds in the noisy setting as well.

Question 3.1.7. *Does a spatially coupled design coming with a* `SPIV` *like algorithm perform better at inference under a noisy setting than currently known algorithms?*

Returning to the noiseless case, we find that even the linear case is not completely understood.

Linear group testing More precisely, while non-adaptive algorithms need to fail with high probability, it is possible to infer the ground-truth within multiple rounds of testing at the universal converse [13, 98]. Nevertheless, one important question is still open.

Question 3.1.8. *Are there any (potentially exponential-time) algorithms on less than* $n-1$ *tests which succeed at inference of the ground-truth if the prevalence is larger than* $1/3$ *in the hypergeometric group testing problem? Or conversely, can we proof that such algorithms do not exist?*

A bit more application-driven question is the following. Which influence do such asymptotic designs have on real world group testing?

Applications Suppose we have $k = 5$ infected individuals within a population of $n = 1000$. Then the following statement sounds correct.

> By the given prevalence of 0.5% we are clearly in the setting of linear group testing. Therefore, any non-adaptive group testing strategy fails due to Aldridge [7] and we need 1000 tests for inference within one round.

While this seems to be indeed true, let us provide a second statement which might also sound plausible.

> As the number of infected individuals k scales like $n^{0.233}$, we require $\frac{1}{\ln^2 2} \cdot 5 \cdot \ln(200) \approx 56$ tests such that `DD` infers the infected individuals correctly.

Finally, the folklore counting bound yields that we require $2^m > \binom{1000}{5}$, thus we need at least $m \geq 43$ tests.

Of course, above's statements do not only contradict each other, they are also false. All provided proofs in all contributions (besides the universal counting bound) are obtained under the assumption that n tends to ∞. Therefore, all we can say is that something in between 43 and 1000 tests will be the correct answer. But this is of course very unsatisfactory.

Question 3.1.9. *Given a real world instance on n individuals with an infection rate of* α *with a realistic false positive and false negative rate, how could an almost optimal non-adaptive or two-stage design look like in order to infer almost all individuals correctly?*

An answer to this question would be of high interest in applications. First intends are done by, for instance, Aldridge [12] and Cuturi et al. [57], but we are far from knowing precise statements. A natural suggestion is that an inference algorithm based on Belief Propagation might facilitate best-possible.

This question directly extends to different inference problems. It might be a fruitful research direction to apply message passing algorithms coming with a spatially coupled design to various inference problems. Unfortunately, in general it is much harder to find combinatorial descriptions of those message passing algorithms as in the group testing problem. Therefore, they might be highly challenging to analyse.

3.2. Random satisfiability

As described earlier, we could exactly pin down the number of satisfying assignments of a random 2−SAT formula in terms of the Bethe functional. A natural question would be if the result can be extended to general random k−SAT formulas. It is at least far from clear if this is possible. Indeed, the core of our proof technique is the possibility to construct worst-case conditions at the boundary of a random tree in order to prove that the Boltzmann distribution is a Bethe state itself. Even in a random 3−SAT formula it is not clear how such worst-case conditions might be constructed as the number of possibilities to nudge the marginal of a parent variable into a certain direction is huge and dependent on decisions in different sub-trees.

Within the setting of random 2−SAT we could furthermore ask if we can pin down the distribution of $\ln Z(\boldsymbol{\Phi})$ exactly.

Question 3.2.1. *Does* $n^{-1/2}(\ln Z(\boldsymbol{\Phi}) - \mathbb{E}\ln Z(\boldsymbol{\Phi}))$ *converge to a Gaussian random variable?*

3.3. The cut-distance, regularity and limits of probability measures

We managed to develop a consistent limit theory for discrete probability measures akin to the graph limit theory. Furthermore, we introduced the pinning operation on the limit objects (Ω−laws) as an elegant and easy algorithm to obtain a decomposition of the phase state on which the respective probability measures are extremal, thus close to product measures under the cut-distance. This is a kind of a regularity lemma which allows to write a probability measure as a convex combination of simple measures.

There is a second string of research, primarily by Austin [17], which develops similar regularity lemmas based on the so-called *dual total correlation (DTC)*. The latter can be seen as a generalisation of the classical mutual information from 2 to n random variables which reads as

$$DTC(\mu) = H(\boldsymbol{X}_1,\dots,\boldsymbol{X}_n) - \sum_{i=1}^{n} H(\boldsymbol{X}_i \mid \boldsymbol{X}_1,\dots,\boldsymbol{X}_{i-1},\boldsymbol{X}_{i+1},\dots,\boldsymbol{X}_n)$$

with $H(\cdot)$ being the entropy and $\boldsymbol{X}_1,\dots,\boldsymbol{X}_n$ random variables with joint distribution μ.

Described very briefly, it says that a small dual total correlation implies that a probability measure is close to a product measure with respect to a specific transportation metric [17, Theorem A]. Further, a regularity lemma is known which guarantees that there is a partition of the phase space which allows to write a probability measure as a convex combination of measures with low dual total correlation [17, Theorem 1.1]. An intriguing question is the following.

Question 3.3.1. *Is there a natural connection between the cut-distance of* μ *and* $\bar{\mu}$ *and the dual total correlation of* μ*?*

A second and a third question regarding regularity arise with respect to the pinning operation itself. First, we need to pin a random number of coordinates in order to achieve an extremal measure. It is not clear whether this part of the statement has a deeper reason or if it is just a relic of the proof technique which is originally based on a contribution of Raghavendra and Tan [148]. Furthermore, the pinning

lemma yields a sufficient condition on how many variables have to be pinned to specific spins. A lower bound on how many variables are necessary is currently unknown.

Question 3.3.2. *What is a lower bound on the number of variables which are necessary to be pinned in the pinning operation? Is it necessary that this number is random?*

3.4. Perturbed graphs

The note on spanning structures in randomly perturbed sparse graphs establishes sufficient results for Hamilton cycles and matchings in graphs that are optimal up to a constant. Furthermore, it gives a meta-theorem which proves that the existence of almost spanning bounded degree trees in $\mathcal{G}(n,\beta/n)$ suffices in order to gain the complete spanning structure in $\mathcal{G}(n,\beta/n)$ for $\beta = -\Omega((\Delta+1)\ln(\alpha))$ if Δ is the maximum degree of the tree. Thus, the first question which was also stated in the note reads as follows.

Question 3.4.1. *Is any given tree with maximum degree Δ on $(1-C\exp(-\beta))n$ vertices contained in $\mathcal{G}(n,\beta/n)$ for $0<\beta\leq\ln n$ and a suitable chosen constant C?*

Moreover, we already discussed that the sufficient conditions proven in the note are probably not tight but the proof technique yields to those expressions. It is likely to be true that the choice of other, more complicated, absorbing structures and a more carefully applied large deviations analysis might yield to better constants in the theorems. Therefore, it is an interesting question to pin down the exact phase transition points.

Question 3.4.2. *Can we establish strict phase transitions for the existence of a Hamilton cycle and a perfect matching in $\mathcal{G}(n,\beta/n)\cup\mathcal{G}_\alpha$?*

Finally, while we only studied certain spanning structures in graphs, the objects of interest can be extended. First, we can study different spanning structures in graphs like for instance triangle factors. Second, one can extend the analysis from perturbed sparse graphs to perturbed sparse hypergraphs.

4. Zusammenfassung

Die Analyse sehr großer diskreter Systeme ist ein wesentlicher Bestandteil aktueller Forschung in, unter anderem, der Diskreten Mathematik, der Informatik sowie der Statistischen Physik. Von besonders hoher Relevanz ist das Phänomen der *Phasenübergänge* [78, 82, 121]. Hierbei handelt es sich um Momente in der Evolution eines Systems, an welchen sich dessen Verhalten dramatisch verändert. Solche Phasenübergänge wurden und werden beispielsweise in zufälligen Graphen, Spin-Gläsern und bezüglich der Performance von Algorithmen untersucht.

Seit einigen Jahrzehnten kristallisiert sich immer weiter heraus, dass einige Ideen der Statistischen Physik auf die Untersuchung von eigentlich rein kombinatorischen Problemen übertragen werden können [168]. Zu solchen Problemen zählen zum Beispiel die Beantwortung der Frage nach Erfüllbarkeit gegebener zufälliger Formeln oder auch die Analyse der algorithmischen und informationstheoretischen Lösbarkeit von Inferenzproblemen wie dem *Group-Testing* Problem [68, 133]. Eine wesentliche Herausforderung aus Sicht der Mathematik besteht darin, die Ideen der physikalischen Heuristiken in rigorose mathematische Verfahren und Beweise zu übersetzen.

In dieser Dissertation untersuchen wir, wie Ideen aus der physikalischen Theorie der *diluted mean-field models* für Spin-Gläser dazu genutzt werden können, zufällige Erfüllbarkeitsprobleme zu analysieren. Insbesondere nutzen wir diese Ideen, um zu berechnen, wie viele Lösungen eine zufällige 2-SAT Formel in der Regel besitzt. Zudem werden wir die sogenannte *planted* Version von solchen zufälligen Erfüllbarkeitsproblemen untersuchen. Es stellt sich durch die Brille der Statistischen Physik betrachtet heraus, dass solche Modelle genutzt werden können, um statistische Inferenzprobleme auszudrücken [168].

Sowohl durch die Analyse von Problemen der statistischen Inferenz als auch durch das Untersuchen der zufälligen Erfüllbarkeit stellen wir fest, dass eine geschickte Kombination von Ideen der statistischen Physik mit ureigenen kombinatorischen Eigenschaften zufälliger Graphen zu rigorosen, neuen Resultaten führt. Wir beginnen mit einer knappen Einführung in wesentliche Begriffe der statistischen Physik.

Grundlagen der statistischen Physik Gegeben sei eine Menge $V = \{x_1, \dots, x_n\}$ von *n Partikeln*. Ein *Spin-System* mit diesen Partikeln besteht aus einer endlichen Menge Ω von möglichen Spins sowie einem $k-$uniformen dekorierten Hypergraphen $G = (V, E, J)$, welcher die Interaktionen zwischen den Partikeln beschreibt. Genauer gesagt bezeichnen wir mit $\sigma \in \Omega^n$ eine *Konfiguration*, welche jedem Partikel einen der möglichen Spins aus Ω zuordnet und mit $H : \Omega^n \to \mathbb{R}$ eine *Energie-Funktion*, die jeder Konfiguration ihre Energie zuweist. Formal definieren wir die Energie-Funktion H als

$$H(\sigma) = - \sum_{(i_1,\dots,i_k)\in E(G)} J_{i_1,\dots,i_k}(\sigma_{i_1}, \dots, \sigma_{i_k}).$$

Im Spezialfall $k = 1$ bezeichnen wir das System als ein nicht-interagierendes System, da die verschiedenen Partikel gar nicht miteinander interagieren, während höhere Werte von k ein sogenanntes $k-$*body interacting* System beschreiben. In dieser Dissertation werden wir nur den Fall $k = 2$ untersuchen, das heißt, der zugrunde liegende Hypergraph G ist ein einfacher Graph. Die Familie $\{J_{i,j}\}_{ij\in E(G)}$ beschreibt die Stärke und Art der Interaktion der interagierenden Partikel.

Jede Wahl der *Coupling-Konstanten* J_{ij} sowie der Energie-Funktion H und des Interaktionsgraphen G beschreibt ein spezifisches Spin-System. Klassische Systeme, wie das Potts-Modell oder das Edwards-Anderson-Modell sind Modelle, in denen G einem Gitter-Graphen (zum Beispiel $\mathbb{Z}^3$) entspricht. Diese Modelle mögen die naheliegensten Interaktionsmodelle für zum Beispiel Ferromagnetismus sein, allerdings sind diese auf Grund der geometrischen Beziehungen im Graphen mathematisch sehr schwer zu analysieren [135]. Eine mögliche Vereinfachung sind sogenannte *mean-field*-Modelle wie das SK-Modell [158], in denen der Interaktionsgraph dem vollständigen Graphen entspricht. Ferner wird eine Energie-Funktion gewählt, deren Wert invariant gegenüber Permutationen der Partikel ist. Auf diese Art und Weise werden einfachere Modelle definiert, welche die lokale und globale Struktur sowie Abhängig-

keit verschiedener Partikel ignorieren. Es stellt sich heraus, dass diese Modelle einfacher zu analysieren sind als ihre zugehörigen Gitter-Pendants [143, 162], allerdings wichtige physikalische Eigenschaften nicht korrekt beschreiben können [157].

Die sogenannten *diluted mean-field* Modelle, wie das Viana-Bray-Modell [166], versuchen die mathematische Einfachheit der mean-field-Theorie möglichst beizubehalten und dennoch wesentliche Eigenschaften realer Systeme möglichst gut zu modellieren. Der Kerngedanke ist, dass der zugrunde liegende Interaktionsgraph ein (zufälliger) dünner Graph ist, das heißt, dass zwar die globale Geometrie sicherlich nicht gegeben ist, der Graph lokal aber an einen Gittergraphen erinnert. Tatsächlich ist es so, dass die Analyse gut gewählter *diluted mean-field* Modelle zu exakten Lösungen von Problemen in (echten) Spin-Glas-Modellen führen kann [144], das heißt, diese Modelle tragen noch wesentliche Informationen von realitätsnahen Modellen in sich, während sie mathematisch analysierbar bleiben.

Außerdem können wesentliche Probleme der Informatik, wie das *zufällige Erfüllbarkeitsproblem* und weitere *Constraint Satisfaction*-Probleme (CSPs), als *diluted mean-field* Modell beschrieben werden.

Bevor wir uns diesem Punkt zuwenden, werden wir noch eine wichtige Wahrscheinlichkeitsverteilung, die sogenannte *Boltzmann-Verteilung*, einführen. Dazu erinnern wir uns an ein wesentliches Prinzip der Physik: Ein System versucht immer einen Zustand minimaler Energie einzunehmen. Auch Modelle von Partikelsystemen sollten diese Eigenschaft widerspiegeln. Wir sagen, dass das System in einem Zustand minimaler Energie ist, wenn $H(\sigma)$ durch die Konfiguration $\sigma \in \Omega^n$ minimiert wird. Außerdem bezeichnen wir mit $\Omega_0^n \subset \Omega^n$ die Menge der Konfigurationen, welche die Energie minimieren. Eine vernünftige und natürliche Wahrscheinlichkeitsverteilung sollte also Konfigurationen in Ω_0^n bevorzugen.

Dazu führen wir die *inverse Temperatur* $\beta > 0$ ein und definieren die Boltzmann-Verteilung auf Ω^n als

$$\mu_\beta(\sigma) = \frac{\exp\left(-\beta H(\sigma)\right)}{Z_\beta}, \qquad \text{wobei} \qquad Z_\beta = \sum_{\sigma \in \Omega^n} \exp\left(-\beta H(\sigma)\right).$$

Die Normalisierungskonstante Z_β der Boltzmann-Verteilung wird auch *Partitionsfunktion* genannt. Die Boltzmann-Verteilung spiegelt die Idee, dass das System einen Zustand minimaler Energie anstrebt, wider. Je geringer die Energie $H(\sigma)$, desto höher ist die Wahrscheinlichkeit σ unter μ_β zu beobachten. Steigt die Temperatur des Systems stark an (β wird klein), so verringert sich dieser Effekt und für $\beta \to 0$, also im *high-temperature limit*, wird die Boltzmann-Verteilung zur uniformen Verteilung auf Ω^n. Wächst β hingegen, wird das System also abgekühlt, so verstärkt sich o.g. Effekt, sodass die Boltzmann-Verteilung im *zero-temperature limit* die uniforme Verteilung auf allen Zuständen minimaler Energie wird. Formal ausgedrückt finden wir die folgenden Zusammenhänge:

$$\lim_{\beta \to 0} \mu_\beta(\sigma) = \frac{1}{|\Omega^n|} \qquad \text{und} \qquad \lim_{\beta \to \infty} \mu_\beta(\sigma) = \frac{\mathbf{1}\left\{\sigma \in \Omega_0^n\right\}}{\left|\Omega_0^n\right|}.$$

Da in der statistischen Physik (und auch in der theoretischen Informatik) oft das makroskopische Verhalten von sehr großen Systemen ($n \to \infty$) von Interesse ist, wird häufig der sogenannte *thermodynamische Grenzwert* eines Systems betrachtet [135]. Wir bezeichnen mit $\phi_{n,\beta} = \ln(Z_\beta)$ die *freie Entropie* und mit $\phi_\beta = \lim_{n\to\infty} \frac{\ln(Z_\beta)}{n}$ die *freie Entropiedichte* (*free entropy density*). Innerhalb eines physikalischen Systems definieren wir nun die nicht-analytischen Punkte von ϕ_β als *Phasenübergänge*. Das sind Punkte, an denen sich das qualitative Verhalten des Systems drastisch ändert [135]. Wir betrachten insbesondere Phasenübergänge von den bereits angesprochenen CSPs.

Constraint Satisfaction und Phasenübergänge Zunächst definieren wir, was wir unter CSPs verstehen. Ein besonders prominentes Beispiel ist das $k-$SAT Problem, also die Frage nach Erfüllbarkeit einer aussagenlogischen Formel in konjunktiver Normalform mit Klauseln der Größe k. Genauer gesagt ist eine $k-$SAT Formel Φ eine Konjunktion von m Klauseln

$$\Phi = \Phi_1 \wedge \ldots \wedge \Phi_m,$$

sodass jede Klausel selbst eine Disjunktion von exakt k Literalen der Variablen $x_1 \ldots x_n$ ist.

Ist eine Formel Φ gegeben, so lautet eine der wichtigsten Fragen offenbar, ob es eine Belegung $\sigma : \{x_1, \ldots, x_n\} \to \{-1, +1\}^n$ gibt, welche jeder der Variablen einen der Wahrheitswerte WAHR und FALSCH zuordnet, sodass jede Klausel Φ_j (und somit die gesamte Formel Φ) erfüllt ist. Wir interpretieren den Spin +1 als WAHR und den Spin -1 als FALSCH. Eine solche Formel Φ kann als *Faktorgraph* visualisiert werden.

Ein Faktorgraph $G = (V \cup F, E, \Psi)$ ist ein bipartiter Graph mit einer Menge von Variablenknoten V und einer Menge von Faktorknoten F, einer Kantenmenge E sowie einer Familie von Gewichtsfunktionen Ψ [81].

Wir beschreiben im Folgenden eine Konstruktionsanweisung für einen Faktorgraphen G^Φ, der die $k-$SAT Formel Φ repräsentiert [135]. Seien die Knotenmengen als

$$V = \{x_1, \ldots, x_n\} \qquad \text{sowie} \qquad F = \left\{a_1^\Phi, \ldots, a_m^\Phi\right\}$$

gegeben. Ferner besteht die Kantenmenge $E = \left\{a_1^\Phi, \ldots, a_m^\Phi\right\}$ aus zwei disjunkten Klassen E^+ und E^-, sodass die Kante $x_i a_j^\Phi$ genau dann in E^- liegt, wenn Variable x_i negiert in Klausel Φ_j vorkommt, und in E^+, falls x_i positiv in Φ_j enthalten ist. An jedem Faktorknoten a_j existiert eine lokale Funktion $\Psi_{a_j} : \{-1, +1\}^{|\partial a_j|} \to \{-1, +1\}$, sodass für eine Belegung $\sigma \in \{-1, +1\}^n$ das Folgende gilt:

$$\Psi_{a_j}(\sigma_{\partial a_j}) = \mathbf{1}\left\{\max_{x_i \in \partial a_j} \{\sigma_i s_{ij} = 1\}\right\}.$$

Eine Visualisierung eines solchen Faktorgraphen findet sich in Abbildung 4.1.

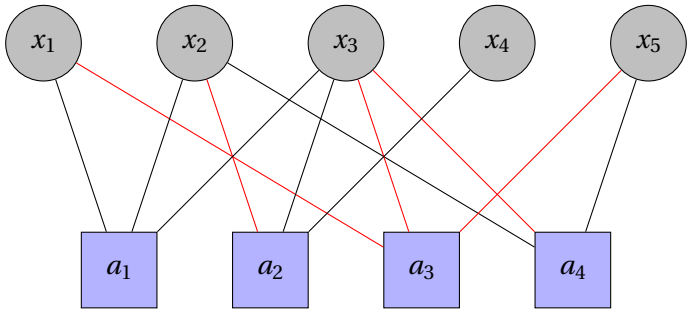

Abbildung 4.1.: Der Faktor-Graph G^Φ zur 3−SAT Formel $\Phi : (x_1 \vee x_2 \vee x_3) \wedge (\neg x_2 \vee x_3 \vee x_4) \wedge (\neg x_1 \vee \neg x_3 \vee \neg x_5) \wedge (x_2 \vee \neg x_3 \vee x_5)$. Die $n = 5$ Variablenknoten sind als Kreise dargestellt, die $m = 4$ Faktorknoten als Rechtecke und die Farbe der Kanten beschreibt, ob eine Variable negiert oder nicht negiert in einer Klausel vorkommt.

Während für $k \geq 3$ das Entscheidungsproblem, ob eine $k-$SAT Formel mindestens eine erfüllende Belegung besitzt, ***NP***-schwer ist [109], kann es leicht als physikalisches System aufgefasst werden [111]. Die n Variablen entsprechen den Partikeln des Systems und die Wahrheitswerte den Spins $\Omega = \{-1, +1\}$. Eine Belegung entspricht nun einer Konfiguration $\sigma \in \Omega^n$ und eine mögliche Energie-Funktion ist

$$H_{k-SAT}(\sigma) = \sum_{a_j^\Phi \in F} 1 - \Psi_{a_j}(\sigma_{\partial a_j}).$$

Also entspricht $H_{k-SAT}(\sigma)$ genau der Anzahl an nicht-erfüllten Klauseln unter einer gegebenen Konfiguration. Betrachten wir nun den *zero-temperature limit*, so ist die entsprechende Boltzmann-Verteilung $\mu_\infty = \lim_{\beta \to \infty} \mu_\beta$ die uniforme Verteilung auf solchen Konfigurationen, die am wenigsten Klauseln verletzen. Das heißt, wenn $\boldsymbol{\sigma} \sim \mu_\infty$ eine zufällige Konfiguration (gezogen von μ_∞) ist, so ist Φ genau dann erfüllbar, wenn $H_{k-SAT}(\boldsymbol{\sigma}) = 0$ gilt. Da wir wissen, dass $k-$SAT ***NP***-schwer ist, folgt direkt, dass es im Allgemeinen auch schwer ist, eine Konfiguration minimaler Energie in einem Spin-System zu finden.

An dieser Stelle fällt auf, dass sich die Boltzmann-Verteilung eines CSPs sehr elegant faktorisieren

lässt, das heißt,

$$\mu_\infty(\sigma) = \lim_{\beta\to\infty} \frac{\prod_{a_j^\Phi \in F} \exp\left(-\beta \mathbf{1}\left\{a_j^\Phi \text{ is not satisfied under } \sigma\right\}\right)}{Z(\Phi)}.$$

Das ist keine Eigenart spezieller CSPs, sondern eine sehr universelle Eigenschaft aller CSPs, die sich als Faktorgraph ausdrücken lassen. Wir bemerken, dass ...

- ... jeder Faktor in der Boltzmann-Verteilung zu einem Faktorknoten in G^Φ korrespondiert,
- ... für $\beta > 0$ jede nicht-erfüllte Klausel eine Art Strafe von $\exp(-\beta)$ auf die Wahrscheinlichkeit, diese Konfiguration zu beobachten, addiert.
- ... falls Φ erfüllbar ist, der Träger der Boltzmann-Verteilung im *low-temperature limit* den erfüllenden Belegungen einer Formel entspricht.
- ... falls Φ erfüllbar ist, die Partitionsfunktion im *low-temperature limit* der Anzahl der erfüllenden Belegungen entspricht.

Oft werden im Kontext von CSPs sogenannte zufällige CSPs betrachtet [156]. Was meint zufällig hierbei? Sind n Variablenknoten und $\boldsymbol{m}$ Faktorknoten (wobei $\boldsymbol{m}$ durchaus auch zufällig sein kann) sowie deren Grade gegeben, so bilden wir einen zufälligen bipartiten Graphen. Je nach CSP können die Gradsequenzen der Knoten selbst zufällig sein. Sind die Gradsequenzen gegeben, so ziehen wir uniform und unabhängig von allem anderen Zufall einen Faktorgraphen, der die entsprechenden Gradsequenzen hat.

Im Falle von $k-$SAT werden n Variablenknoten und m Faktorknoten gegeben. Jeder Faktorknoten hat Grad k und jede Variable hat Grad $\mathbf{Po}(mk/n)$. Wir wählen, gegeben $\left\{\sum_{i=1}^n \boldsymbol{d}_i = mk\right\}$, einen zufälligen einfachen Graphen mit den gegebenen Gradsequenzen. Ferner markieren wir jede Kante unabhängig und uniform mit $+1$ mit Wahrscheinlichkeit $1/2$ und mit -1 mit Wahrscheinlichkeit $1/2$. Eine sehr einfache Frage lautet: Ist für $n \to \infty$ die entstandene zufällige Formel mit hoher Wahrscheinlichkeit erfüllbar? Diese Frage wurde exzessiv untersucht, unter anderem auch mit Methoden der statistischen Physik [37, 111], und es wurde eine präzise Vermutung für ein kritisches Verhältnis $\alpha_s = m_s/n$ zwischen Klauseln und Variablen formuliert, sodass eine zufällige Formel mit geringerem Verhältnis erfüllbar und mit höherem Verhältnis nicht erfüllbar ist. Wir haben also einen Phasenübergang gefunden. Genauer gesagt wurde dieses Problem aufbauend auf vielfältigen Arbeiten [4, 51, 54, 91] schlussendlich von Ding, Sly und Sun [64] gelöst, die beweisen konnten, dass für k groß genug der *Erfüllbarkeits-Schwellenwert* bei

$$\alpha_s = 2^k \ln 2 - \frac{1+\ln 2}{2} + O(2^{-k})$$

liegt. Natürlich finden solche Phasenübergänge auch in allgemeinen zufälligen CSPs statt [136] und der Erfüllbarkeits-Schwellenwert ist nicht der einzige interessante Schwellenwert. Sei dazu $\mathscr{S}$ die Menge aller erfüllenden Belegungen einer zufälligen Formel. Wir sagen, dass zwei Lösungen verbunden sind, wenn ihr Hamming-Abstand 1 ist, und bezeichnen die Menge aller verbundenen Lösungen als Cluster. Es zeigt sich, dass die Geometrie von $\mathscr{S}$ hoch komplex ist, aber glücklicherweise liefert der 1-RSB-Ansatz der statistischen Physik ein nicht-rigoroses aber detailliertes Bild, wie sich $\mathscr{S}$ mit wachsendem Faktor-zu-Variablen-Verhältnis α entwickelt (siehe Abbildung 4.2).

Wir starten bei $\alpha = 0$ und lassen α stetig wachsen. Dann beobachten wir die Existenz von vier Schwellenwerten $\alpha_u \le \alpha_{\text{clus}} \le \alpha_{\text{cond}} \le \alpha_s$, an denen sich die Struktur von $\mathscr{S}$ drastisch verändert [121, 138, 169, 170]. Wir bezeichnen $\mathscr{S}$ manchmal als Lösungsraum und die erfüllenden Belegungen analog als Lösungen.

1. Ist $\alpha < \alpha_u$, so existiert exakt ein Cluster von Lösungen. Diese Phase wird als *unique phase* bezeichnet.

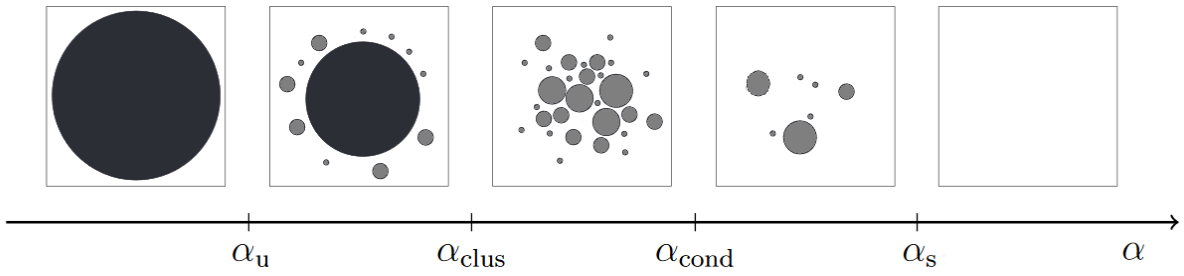

Abbildung 4.2.: Übersicht über die Geometrie des Lösungsraums in zufälligen CSPs. Der 1-RSB-Ansatz sagt die Existenz von vier wichtigen Phasenübergängen vorher. Die Abbildung ist modifiziert nach [138, 169, 170].

2. Sobald α größer wird als α_u, befindet sich das System in der *extremalen Phase*. Wenige und exponentiell kleine Cluster von Lösungen entstehen, die neben einem großen Cluster einige wenige Lösungen enthalten.

3. Am *clustering*-Schwellenwert α_{clus} zerfällt die Menge der Lösungen in exponentiell viele, exponentiell kleine Cluster.

4. Wenn α nun größer wird als der *condensation*-Schwellenwert, so befinden sich fast alle Lösungen in endlich vielen verschiedenen Clustern.

5. Schlussendlich, sobald α größer wird als der Erfüllbarkeits-Schwellenwert α_s, enthält $\mathscr{S}$ keine Lösungen mehr.

Formal betrachtet existieren strikte und schwache Schwellenwerte. Ist $\mathscr{P}$ irgendeine Eigenschaft und α eine Parametrisierung eines zufälligen Systems (im obigen Beispiel ist $\mathscr{P}$ die Menge der nicht erfüllbaren Formeln und α die Faktoren-zu-Variablen-Dichte), dann durchläuft das System $\mathscr{G}$ (der Faktorgraph) einen strikten Phasenübergang am strikten Schwellenwert α^*, falls für jedes $\varepsilon > 0$:

$$\mathbb{P}\left(\mathscr{G} \in \mathscr{P} \mid \alpha \leq (1-\varepsilon)\alpha^*\right) = o(1) \qquad \text{und} \qquad \mathbb{P}\left(\mathscr{G} \in \mathscr{P} \mid \alpha \geq (1+\varepsilon)\alpha^*\right) = 1 - o(1).$$

Analog findet ein schwacher Phasenübergang statt, falls

$$\mathbb{P}\left(\mathscr{G} \in \mathscr{P} \mid \alpha = o(\alpha^*)\right) = o(1) \qquad \text{sowie} \qquad \mathbb{P}\left(\mathscr{G} \in \mathscr{P} \mid \alpha = \omega(\alpha^*)\right) = 1 - o(1).$$

Die Definition verlangt offenbar, dass $\mathscr{P}$ mit wachsendem α wahrscheinlicher wird. Natürlich kann es analog auch für Eigenschaften definiert werden, die mit fallendem α wahrscheinlicher werden.

Friedgut und Bourgain [82] haben gezeigt, dass jedes zufällige System bezüglich einer monoton steigenden oder fallenden Eigenschaft einen nicht-uniformen strikten Phasenübergang durchläuft. Beispielsweise fällt die Frage, ob eine zufällige $k-$SAT Formel erfüllbar ist, darunter, da das Hinzufügen von Klauseln die Wahrscheinlichkeit reduziert, dass die Formel erfüllbar bleibt. Selbstverständlich liefert das Resultat nur eine Existenzaussage, das heißt, es bleibt unklar, was die entsprechenden Schwellenwerte sind. Für viele Eigenschaften bezüglich des zufälligen Graphen $\mathscr{G}(n,p)$[1] ist der Schwellenwert bekannt [103], aber in Bezug auf zufällige CSPs sieht die Welt anders aus, auch wenn in den letzten Jahren Schwellenwerte für manche zufälligen CSPs gefunden wurden [25, 43, 55, 64].

In dieser Dissertation beantworten wir eine offene Frage bezüglich des Lösungsraums des zufälligen $2-$SAT Problems. Das $2-$SAT Problem ist ein Spezialfall des bereits diskutierten $k-$SAT und auf eine gewisse Art und Weise eine Besonderheit. Es ist das einzige $k-$SAT Problem, in welchem es komplexitätstheoretisch einfach ist, eine erfüllende Belegung zu finden (sofern eine solche existiert) [120]. Auch

[1] Wir definieren $\mathscr{G}(n,p)$ wie Gilbert [90], das heißt, jede Kante ist existent mit Wahrscheinlichkeit p unabhängig von allem anderen Zufall.

der Erfüllbarkeits-Schwellenwert des zufälligen 2−SAT Problems ist seit den frühen 1990er Jahren bekannt [40, 91] und hängt eng mit dem Perkulationsphasenübergang im zufälligen Graphen zusammen [30]. Dennoch blieb eine sehr unschuldig vermutende, aber zentrale Frage offen [78]: Falls eine zufällige Formel erfüllbar ist, wie viele erfüllbare Belegungen gibt es? Es stellt sich heraus, dass dies schwer zu beantworten ist, insbesondere ist im (nicht-zufälligen) 2−SAT das Zählen der Lösungen komplexitätstheoretisch schwierig, es liegt in #**P** [165]. Die in der vorliegenden Dissertation enthaltene Publikation

The number of satisfying assignments of random 2-SAT formulas [2]

beantwortet diese Frage vollständig. Dies gelingt, indem wir zeigen, dass im Falle des zufälligen 2−SAT Problems bestimmte Heuristiken der statistischen Physik in einen rigorosen mathematischen Beweis verwandelt werden können. Genauer gesagt zeigen wir, dass die durch Belief Propagation berechnete Lösung der Marginale der zugehörigen Boltzmann-Verteilung korrekt ist.

Die Cut-Distanz Dies führt uns zum nächsten, in dieser Dissertation enthaltenen Beitrag mit dem Titel

The cut metric for probability distributions [47].

Die Boltzmann-Verteilung ist eine diskrete Wahrscheinlichkeitsverteilung auf Ω^n und in vielen zu studierenden Problemen betrachten wir den sogenannten *thermodynamischen Grenzwert* $n \to \infty$. Es liegt daher nahe, kontinuierliche Grenzobjekte von solchen diskreten Wahrscheinlichkeitsverteilungen zu betrachten. Die $\Omega-$*laws*, ursprünglich eingeführt durch Coja-Oghlan, Perkins und Skubch [42] beschreiben solche Grenzobjekte. Genauer gesagt verwandeln wir eine Konfiguration $\sigma \in \Omega^n$ in eine messbare Funktion $\hat{\sigma}$ von $[0,1)$ in die Menge der Wahrscheinlichkeitsmaße über Ω, wobei diese Menge als $\mathscr{P}(\Omega)$ bezeichnet wird. Es sei ferner Σ_Ω der Raum aller messbaren Funktionen $f : [0,1) \to \mathscr{P}(\Omega)$ bis auf Gleichheit fast überall. Nun definieren wir $\hat{\sigma}$ als

$$\hat{\sigma} : [0,1) \to \mathscr{P}(\Omega), \qquad \text{sodass} \qquad x \mapsto \sum_{i=1}^{n} \delta_{\sigma_i} \mathbf{1}\left\{x \in \left[\frac{i-1}{n}, \frac{i}{n}\right)\right\}.$$

Das zugehörige Wahrscheinlichkeitsmaß $\mu \in \mathscr{P}(\Omega^n)$ wird nun in den Raum der Wahrscheinlichkeitsmaße $\mathscr{P}(\Sigma_\Omega)$ wie folgt eingebettet. Wir definieren

$$\hat{\mu} = \sum_{\sigma \in \Omega^n} \mu(\sigma) \delta_{\hat{\sigma}}, \qquad \text{sodass} \qquad \hat{\mu} \in \mathscr{P}(\Sigma_\Omega).$$

Offenbar besteht eine 1-zu-1-Beziehung zwischen μ und $\hat{\mu}$. Es ist weiterhin möglich eine sehr schwache Metrik, die Cut-Distanz, auf $\mathscr{P}(\Sigma_\Omega)$ zu definieren. Dazu bezeichnen wir mit $\mathbb{S}_{[0,1)}$ die Menge der maßerhaltenden, invertierbaren Bijektionen auf $[0,1)$ sowie mit $\Gamma(\mu,\nu)$ die Menge der *Couplings* von μ, ν, also gemeinsamen Wahrscheinlichkeitsverteilungen mit Marginalen μ und ν. Die Cut-Distanz ist nun definiert als

$$D_{\boxtimes}(\mu,\nu) = \inf_{\substack{\gamma \in \Gamma(\mu,\nu),\\ \varphi \in \mathbb{S}_{[0,1)}}} \sup_{\substack{B \subset \Sigma_\Omega^2,\\ U \subset [0,1),\\ \omega \in \Omega}} \left| \int_B \int_U \sigma_x(\omega) - \tau_{\varphi(x)}(\omega) \mathrm{d}x \mathrm{d}\gamma(\sigma,\tau) \right|$$

und ist als eine Art 2-Spieler Spiel zu verstehen. Spieler 1 wählt ein mögliches Coupling von zwei Wahrscheinlichkeitsverteilungen, unter welchem sich die Verteilungen möglichst ähnlich sehen. Nun wählt Spieler 2 eine Menge von (verallgemeinerten) Koordinaten und (verallgemeinerten) Konfigurationen, an denen sich μ und ν stark unterscheiden. Selbstverständlich gibt es von der Cut-Distanz auch eine diskrete Variante, nämlich

$$\Delta_{\boxtimes}(\mu,\nu) = \inf_{\substack{\gamma \in \Gamma(\mu,\nu),\\ \phi \in \mathbb{S}_n}} \sup_{\substack{S \subset \Omega^n \times \Omega^n,\\ X \subset [n],\\ \omega \in \Omega}} \left| \sum_{\substack{(\sigma,\tau) \in S,\\ x \in X}} \gamma(\sigma,\tau) \left(\mathbf{1}\{\sigma_x = \omega\} - \mathbf{1}\{\tau_{\phi(x)} = \omega\} \right) \right|.$$

Hierbei bezeichnet $\mathbb{S}_n$ die Menge der Permutationen von $[n]$. In der o.g. Publikation zeigen wir unter anderem, dass für Maße $\mu, \nu \in \mathscr{P}(\Omega^n)$ die Abstände $\Delta_{\boxtimes}(\mu, \nu)$ und $D_{\boxtimes}(\hat{\mu}, \hat{\nu})$ eng zusammenhängen, wie man es intuitiv auch erwarten sollte.

Mit Hilfe der Cut-Distanz können wir $\mathscr{P}(\Sigma_\Omega)$ in einen kompakten metrischen Raum verwandeln. Genauer gesagt müssen wir Maße mit Cut-Distanz 0 identifizieren und definieren den Raum der $\Omega-$laws $\mathscr{L}(\Omega)$ als den Raum der Äquivalenzklassen unter dieser Identifikation.

Während $\mathscr{L}(\Omega)$, wie bereits erwähnt, durch Coja-Oghlan, Perkins und Skubch [42] eingeführt wurde, etablieren wir eine komplette und in sich konsistente Grenzwerttheorie für diskrete Wahrscheinlichkeitsmaße, die an die Theorie der Graph-Grenzwerte [31, 32, 128] angelehnt ist. Unter anderem beweisen wir, dass $(\mathscr{L}(\Omega), D_{\boxtimes}(\cdot,\cdot))$ ein kompakter metrischer Raum ist, und geben eine Art *schwaches Regularitätslemma* [84] für $\Omega-$laws, was ein zuvor bekanntes Resultat auf diskreten Maßen verallgemeinert [49]. Besonders elegant hierbei ist, dass das Regularitätslemma nicht nur eine Existenzaussage umfasst, sondern auch einen einfachen Algorithmus, das sogenannte *Pinning*, liefert, mit welchem eine entsprechende Partition des Phasenraums Σ_Ω bzw. Ω^n gefunden wird. Ohne zu sehr ins Detail zu gehen, verstehen wir hierbei unter Regularität, dass das Wahrscheinlichkeitsmaß - eingeschränkt auf die Partition - unter der Cut-Distanz sehr ähnlich wie das Produktmaß mit denselben Marginalen aussieht. Wir zeigen unter anderem, dass diese Eigenschaft, welche wir *Extremität* nennen, mit der schwachen Regularität der Graphentheorie eng verknüpft ist. Ferner zeigen wir, dass verschiedene wesentliche Operationen der statistischen Physik (wie zum Beispiel das Bilden von *Overlaps*) unter der Cut-Distanz stetig sind.

Insgesamt hilft die rigorose Analyse der Cut-Distanz in Spezialfällen dabei, Heuristiken der statistischen Physik in mathematische Beweise zu verwandeln. An dieser Stelle kehren wir zurück zur direkten Analyse von CSPs. Bislang haben wir Faktorgraphen von (zufälligen) CSPs betrachtet und uns deren Phasenraum angeschaut. Dieser Lösungsraum ändert sich drastisch, wenn wir die sogenannte *planted*-Version von CSPs betrachten.

Planted-Modelle und statistische Inferenz Wir betrachten dazu zunächst ein Beispiel. Nehmen wir an, wir haben n Variablenknoten gegeben sowie eine Färbung $\boldsymbol{\sigma}$ der Knoten mit q Farben. Nun erzeugen wir einen zufälligen Graphen derart, dass jede Kante ij mit Wahrscheinlichkeit p_1 existiert, falls $\boldsymbol{\sigma}_i = \boldsymbol{\sigma}_j$ (das heißt, ij ist monochromatisch) beziehungsweise mit Wahrscheinlichkeit p_2, falls $\boldsymbol{\sigma}_i \neq \boldsymbol{\sigma}_j$. Nach dem Einfügen der Kanten vergessen wir die zugrunde liegende Färbung $\boldsymbol{\sigma}$. Je nach Wahl von p_1 und p_2 sieht der Graph sehr unterschiedlich aus. Ist $p_1 = p_2$, so kann der entstandene Graph nicht von einem rein zufälligen $\mathscr{G}(n, p_1)$ unterschieden werden, ist hingegen $p_1 \ll p_2$ oder $p_1 \gg p_2$, so sollten wir - gegeben der zufällige Graph - in der Lage sein, eine Färbung $\tilde{\sigma}$ zu finden, die $\boldsymbol{\sigma}$ ähnelt.

Etwas formaler erklärt dieses Vorgehen das *Lehrer-Schüler-Modell* der statistischen Inferenz [168]. Im einfachsten Fall erzeugt ein Lehrer eine Grundwahrheit $\boldsymbol{\sigma}$, ein *planted*-Modell basierend auf dieser Grundwahrheit, und übermittelt einem Schüler das Modell und die Information, wie $\boldsymbol{\sigma}$ und das Modell erzeugt wurden. Die Aufgabe des Schülers ist es nun, eine Vermutung $\tilde{\sigma}$ zu formulieren, die möglichst nah an der Grundwahrheit liegt. Im obigen Beispiel, was eine simple Form des stochastischen Blockmodells [62, 85, 95] darstellt, muss der Schüler die Färbung $\boldsymbol{\sigma}$ möglichst genau aus dem zufälligen Graphen rekonstruieren.

Es stellt sich heraus, dass solche Inferenzprobleme als physikalisches Modell wie zuvor ausgedrückt werden können und sich somit Heuristiken zur Lösung von CSPs auch auf statistische Inferenzprobleme übertragen [168]. Daher ist es nicht überraschend, dass wir auch hier Phasenübergänge untersuchen können. Wir betrachten im Wesentlichen zwei Phasenübergänge. Sei $\mathscr{I}$ die Information, die der Schüler erhält (z.B. der zufällige Graph sowie p_1 und p_2). Die Menge der Information sei durch α parametrisiert. Beispielsweise kann α die Differenz von p_1 und p_2 sein oder die Anzahl von Messungen eines komprimierten Signals.

- Der *informationstheoretische* Schwellenwert bezeichnet den Moment, ab welchem der Schüler $\boldsymbol{\sigma}$ aus $\mathscr{I}(\alpha)$ rekonstruieren kann.
- Der *algorithmische* Schwellenwert bezeichnet die Menge an Information, die notwendig ist, damit ein effizienter Algorithmus bekannt ist, welcher $\boldsymbol{\sigma}$ aus $\mathscr{I}(\alpha)$ rekonstruieren kann.

In dieser Disseration betrachten wir ein spezielles Inferenzproblem – das Group-Testing – und studieren sowohl informationstheoretische Schwellenwerte als auch algorithmische Schwellenwerte.

Group Testing Das Group-Testing Problem fand in den 1940er Jahren den Weg in die mathematische Literatur [68] und wurde über die Jahrzehnte hinweg stetig untersucht [8, 9, 58, 59, 73, 80, 86, 98, 108, 132, 133, 154, 164]. In einer Population von $n \gg 1$ Individuen sind k mit einer Krankheit infiziert. Es ist möglich mehrere Individuen auf einmal zu testen und das Ergebnis eines solchen Gruppentests ist positiv, genau dann wenn mindestens ein infiziertes Individuum in dem Test enthalten ist. Gesucht ist nun eine Strategie, die möglichst wenige Tests benötigt um die infizierten Individuen (mit hoher Wahrscheinlichkeit) korrekt zu identifizieren. Eine Visualisierung findet sich in Abbildung 4.3.

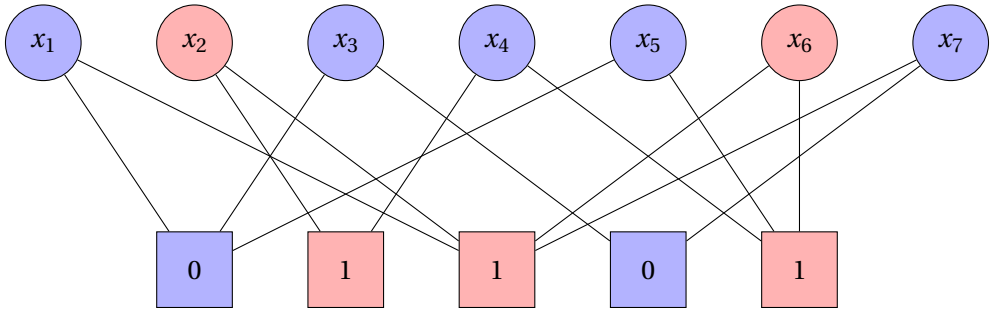

Abbildung 4.3.: Faktorgraph-Darstellung einer Group-Testing Strategie mit $n = 7$ Individuen von denen $k = 2$ infiziert sind. Blaue Individuen sind nicht infiziert und rote Individuen sind infiziert.

Das Group-Testing Problem kann unter vielen verschiedenen Modellen studiert werden. Beispielsweise können wir annehmen, dass jedes Individuum mit Wahrscheinlichkeit k/n infiziert ist oder dass es exakt k Infizierte gibt. Ferner können wir die Teststrategien insofern beschränken, als dass nur eine, zwei oder drei Runden Tests durchgeführt werden können, oder derart, dass Individuen nicht öfter als Δ mal getestet werden dürfen bzw. ein Gruppentest eine Maximalkapazität aufweist. Schlussendlich können die Tests immer ein korrektes Ergebnis liefern oder aber mit einer bestimmten Wahrscheinlichkeit ein falsches Ergebnis liefern. Die hier aufgelisteten Varianten sind weit entfernt davon vollständig zu sein und wir verweisen interessierte Leser auf einen Übersichtsartikel von Aldridge, Johnson und Scarlett [10].

In dieser Disseration beschäftigen wir uns mit dem sogenannten *hypergeometrischen sublinearen probabilistischen* Group-Testing. Das heißt, wir nehmen an, dass wir die Anzahl der infizierten Individuen k exakt kennen und sich k sublinear in n verhält, also $k = n^\theta (\theta \in (0,1))$. Ferner möchten wir die infizierten Individuen mit hoher Wahrscheinlichkeit rekonstruieren. Hierbei betrachten wir sowohl das *uneingeschränkte* Group-Testing Problem, in welchem Individuen beliebig oft getestet werden können und Tests beliebig groß werden dürfen, als auch das *eingeschränkte* Group-Testing Problem, in welchem dies nicht der Fall ist. Insbesondere sind wir an Schwellenwerten interessiert, welche die Anzahl der Tests $m = m(n,k)$ beschreiben, die notwendig bzw. hinreichend sind, um den Infektionsstatus aller Individuen zu rekonstruieren. Genauer gesagt enthält diese Dissertation drei Prublikationen, in denen wir uns mit dem Group-Testing Problem auseinandersetzen, nämlich

Information-Theoretic and Algorithmic Thresholds for Group Testing [41]

und sowohl

Optimal group testing [46]

als auch

Near optimal sparsity-constrained group testing: improved bounds and algorithms [88].

Im nicht-adaptiven Fall, das heißt, dass alle betrachteten Teststrategien alle Tests parallel ausführen müssen, war vor den Beiträgen der Dissertation die beste bekannte Strategie das sogenannte *zufällige reguläre Modell* [9]. In diesem Modell wählt jedes Individuum $\Delta = \Theta(\ln n)$ Tests zufällig aus und nimmt an ihnen teil. Zudem war der beste bekannte Algorithmus der DD-Algorithmus, in welchem zunächst alle Individuen in negativen Tests als gesund deklariert werden und alle Individuen, die nun alleine in einem positiven Test vorkommen, als infiziert. Alle übrigen Individuen werden ebenfalls als nicht-infiziert deklariert [108]. In Abbildung 4.4 sind die Ergebnisse der Publikationen [41, 46] zusammengefasst.

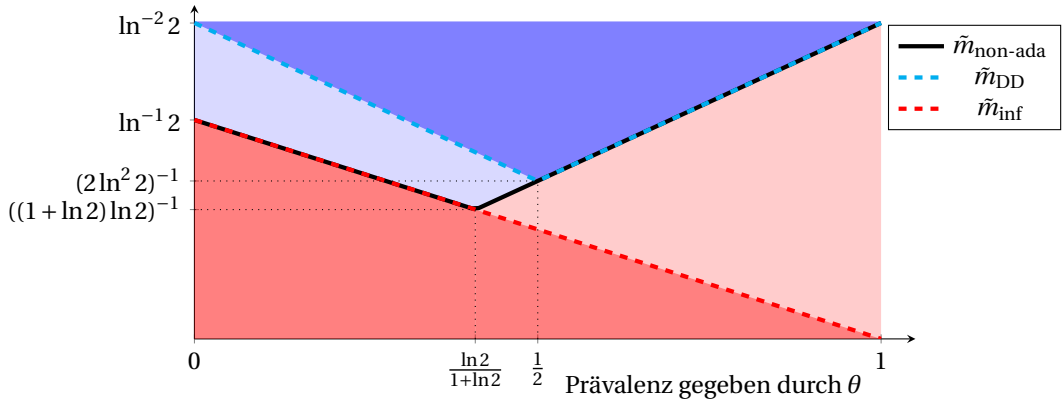

Abbildung 4.4.: Die Phasenübergänge im hypergeometrischen probabilistischen Group-Testing. Die Abbildung ist modifiziert nach [46, Abbildung 1]. Wir definieren $\tilde{m} = m \cdot (k \ln(n))^{-1}$.

Beginnen wir mit den bereits zuvor bekannten Tatsachen. Im dunkelroten Bereich, unter m_{inf}, ist es weder nicht-adaptiv noch adaptiv (in mehreren Runden) möglich, die infizierten Individuen zu identifizieren, wie aus dem Zählargument folgt, dass die Anzahl der möglichen Testergebnisse 2^m mindestens der Anzahl der Konfigurationen mit k infizierten Individuen $\binom{n}{k}$ entsprechen muss. Überhalb dieser Schwellenwertfunktion gibt es Algorithmen, die in mehreren Runden die infizierten Individuen rekonstruieren können [13, 60, 98, 153]. Ferner beschreibt die dunkelblaue Fläche (über m_{DD}) den Bereich, in welchem der DD-Algorithmus auf dem zufälligen regulären Modell funktioniert [108]. Es war ferner bekannt, dass es informationstheoretisch auf selbigem Modell unterhalb von $m_{\text{non-ada}}$ unmöglich ist, die Infizierten zu rekonstruieren [8]. Unsere Beiträge lassen sich wie folgt zusammenfassen, wobei alle angegebenen Schwellenwertfunktionen strikte Phasenübergänge ausdrücken.

- Unterhalb von m_{DD} ist keine Inferenz durch den DD-Algorithmus möglich.
- Unterhalb von $m_{\text{non-ada}}$ kann es keine nicht-adaptive Teststrategie geben, welche die Inferenz der infizierten Individuen ermöglicht. Das heißt, dass die hellrote Fläche einen Bereich darstellt, in welchem adaptive Algorithmen bekannt sind und funktionieren, während nicht-adaptive Strategien keinen Erfolg haben.
- Das zufällige reguläre Modell ermöglicht informationstheoretisch die Inferenz ab $m_{\text{non-ada}}$, ist also informationstheoretisch optimal.
- Wir definieren basierend auf Ideen der Coding-Theorie eine neuartige nicht-adaptive, *spatially-coupled* Teststrategie und einen effizienten Algorithmus, der ab $m_{\text{non-ada}}$, also auch bereits im hellblauen Bereich, funktioniert.
- Die Inferenz von allen bis auf $o(k)$ Individuen ist ab m_{inf} durch das eben genannte Modell mit demselben Algorithmus möglich. Ebenso kann der Algorithmus leicht zu einem zweistufigen Algorithmus verändert werden, der die Inferenz aller Individuen ab m_{inf} ermöglicht.

Zusammenfassend ist somit das hypergeometrische probabilistische Group-Testing im sublinearen Fall vollständig verstanden. Schränken wir die zulässige Anzahl an Tests pro Individuum ein, das heißt, jedes Individuum ist nur in maximal $\Delta = O\left(\ln^{1-\delta} n\right)$ ($\delta \in (0,1]$) Tests enthalten, so ist das Problem noch

nicht vollständig erforscht. Bekannte Resultate sowie unsere erzielten Ergebnisse sind in Abbildung 4.5 dargestellt.

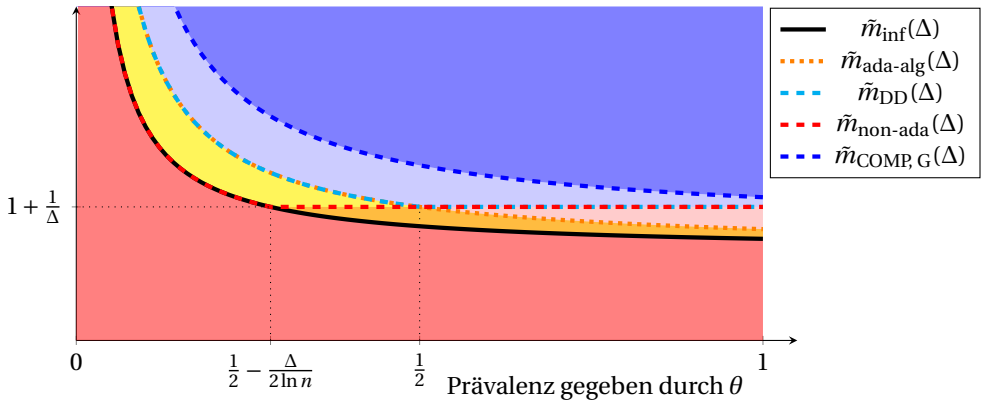

Abbildung 4.5.: Wichtige Phasenübergänge im hypergeometrischen probabilistischen Group-Testing Problem, falls jedes Individuum höchstens Δ-mal getestet werden darf mit $\Delta = O\left(\ln^{1-\delta} n\right)$ für ein $\delta \in (0,1]$. Die Abbildung zeigt die Phasenübergänge für die Wahl $\Delta = 5$ und $n = 10^5$. Wir parametrisieren die eigentliche Schwellenwertfunktion als $m = \Delta k^{\tilde{m}}$.

Die relevantesten zuvor bekannten Resultate gehen auf Gandikota et al. [86] zurück. Insbesondere analysieren die Autoren das zufällige reguläre Modell und einen einfachen Algorithmus `COMP`, der im dunkelblauen Bereich die Inferenz ermöglicht. Dieser sehr einfache Algorithmus deklariert alle Individuen, die in einem negativen Test vorkommen, als uninfiziert während alle weiteren Individuen als infiziert deklariert werden. Dieselben Autoren zeigen, dass jede nicht-adaptive Teststrategie mit höchstens $(\mathrm{e} m_{\text{inf}}(\Delta))^{1-\varepsilon}$ Tests scheitert. Wir erweitern diese Resultate wie folgt und nehmen wiederum implizit an, dass Schwellenwertfunktionen zu strikten Phasenübergängen korrespondieren.

- Der `DD`-Algorithmus auf dem zufälligen regulären Modell ermöglicht Inferenz im blauen Bereich, ist also strikt besser als `COMP`. Außerdem ermöglicht der `DD`-Algorithmus unterhalb von $\tilde{m}_{\text{DD}}$ mit hoher Wahrscheinlichkeit keine Inferenz auf dem zufälligen regulären Modell.
- Unterhalb von $m_{\text{inf}}(\Delta)$ kann keine – auch keine adaptive – Teststrategie erfolgreich sein.
- Unterhalb von $m_{\text{non-ada}}(\Delta)$ kann keine nicht-adaptive Teststrategie die Inferenz der infizierten Individuen ermöglichen.
- Überhalb von $m_{\text{adap-alg}}(\Delta)$ existiert ein effizienter adaptiver Algorithmus, der die Inferenz mit hoher Wahrscheinlichkeit ermöglicht. Das heißt insbesondere, dass adaptive Algorithmen im hellroten Bereich wie zuvor eine bessere Performance als nicht-adaptive Strategien liefern.

In diesem Setting bleibt somit offen, wo der informationstheoretische Phasenübergang für adaptive Algorithmen stattfindet (orangener Bereich) und ob es nicht-adaptive Teststrategien (ggf. mit effizienten Algorithmen) gibt, welche Inferenz im gelben Bereich ermöglichen.

Zuletzt haben wir uns ebenfalls mit einer anderen Art der Einschränkung im Group-Testing Problem beschäftigt. Falls jeder Test nur $\Gamma = \Theta(1)$ Individuen beinhalten darf, so war vor dem Beitrag der Dissertation nur wenig bekannt. Die Ergebnisse sind in Abbildung 4.6 visualisiert.

Gandikota et al. [86] analysieren wiederum den `COMP`-Algorithmus auf dem zufälligen regulären Modell, welcher ab der blauen Linie Inferenz ermöglicht. Ferner zeigt eine einfache Zählschranke, dass mindestens n/Γ Tests in jeder (adaptiven) Teststrategie benötigt werden (schwarze Linie). Wir erzielen die folgenden Resultate.

- Wir etablieren eine universelle informationstheoretische Schranke für alle nicht-adaptiven Teststrategien, das heißt, jede nicht-adaptive Teststrategie kann unterhalb der roten Linie mit hoher Wahrscheinlichkeit nicht die infizierten Individuen rekonstruieren.

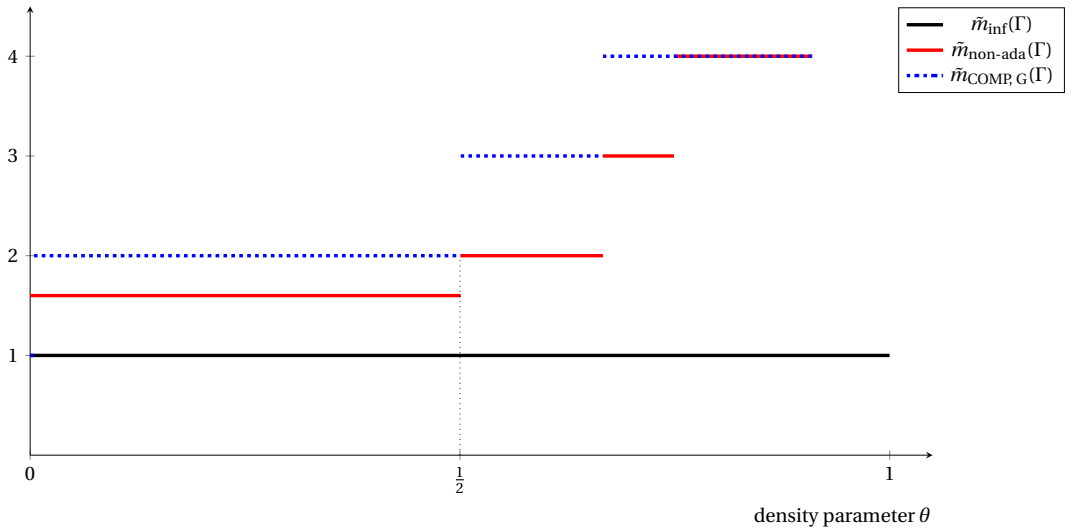

Abbildung 4.6.: Übersicht über die Phasenübergänge im hypergeometrischen probabilistischen Group-Testing, sofern jeder Test höchstens $\Gamma = 4$ Individuen beinhalten darf. Wir parametrisieren die Anzahl der Tests $m = \tilde{m}\frac{n}{\Gamma}$ und nehmen implizit an, dass $n/\Gamma \in \mathbb{Z}$.

- Wir definieren eine neue nicht-adaptive Teststrategie, die bei niedriger Prävalenz ($\theta < 1/2$) vom regulären Modell abweicht, und zeigen, dass der `DD`-Algorithmus für alle Werte von θ außerhalb einer Nullmenge auf diesem Modell ab der roten Linie die Infizierten mit hoher Wahrscheinlichkeit rekonstruiert. Dieses Modell mit dem `DD`-Algorithmus ist also optimal.
- Es existiert ein effizienter adaptiver Algorithmus, der – bis auf Rundung – oberhalb der schwarzen Linie die Infizierten erfolgreich rekonstruiert.

Entsprechend ist dieser Fall des Group-Testings auch nahezu vollständig verstanden. Es bleibt allerdings offen, ob ähnliche Analysen auch für $\omega(1) = \Gamma = o(n/k)$ durchgeführt werden können.

Schlussendlich befasst sich die vorliegende Dissertation mit sogenannten *zufällig perturbierten* Graphmodellen.

Zufällig perturbierte Graphen Ursprünglich stammt die Idee der zufälligen Perturbation deterministischer Systeme aus der *smooth analysis of algorithms*, also der *stufenlosen Analyse von Algorithmen* [159]. Während viele Algorithmen eine exponentielle Laufzeit im schlimmsten Fall (*Worst-Case-Analyse*) aufweisen, so zeigt sich in realen Anwendungen, dass sie meistens sehr effizient funktionieren. Das bekannteste Beispiel ist vermutlich der Simplex-Algorithmus [61] zum Lösen linearer Optimierungsprobleme. In Anwendungen scheinen somit oft die Worst-Case-Bedingungen nicht einzutreten, allerdings ist es wichtig zu verstehen, wie hoch die Laufzeit auf einer anwendungsspezifischen Eingabe vermutlich werden kann. Die Analyse der typischen Laufzeit auf einer zufälligen Eingabe (*Average-Case-Analyse*) beantwortet diese Frage nur ungenügend, da in einer Anwendung durchaus Konstellationen auftreten können, die weit entfernt von einer durchschnittlichen Eingabe sind. Allerdings ist es ebenso unwahrscheinlich, eine Worst-Case-Konfiguration zu beobachten. Aus diesem Grund werden perturbierte Modelle untersucht. Wir beginnen bei einer Worst-Case-Konfiguration, fügen (ein wenig) Zufall hinzu und möchten die Laufzeit des Algorithmus in Abhängigkeit der Menge des hinzugefügten Zufalls untersuchen [159].

Kurz nach Einführung dieser Betrachtungsweise von Algorithmen wurde das Prinzip auf die Existenz aufspannender Strukturen in Graphen übertragen. Eine zentrale aufspannende Struktur ist beispielweise der Hamiltonkreis, also ein Rundweg, der jeden Knoten exakt einmal besucht. Es sind zahlreiche hinreichende Bedingungen, wie Diracs Theorem [65] bekannt, welche zum Beispiel durch Minimalgradbedingungen die Existenz eines Hamiltonkreises in einem beliebigen deterministischen Graphen garantieren. Diese Art von Theoremen kann mit der Worst-Case-Analyse verglichen werden. Auf der anderen

Seite ist der exakte Schwellenwert $p = p(n)$ bekannt, ab welchem der zufällige Graph $\mathcal{G}(n.p)$ mit hoher Wahrscheinlichkeit einen solchen aufspannenden Kreis besitzt (Average-Case-Analyse) [116, 117, 147]. Selbstverständlich existieren solche Analysen nicht nur für Hamiltonkreise, sondern auch unter anderem für Spannbäume, Matchings, Potenzen von Hamiltonkreisen sowie für allgemeine aufspannende Graphen mit beschränktem Maximalgrad, sowohl in deterministischen Graphen [36, 94, 113, 114, 115] als auch in zufälligen Graphen [14, 33, 72, 76, 107, 118, 125, 139, 141, 150].

Das Modell der zufälligen perturbierten Graphen ist nun wie folgt zu verstehen. Gegeben sei ein beliebiger Graph $\mathcal{G}_\alpha$ mit Minimalgrad αn sowie ein zufälliger Graph $\mathcal{G}(n,p)$ mit $p = p(\alpha)$. Nun stellt sich die Frage, wann $\mathcal{G}_\alpha \cup \mathcal{G}(n,p)$ mit hoher Wahrscheinlichkeit einen Hamiltonkreis besitzt. Offenbar ist der Schwellenwert nicht nur abhängig von α, sondern auch von dem gegebenen Graphen $\mathcal{G}_\alpha$, das heißt, wir suchen einen Schwellenwert p^* im folgenden Sinne. Sofern $p > p^*$, so enthält $\mathcal{G}_\alpha \cup \mathcal{G}(n,p)$ einen Hamiltonkreis mit hoher Wahrscheinlichkeit unabhängig von der Wahl von $\mathcal{G}_\alpha$. Ist andererseits $p < p^*$, so existiert mindestens ein Graph $\mathcal{G}_\alpha$ mit Minimalgrad αn, sodass die Vereinigung mit dem zufälligen Graphen keinen Hamiltonkreis besitzt.

Dieses Modell wurde erstmalig von Bohman, Frieze und Martin [28] diskutiert. Es folgten mehrere Publikationen, welche hinreichende Bedingungen für die Existenz verschiedener aufspannender Strukturen in $\mathcal{G}_\alpha \cup \mathcal{G}(n,p)$ geben [20, 26, 35, 34, 119], allerdings liegt all diesen Beiträgen zu Grunde, dass der betrachtete deterministische Graph ein dichter Graph ist, das heißt, dass $\alpha = \Theta(1)$ eine Konstante ist. Das letzte in dieser Dissertation enthaltene Manuskript,

Random perturbation of sparse graphs [93],

beschäftigt sich mit dem Fall $\alpha = o(1)$, also mit dem Fall, dass der zugrunde liegende deterministische Graph dünn ist. In diesem Beitrag geben wir hinreichende Bedingungen für die Existenz von Matchings und Hamiltonkreisen sowie für die Existenz aufspannender Bäume mit beschränktem Maximalgrad, indem wir die Frage nach der Existenz einer aufspannenden Struktur in $\mathcal{G}_\alpha \cup \mathcal{G}(n,p)$ auf die Existenz nahezu-aufspannender Strukturen in $\mathcal{G}(n,p)$ zurückführen.

References

[1] D. Achlioptas and A. Coja-Oghlan. 'Algorithmic Barriers from Phase Transitions'. In: *2008 49th Annual IEEE Symposium on Foundations of Computer Science*. 2008, pp. 793–802.

[2] D. Achlioptas, A. Coja-Oghlan, M. Hahn-Klimroth, J. Lee, N. Müller, M. Penschuck and G. Zhou. 'The number of satisfying assignments of random 2-SAT formulas'. In: *Random Structures & Algorithms* (2021), pp. 1–39.

[3] D. Achlioptas and C. Moore. 'Random k-SAT: Two Moments Suffice to Cross a Sharp Threshold'. In: *SIAM Journal on Computing* 36.3 (2006), pp. 740–762.

[4] D. Achlioptas and Y. Peres. 'The threshold for random k-SAT is 2k (ln 2 - O(k))'. In: *Proceedings of the thirty-fifth ACM symposium on Theory of computing - STOC '03*. ACM Press, 2003.

[5] M. Aizenman, R. Sims and S. L. Starr. 'Extended variational principle for the Sherrington-Kirkpatrick spin-glass model'. In: *Phys. Rev. B* 68 (21 2003), p. 214403.

[6] D. J. Aldous. 'Representations for partially exchangeable arrays of random variables'. In: *Journal of Multivariate Analysis* 11.4 (1981), pp. 581–598.

[7] M. Aldridge. 'Individual Testing Is Optimal for Nonadaptive Group Testing in the Linear Regime'. In: *IEEE Transactions on Information Theory* 65.4 (2019), pp. 2058–2061.

[8] M. Aldridge, L. Baldassini and O. Johnson. 'Group Testing Algorithms: Bounds and Simulations'. In: *IEEE Transactions on Information Theory* 60.6 (2014), pp. 3671–3687.

[9] M. Aldridge, O. Johnson and J. Scarlett. 'Improved group testing rates with constant column weight designs'. In: *2016 IEEE International Symposium on Information Theory (ISIT)*. 2016, pp. 1381–1385.

[10] M. Aldridge, O. Johnson and J. Scarlett. *Group Testing: An Information Theory Perspective*. 2019.

[11] M. Aldridge. 'The Capacity of Bernoulli Nonadaptive Group Testing'. In: *IEEE Transactions on Information Theory* 63.11 (2017), pp. 7142–7148.

[12] M. Aldridge. *Conservative two-stage group testing*. 2020. arXiv: 2005.06617 [stat.AP].

[13] A. Allemann. 'An Efficient Algorithm for Combinatorial Group Testing'. In: *Information Theory, Combinatorics, and Search Theory*. Springer Berlin Heidelberg, 2013, pp. 569–596.

[14] N. Alon and Z. Füredi. 'Spanning subgraphs of random graphs'. In: *Graphs and Combinatorics* 8.1 (1992), pp. 91–94.

[15] E. Arikan. 'Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels'. In: *IEEE Transactions on Information Theory* 55.7 (2009), pp. 3051–3073.

[16] J. Ashkin and E. Teller. 'Statistics of Two-Dimensional Lattices with Four Components'. In: *Physical Review* 64.5-6 (1943), pp. 178–184.

[17] T. Austin. 'Multi-variate correlation and mixtures of product measures'. In: *Kybernetika* (2020), pp. 459–499.

[18] P. Ayre, A. Coja-Oghlan, P. Gao and N. Müller. 'The Satisfiability Threshold For Random Linear Equations'. In: *Combinatorica* 40.2 (2020), pp. 179–235.

[19] J. Balogh, B. Csaba, M. Pei and W. Samotij. 'Large bounded degree trees in expanding graphs'. In: *the electronic journal of combinatorics* 17.1 (2010), p. 6.

[20] J. Balogh, A. Treglown and A. Z. Wagner. 'Tilings in randomly perturbed dense graphs'. In: *Combinatorics, Probability and Computing* 28.2 (2019), pp. 159–176.

[21] V. Bapst and A. Coja-Oghlan. 'Harnessing the Bethe free energy'. In: *Random Structures & Algorithms* 49.4 (2016), pp. 694–741.

[22] F. Barahona. 'On the computational complexity of Ising spin glass models'. In: *Journal of Physics A: Mathematical and General* 15.10 (1982), pp. 3241–3253.

[23] A. Barra, G. Genovese, F. Guerra and D. Tantari. 'About a solvable mean field model of a Gaussian spin glass'. In: *Journal of Physics A: Mathematical and Theoretical* 47.15 (2014), p. 155002.

[24] W. H. Bay, E. Price and J. Scarlett. *Optimal Non-Adaptive Probabilistic Group Testing Requires* $\Theta(\min\{k \ln n, n\})$ *Tests*. 2020. arXiv: 2006.01325 [cs.IT].

[25] M. Bayati, D. Gamarnik and P. Tetali. 'Combinatorial Approach to the Interpolation Method and Scaling Limits in Sparse Random Graphs'. In: *Proceedings of the Forty-Second ACM Symposium on Theory of Computing*. STOC '10. Cambridge, Massachusetts, USA: Association for Computing Machinery, 2010, pp. 105–114.

[26] W. Bedenknecht, J. Han, Y. Kohayakawa and G. O. Mota. 'Powers of tight Hamilton cycles in randomly perturbed hypergraphs'. In: *Random Structures & Algorithms* 55.4 (2019), pp. 795–807.

[27] H. A. Bethe and W. L. Bragg. 'Statistical theory of superlattices'. In: *Proceedings of the Royal Society of London. Series A - Mathematical and Physical Sciences* 150.871 (1935), pp. 552–575.

[28] T. Bohman, A. Frieze and R. Martin. 'How many random edges make a dense graph hamiltonian?' In: *Random Structures & Algorithms* 22.1 (2003), pp. 33–42.

[29] B. Bollobás. 'The Evolution of Random Graphs'. In: *Transactions of the American Mathematical Society* 286.1 (1984), pp. 257–274.

[30] B. Bollobás, C. Borgs, J. T. Chayes, J. H. Kim and D. B. Wilson. 'The Scaling Window of the 2-SAT Transition'. In: *Random Struct. Algorithms* 18.3 (2001), pp. 201–256.

[31] C. Borgs, J. Chayes, L. Lovász, V. Sós and K. Vesztergombi. 'Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing'. In: *Advances in Mathematics* 219.6 (2008), pp. 1801–1851.

[32] C. Borgs, J. Chayes, L. Lovász, V. Sós and K. Vesztergombi. 'Convergent sequences of dense graphs II. Multiway cuts and statistical physics'. In: *Annals of Mathematics* 176.1 (2012), pp. 151–219.

[33] J. Böttcher. 'Large-scale structures in random graphs'. In: *Surveys in Combinatorics* 440 (2017), pp. 87–140.

[34] J. Böttcher, J. Han, Y. Kohayakawa, R. Montgomery, O. Parczyk and Y. Person. 'Universality for bounded degree spanning trees in randomly perturbed graphs'. In: *Random Structures & Algorithms* (2019).

[35] J. Böttcher, R. Montgomery, O. Parczyk and Y. Person. 'Embedding spanning bounded degree subgraphs in randomly perturbed graphs'. In: *Mathematika* (2019), pp. 1–25.

[36] J. Böttcher, M. Schacht and A. Taraz. 'Proof of the bandwidth conjecture of Bollobás and Komlós'. In: *Mathematische Annalen* 343.1 (2009), pp. 175–205.

[37] A. Braunstein, M. Mézard, M. Weigt and R. Zecchina. 'Constraint Satisfaction by Survey Propagation'. In: *Advances in Neural Information Processing Systems* 9 (2005), p. 424.

[38] A. Braunstein, M. Mézard and R. Zecchina. 'Survey propagation: An algorithm for satisfiability'. In: *Random Structures & Algorithms* 27.2 (2005), pp. 201–226.

[39] C. L. Chan, S. Jaggi, V. Saligrama and S. Agnihotri. 'Non-Adaptive Group Testing: Explicit Bounds and Novel Algorithms'. In: *IEEE Transactions on Information Theory* 60.5 (2014), pp. 3019–3035.

[40] V. Chvatal and B. Reed. 'Mick gets some (the odds are on his side) (satisfiability)'. In: *Proceedings., 33rd Annual Symposium on Foundations of Computer Science*. 1992, pp. 620–627.

[41] A. Coja-Oghlan, O. Gebhard, M. Hahn-Klimroth and P. Loick. 'Information-Theoretic and Algorithmic Thresholds for Group Testing'. In: *IEEE Transactions on Information Theory* 66.12 (2020), pp. 7911–7928.

[42] A. Coja-Oghlan, W. Perkins and K. Skubch. 'Limits of discrete distributions and Gibbs measures on random graphs'. In: *European Journal of Combinatorics* 66 (2017), pp. 37–59.

[43] A. Coja-Oghlan, C. Efthymiou and S. Hetterich. 'On the chromatic number of random regular graphs'. In: *Journal of Combinatorial Theory, Series B* 116 (2016), pp. 367–439.

[44] A. Coja-Oghlan, O. Gebhard, M. Hahn-Klimroth and P. Loick. 'Information-Theoretic and Algorithmic Thresholds for Group Testing'. In: *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)* (2019), 43:1–43:14.

[45] A. Coja-Oghlan, O. Gebhard, M. Hahn-Klimroth and P. Loick. 'Optimal Group Testing'. In: *Proceedings of 33rd Conference on Learning Theory* (2020), pp. 1374–1388.

[46] A. Coja-Oghlan, O. Gebhard, M. Hahn-Klimroth and P. Loick. 'Optimal group testing'. In: *Combinatorics, Probability and Computing* (2021), pp. 1–38.

[47] A. Coja-Oghlan and M. Hahn-Klimroth. *The cut metric for probability distributions*. 2020. arXiv: 1905.13619 [math.CO].

[48] A. Coja-Oghlan, M. Hahn-Klimroth, P. Loick, N. Müller, K. Panagiotou and M. Pasch. *Inference and mutual information on random factor graphs*. 2020. arXiv: 2007.07494 [cs.DM].

[49] A. Coja-Oghlan, F. Krzakala, W. Perkins and L. Zdeborová. 'Information-theoretic thresholds from the cavity method'. In: *Advances in Mathematics* 333 (2018), pp. 694–795.

[50] A. Coja-Oghlan, N. Müller and J. B. Ravelomanana. *Belief Propagation on the random k-SAT model*. 2020. arXiv: 2011.02303 [math.PR].

[51] A. Coja-Oghlan and K. Panagiotou. 'The asymptotic k-SAT threshold'. In: *Advances in Mathematics* 288 (2016), pp. 985–1068.

[52] A. Coja-Oghlan and W. Perkins. 'Belief propagation on replica symmetric random factor graph models'. In: *Annales de l'Institut Henri Poincaré D* 5.2 (2018), pp. 211–249.

[53] A. Coja-Oghlan and W. Perkins. 'Bethe States of Random Factor Graphs'. In: *Communications in Mathematical Physics* 366.1 (2019), pp. 173–201.

[54] A. Coja-Oghlan and N. Wormald. 'The Number of Satisfying Assignments of Random Regular k-SAT Formulas'. In: *Combinatorics, Probability and Computing* 27.4 (2018), pp. 496–530.

[55] A. Coja-Oghlan and L. Zdeborová. 'The condensation transition in random hypergraph 2-coloring'. In: *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms* (2012).

[56] D. Conlon and J. Fox. 'Bounds for graph regularity and removal lemmas'. In: *Geometric and Functional Analysis* 22.5 (2012), pp. 1191–1256.

[57] M. Cuturi, O. Teboul, Q. Berthet, A. Doucet and J.-P. Vert. *Noisy Adaptive Group Testing using Bayesian Sequential Experimental Design*. 2020. arXiv: 2004.12508 [stat.ME].

[58] A. G. D'yachkov, I. V. Vorob'ev, N. A. Polyansky and V. Y. Shchukin. 'Bounds on the rate of disjunctive codes'. In: *Problems of Information Transmission* 50.1 (2014), pp. 27–56.

[59] P. Damaschke. 'Threshold Group Testing'. In: *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2006, pp. 707–718.

[60] P. Damaschke and A. S. Muhammad. 'Randomized Group Testing Both Query-Optimal and Minimal Adaptive'. In: *SOFSEM 2012: Theory and Practice of Computer Science*. Springer Berlin Heidelberg, 2012, pp. 214–225.

[61] G. B. Dantzig. 'Maximization of a Linear Function of Variables Subject to Linear Inequalities'. In: *Activity Analysis of Production and Allocation, Cowles Commission Monograph* 13 (1951).

[62] R. David and U. Feige. 'On the effect of randomness on planted 3-coloring models'. In: *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*. Ed. by D. Wichs and Y. Mansour. ACM, 2016, pp. 77–90.

[63] P. Diaconis and S. Janson. 'Graph limits and exchangeable random graphs'. In: *Rendiconti di Matematica e delle sue Applicazioni* 28 (2008).

[64] J. Ding, A. Sly and N. Sun. 'Proof of the Satisfiability Conjecture for Large k'. In: *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*. STOC '15. Portland, Oregon, USA: Association for Computing Machinery, 2015, pp. 59–68.

[65] G. A. Dirac. 'Some Theorems on Abstract Graphs'. In: *Proceedings of the London Mathematical Society* s3-2.1 (1952), pp. 69–81.

[66] D. L. Donoho, A. Maleki and A. Montanari. 'Message passing algorithms for compressed sensing: I. motivation and construction'. In: *2010 IEEE Information Theory Workshop on Information Theory (ITW 2010, Cairo)*. 2010, pp. 1–5.

[67] D. Donoho, A. Maleki and A. Montanari. 'Message Passing Algorithms for Compressed Sensing'. In: *Proceedings of the National Academy of Sciences of the United States of America* 106 (2009), pp. 18914–9.

[68] R. Dorfman. 'The Detection of Defective Members of Large Populations'. In: *Ann. Math. Statist.* 14.4 (1943), pp. 436–440.

[69] S. F. Edwards and P. W. Anderson. 'Theory of spin glasses'. In: *Journal of Physics F: Metal Physics* 5.5 (1975), pp. 965–974.

[70] A. El Alaoui, A. Ramdas, F. Krzakala, L. Zdeborová and M. I. Jordan. 'Decoding From Pooled Data: Phase Transitions of Message Passing'. In: *IEEE Transactions on Information Theory* 65.1 (2019), pp. 572–585.

[71] R. S. Ellis and C. M. Newman. 'The statistics of Curie-Weiss models'. In: *Journal of Statistical Physics* 19.2 (1978), pp. 149–161.

[72] P. Erdős and A. Rényi. 'On the existence of a factor of degree one of a connected random graph'. In: *Acta Mathematica Hungarica* 17.3-4 (1966), pp. 359–368.

[73] P. Erdős and A. Rényi. 'On Two Problems of Information Theory'. In: *Magyar Tud. Akad. Mat. Kutató Int. Közl* 8 (1963), pp. 229–243.

[74] U. Feige, S. Goldwasser, L. Lovász, S. Safra and M. Szegedy. 'Interactive Proofs and the Hardness of Approximating Cliques'. In: *J. ACM* 43.2 (1996), pp. 268–292.

[75] U. Feige, E. Mossel and D. Vilenchik. 'Complete Convergence of Message Passing Algorithms for Some Satisfiability Problems'. In: *Theory of Computing* 9.19 (2013), pp. 617–651.

[76] A. Ferber, K. Luh and O. Nguyen. 'Embedding large graphs into a random graph'. In: *Bulletin of the London Mathematical Society* 49.5 (2017), pp. 784–797.

[77] A. Ferber and R. Nenadov. 'Spanning universality in random graphs'. In: *Random Structures & Algorithms* 53.4 (2018), pp. 604–637.

[78] W. Fernandez de la Vega. 'Random 2-SAT: results and problems'. In: *Theoretical Computer Science* 265.1 (2001). Phase Transitions in Combinatorial Problems, pp. 131–146.

[79] E. Fischer, A. Matsliah and A. Shapira. 'Approximate Hypergraph Partitioning and Applications'. In: *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*. 2007, pp. 579–589.

[80] P. Fischer, N. Klasner and I. Wegener. 'On the cut-off point for combinatorial group testing'. In: *Discrete Applied Mathematics* 91.1 (1999), pp. 83–92.

[81] B. J. Frey, F. R. Kschischang, H.-a. Loeliger and N. Wiberg. 'Factor Graphs and Algorithms'. In: *Proc. 35th Allerton Conf. on Communications, Control, and Computing, (Allerton)*. 1998, pp. 666–680.

[82] E. Friedgut and J. Bourgain. 'Sharp thresholds of graph properties, and the k-sat problem'. In: *Journal of the American Mathematical Society* 12.4 (1999), pp. 1017–1054.

[83] A. Frieze. 'On large matchings and cycles in sparse random graphs'. In: *Discrete Mathematics* 59.3 (1986), pp. 243–256.

[84] A. Frieze and R. Kannan. 'Quick Approximation to Matrices and Applications'. In: *Combinatorica* 19.2 (1999), pp. 175–220.

[85] T. Funke and T. Becker. 'Stochastic block models: A comparison of variants and inference methods'. In: *PLOS ONE* 14.4 (2019), pp. 1–40.

[86] V. Gandikota, E. Grigorescu, S. Jaggi and S. Zhou. 'Nearly Optimal Sparse Group Testing'. In: *IEEE Transactions on Information Theory* 65.5 (2019), pp. 2760–2773.

[87] E. Gardner and B. Derrida. 'Three unfinished works on the optimal storage capacity of networks'. In: *Journal of Physics A: Mathematical and General* 22.12 (1989), pp. 1983–1994.

[88] O. Gebhard, M. Hahn-Klimroth, O. Parczyk, M. Penschuck, M. Rolvien, J. Scarlett and N. Tan. *Near optimal sparsity-constrained group testing: improved bounds and algorithms*. 2020. arXiv: 2004.11860 [cs.DS].

[89] O. Gebhard, O. Johnson, P. Loick and M. Rolvien. *Improved bounds for noisy group testing with constant tests per item*. 2020. arXiv: 2007.01376 [cs.IT].

[90] E. N. Gilbert. 'Random Graphs'. In: *The Annals of Mathematical Statistics* 30.4 (1959), pp. 1141–1144.

[91] A. Goerdt. 'A Threshold for Unsatisfiability'. In: *Journal of Computer and System Sciences* 53.3 (1996), pp. 469–486.

[92] W. Gowers. 'Lower bounds of tower type for Szemerédi's uniformity lemma'. In: *Geometric and Functional Analysis* 7.2 (1997), pp. 322–337.

[93] M. Hahn-Klimroth, G. S. Maesaka, Y. Mogge, S. Mohr and O. Parczyk. *Random perturbation of sparse graphs*. 2020. arXiv: 2004.04672 [math.CO].

[94] A. Hajnal and E. Szemerédi. 'Proof of a conjecture of P. Erdős'. In: *Combinatorial theory and its applications* 2 (1970), pp. 601–623.

[95] P. W. Holland, K. B. Laskey and S. Leinhardt. 'Stochastic blockmodels: First steps'. In: *Social Networks* 5.2 (1983), pp. 109–137.

[96] D. Hoover. *Relations on Probability Spaces and Arrays of Random Variables*. 1979.

[97] M. C. Hu, F. K. Hwang and J. K. Wang. 'A Boundary Problem for Group Testing'. In: *SIAM Journal on Algebraic Discrete Methods* 2.2 (1981), pp. 81–87.

[98] F. K. Hwang. 'A Method for Detecting All Defective Members in a Population by Group Testing'. In: *Journal of the American Statistical Association* 67.339 (1972), pp. 605–608.

[99] F. Iliopoulos and I. Zadik. *Group testing and local search: is there a computational-statistical gap?* 2020. arXiv: 2011.05258 [math.ST].

[100] E. Ising. 'Beitrag zur Theorie des Ferromagnetismus'. In: *Zeitschrift für Physik* 31.1 (1925), pp. 253–258.

[101] A. Jagannath. 'Approximate Ultrametricity for Random Measures and Applications to Spin Glasses'. In: *Communications on Pure and Applied Mathematics* 70.4 (2017), pp. 611–664.

[102] S. Janson. *Graphons, cut norm and distance, couplings and rearrangements*. Vol. 4. NYJM Monographs. New York Journal of Mathematics, 2013.

[103] S. Janson, T. Luczak, A. Rucinski and R. Rucinski. *Random Graphs*. New York: Wiley, 2000.

[104] E. T. Jaynes. 'Information Theory and Statistical Mechanics'. In: *Phys. Rev.* 106 (4 1957), pp. 620–630.

[105] M. Jerrum. 'Large Cliques Elude the Metropolis Process'. In: *Random Structures and Algorithms* 3.4 (1992), pp. 347–359.

[106] A. Jimenez Felstrom and K. S. Zigangirov. 'Time-varying periodic convolutional codes with low-density parity-check matrix'. In: *IEEE Transactions on Information Theory* 45.6 (1999), pp. 2181–2191.

[107] A. Johansson, J. Kahn and V. Vu. 'Factors in random graphs'. In: *Random Structures & Algorithms* 33.1 (2008), pp. 1–28.

[108] O. Johnson, M. Aldridge and J. Scarlett. 'Performance of Group Testing Algorithms With Near-Constant Tests Per Item'. In: *IEEE Transactions on Information Theory* 65.2 (2019), pp. 707–723.

[109] R. M. Karp. 'Reducibility among Combinatorial Problems'. In: *Complexity of Computer Computations*. Springer US, 1972, pp. 85–103.

[110] S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi. 'Optimization by Simulated Annealing'. In: *Science* 220.4598 (1983), pp. 671–680.

[111] S. Kirkpatrick, G. Györgyi, N. Tishby and L. Troyansky. 'The Statistical Mechanics of k-Satisfaction'. In: *Advances in Neural Information Processing Systems 6*. Ed. by J. D. Cowan, G. Tesauro and J. Alspector. Morgan-Kaufmann, 1994, pp. 439–446.

[112] J. Komlos, A. Shokoufandeh, M. Simonovits and E. Szemerédi. 'The Regularity Lemma and Its Applications in Graph Theory'. In: vol. 2292. 2000, pp. 84–112.

[113] J. Komlós, G. N. Sárközy and E. Szemerédi. 'On the Pósa-Seymour conjecture'. In: *Journal of Graph Theory* 29.3 (1998), pp. 167–176.

[114] J. Komlós, G. N. Sárközy and E. Szemerédi. 'Proof of the Seymour conjecture for large graphs'. In: *Annals of Combinatorics* 2.1 (1998), pp. 43–60.

[115] J. Komlós, G. N. Sárközy and E. Szemerédi. 'Spanning trees in dense graphs'. In: *Combinatorics, Probability and Computing* 10.5 (2001), pp. 397–416.

[116] J. Komlós and E. Szemerédi. 'Limit distribution for the existence of hamiltonian cycles in a random graph'. In: *Discrete Mathematics* 43.1 (1983), pp. 55–63.

[117] A. D. Korshunov. 'Solution of a problem of Erdős and Rényi on Hamiltonian cycles in nonoriented graphs'. In: *Dokl. Akad. Nauk SSSR* 228 (3 1976), pp. 529–532.

[118] M. Krivelevich. 'Embedding spanning trees in random graphs'. In: *SIAM Journal on Discrete Mathematics* 24.4 (2010), pp. 1495–1500.

[119] M. Krivelevich, M. Kwan and B. Sudakov. 'Bounded-degree spanning trees in randomly perturbed graphs'. In: *SIAM Journal on Discrete Mathematics* 31.1 (2017), pp. 155–171.

[120] M. R. Krom. 'The Decision Problem for a Class of First-Order Formulas in Which all Disjunctions are Binary'. In: *Mathematical Logic Quarterly* 13.1-2 (1967), pp. 15–20.

[121] F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian and L. Zdeborová. 'Gibbs states and the set of solutions of random constraint satisfaction problems'. In: *Proceedings of the National Academy of Sciences* 104.25 (2007), pp. 10318–10323.

[122] F. R. Kschischang, B. J. Frey and H. Loeliger. 'Factor graphs and the sum-product algorithm'. In: *IEEE Transactions on Information Theory* 47.2 (2001), pp. 498–519.

[123] S. Kudekar, T. Richardson and R. L. Urbanke. 'Spatially Coupled Ensembles Universally Achieve Capacity Under Belief Propagation'. In: *IEEE Transactions on Information Theory* 59.12 (2013), pp. 7761–7813.

[124] S. Kudekar, T. J. Richardson and R. L. Urbanke. 'Threshold Saturation via Spatial Coupling: Why Convolutional LDPC Ensembles Perform So Well over the BEC'. In: *IEEE Transactions on Information Theory* 57.2 (2011), pp. 803–834.

[125] D. Kühn and D. Osthus. 'On Pósa's conjecture for random graphs'. In: *SIAM Journal on Discrete Mathematics* 26.3 (2012), pp. 1440–1457.

[126] M. Lelarge. 'Bypassing correlation decay for matchings with an application to XORSAT'. In: *2013 IEEE Information Theory Workshop (ITW)*. 2013, pp. 1–5.

[127] L. Lovász. *Large networks and graph limits*. Providence, Rhode Island: American Mathematical Society, 2012.

[128] L. Lovász and B. Szegedy. 'Limits of Dense Graph Sequences'. In: *J. Comb. Theory Ser. B* 96.6 (2006), pp. 933–957.

[129] A. Maleki and A. Montanari. 'Analysis of approximate message passing algorithm'. In: *2010 44th Annual Conference on Information Sciences and Systems (CISS)*. 2010, pp. 1–7.

[130] S. Mallapaty. 'The mathematical strategy that could transform coronavirus testing'. In: *Nature* 583.7817 (2020), pp. 504–505.

[131] E. Marinari, G. Parisi, F. Ricci-Tersenghi, J. J. Ruiz-Lorenzo and F. Zuliani. In: *Journal of Statistical Physics* 98.5/6 (2000), pp. 973–1074.

[132] M. Mezard and C. Toninelli. 'Group Testing With Random Pools: Optimal Two-Stage Algorithms'. In: *Information Theory, IEEE Transactions on* 57 (2011), pp. 1736–1745.

[133] M. Mézard, M. Tarzia and C. Toninelli. 'Statistical physics of group testing'. In: *Journal of Physics: Conference Series* 95 (2008), p. 012019.

[134] M. Mézard and G. Parisi. 'The Bethe lattice spin glass revisited'. In: *The European Physical Journal B* 20.2 (2001), pp. 217–233.

[135] M. Mézard and A. Montanari. *Information, Physics, and Computation*. Oxford University Press, Inc., 2009.

[136] M. Molloy. 'Models and thresholds for random Constraint Satisfaction Problems.' In: *Conference Proceedings of the Annual ACM Symposium on Theory of Computing* (2002).

[137] A. Montanari. 'Estimating random variables from random sparse observations'. In: *European Transactions on Telecommunications* 19.4 (2008), pp. 385–403.

[138] A. Montanari, F. Ricci-Tersenghi and G. Semerjian. 'Clusters of solutions and replica symmetry breaking in random k-satisfiability'. In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.04 (2008), P04004.

[139] R. Montgomery. 'Spanning trees in random graphs'. In: *Advances in Mathematics* 356 (2019), p. 106793.

[140] T. Mora and L. Zdeborová. 'Random Subcubes as a Toy Model for Constraint Satisfaction Problems'. In: *Journal of Statistical Physics* 131.6 (2008), pp. 1121–1138.

[141] R. Nenadov and N. Škorić. 'Powers of Hamilton cycles in random graphs and tight Hamilton cycles in random hypergraphs'. In: *Random Structures & Algorithms* 54.1 (2019), pp. 187–208.

[142] D. Panchenko and M. Talagrand. 'Bounds for diluted mean-fields spin glass models'. In: *Probability Theory and Related Fields* 130.3 (2004), pp. 319–336.

[143] G. Parisi. 'A sequence of approximated solutions to the S-K model for spin glasses'. In: *Journal of Physics A: Mathematical and General* 13.4 (1980), pp. L115–L121.

[144] G. Parisi, F. Ricci-Tersenghi and T. Rizzo. 'Diluted Mean-Field Spin-Glass Models at Criticality'. In: *Journal of Statistical Mechanics: Theory and Experiment* 2014 (2014).

[145] J. Pearl. 'Reverend Bayes on Inference Engines: A Distributed Hierarchical Approach'. In: *Proceedings of the Second AAAI Conference on Artificial Intelligence*. AAAI'82. Pittsburgh, Pennsylvania: AAAI Press, 1982, pp. 133–136.

[146] J. Pearl. 'Chapter 4 - Belief Updating by Network Propagation'. In: *Probabilistic Reasoning in Intelligent Systems*. Ed. by J. Pearl. San Francisco (CA): Morgan Kaufmann, 1988, pp. 143–237.

[147] L. Pósa. 'Hamiltonian circuits in random graphs'. In: *Discrete Mathematics* 14.4 (1976), pp. 359–364.

[148] P. Raghavendra and N. Tan. 'Approximating CSPs with Global Cardinality Constraints Using SDP Hierarchies'. In: *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms* (2011).

[149] L. Riccio and C. J. Colbourn. 'Sharper Bounds in Adaptive Group Testing'. In: *Taiwanese Journal of Mathematics* 4.4 (2000), pp. 669–673.

[150] O. Riordan. 'Spanning subgraphs of random graphs'. In: *Combinatorics, Probability and Computing* 9.2 (2000), pp. 125–148.

[151] V. Rödl and M. Schacht. 'Regularity Lemmas for Graphs'. In: *Fete of Combinatorics and Computer Science*. Springer Berlin Heidelberg, 2010, pp. 287–325.

[152] J. Scarlett and V. Cevher. 'Limits on support recovery with probabilistic models: An information-theoretic framework'. In: *2015 IEEE International Symposium on Information Theory (ISIT)*. 2015, pp. 2331–2335.

[153] J. Scarlett. 'Noisy Adaptive Group Testing: Bounds and Algorithms'. In: *IEEE Transactions on Information Theory* 65.6 (2019), pp. 3646–3661.

[154] J. Scarlett and V. Cevher. 'Phase Transitions in Group Testing'. In: *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics, 2015.

[155] J. Scarlett and V. Cevher. 'Phase Transitions in the Pooled Data Problem'. In: *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems*. 2017, pp. 376–384.

[156] G. Semerjian and R. Monasson. 'A Study of Pure Random Walk on Random Satisfiability Problems with "Physical" Methods'. In: *Theory and Applications of Satisfiability Testing*. Springer Berlin Heidelberg, 2004, pp. 120–134.

[157] J. M. L. Sengers. 'Mean-field theories, their weaknesses and strength'. In: *Fluid Phase Equilibria* 158-160 (1999), pp. 3–17.

[158] D. Sherrington and S. Kirkpatrick. 'Solvable Model of a Spin-Glass'. In: *Phys. Rev. Lett.* 35 (26 1975), pp. 1792–1796.

[159] D. A. Spielman and S.-H. Teng. 'Smoothed Analysis of Algorithms: Why the Simplex Algorithm Usually Takes Polynomial Time'. In: *J. ACM* 51.3 (2004), pp. 385–463.

[160] E. Szemerédi. 'On sets of integers containing no k elements in arithmetic progression.' In: *Acta Arithmetica* 27 (1975), pp. 199–245.

[161] M. Talagrand. 'Multiple levels of symmetry breaking'. In: *Probability Theory and Related Fields* 117.4 (2000), pp. 449–466.

[162] M. Talagrand. 'The Parisi Formula'. In: *Annals of Mathematics* 163.1 (2006), pp. 221–263.

[163] T. Tao. *Szemeredi's regularity lemma via random partitions*. 2019. URL: https://terrytao.wordpress.com/2009/04/26/szemeredis-regularity-lemma-via-random-partitions/ (visited on 30/11/2020).

[164] P. Ungar. 'The cutoff point for group testing'. In: *Communications on Pure and Applied Mathematics* 13.1 (1960), pp. 49–54.

[165] L. G. Valiant. 'The Complexity of Enumeration and Reliability Problems'. In: *SIAM Journal on Computing* 8.3 (1979), pp. 410–421.

[166] L. Viana and A. J. Bray. 'Phase diagrams for dilute spin glasses'. In: *Journal of Physics C: Solid State Physics* 18.15 (1985), pp. 3037–3051.

[167] E. Vincent and V. Dupuis. 'Spin Glasses: Experimental Signatures and Salient Outcomes'. In: *Frustrated Materials and Ferroic Glasses*. Springer International Publishing, 2018, pp. 31–56.

[168] L. Zdeborová and F. Krzakala. 'Statistical physics of inference: thresholds and algorithms'. In: *Advances in Physics* 65.5 (2016), pp. 453–552.

[169] L. Zdeborová and M. Mézard. ‘Constraint satisfaction problems with isolated solutions are hard’. In: *Journal of Statistical Mechanics: Theory and Experiment* 2008.12 (2008), P12004.

[170] Y. Zhang. *Phase Transitions of Random Constraints Satisfaction Problem*. UC Berkley, 2017.

A. Contained publications and the author's contributions

This section provides a detailed overview about the author's contributions as well as the publication status of each of those manuscripts. We will explicitly only state the contributions of this thesis's author (MHK). Therefore, the papers might contain results achieved by different authors whose contributions are not discussed below. For the sake of readability, we will abbreviate all author's names to their initials.

Information-theoretic and algorithmic thresholds for group testing This manuscript by A. Coja-Oghlan, O. Gebhard, M. Hahn-Klimroth and P. Loick appeared in the *IEEE Transactions on Information Theory* [41] and a short version appeared in the *Proceedings of the 46th ICALP* [44].

While the idea of this paper was found during the master's thesis of OG (supervised by MHK and PL), the further main contributions of MHK are the development and formalisation of Theorem 1.1 and its proof in joint work with PL as well as the formalisation of the proof of Theorem 1.2 based on an idea of PL.

Optimal Group Testing A short version of this article by A. Coja-Oghlan, O. Gebhard, M. Hahn-Klimroth and P. Loick appeared in the *Proceedings of 33rd Conference on Learning Theory (COLT)* [45] and the full version is accepted for publication at *Combinatorics, Probability and Computing*.

The algorithmic achievability result with respect to adaptive group testing (Theorem 1.3) was formally proven by MHK and PL while the corresponding non-adaptive result (Theorem 1.2) is joint work of all authors. The main technical contribution to Theorem 1.2, thus the derivation of the correct weights as well as the formal proof are due to ACO, MHK and PL. Furthermore, the proof idea of Theorem 1.1 based on a generalised argument of Aldridge was discussed by MHK and PL while the formal derivation, i.e. the reduction on a different prevalence is due to ACO, MHK and PL. Finally, all authors contributed to the writing of the manuscript.

Near-Optimal Sparsity-Constrained Group Testing: Improved Bounds and Algorithms This paper by O. Gebhard, M. Hahn-Klimroth, O. Parczyk, M. Penschuck, M. Rolvien, J. Scarlett and N. Tan is under review at *IEEE Transactions on Information Theory*.

To be more precise, OG, MHK, OP, MP, MR uploaded a draft on non-adaptive sparsity constrained group testing to arXiv and contemporaneously JS and NT published a preprint on adaptive methods. As those contributions could be perfectly merged all authors decided to extend their results and combine them.

While the writing of the manuscript was led by OG, MHK and JS, the development and formal elaboration of the non-adaptive converse bounds with respect to information theory (Theorem 3.2 and Theorem 4.1) are joint equal work of MHK and OP. A further contribution of MHK is the development of the proof idea of Theorem 3.4 (`DD` converse in the $\Delta-$divisible model) and its formal justification was joint work of MHK and OP. The corresponding achievability statement (Theorem 3.3) is joint work of OG and MHK. Moreover, the achievability statement for `DD` in the $\Gamma-$sparse model is joint work of OG and MHK (Theorem 4.10) and MHK and OP (Theorem 4.18) respectively.

The number of satisfying assignments of random 2-SAT formulas. This paper is joint work of D. Achlioptas, A. Coja-Oghlan, M. Hahn-Klimroth, J. Lee, N. Müller, M. Penschuck and G. Zhou and it appeared in *Random Structures & Algorithms*. The current arXiv version is entitled *The random 2-SAT partition function*. The main contribution of MHK is the intuitive description of how to obtain a worst-case boundary condition in order to prove that the Boltzmann distribution is a Bethe state while its formal justification is due to NM. A further contribution of MHK are some technical derivations in order

to extend the positive temperature concentration result to the zero temperature limit while the main idea of this approach and the detailed proof were provided by ACO.

The cut metric for probability distributions This article by A. Coja-Oghlan and M. Hahn-Klimroth is accepted for publication at the *SIAM Journal of Discrete Mathematics*. It is a joint project of MHK with his PhD advisor ACO. While the latter led the writing of the manuscript, the theorems and their proofs were obtained and discussed jointly.

Random perturbation of sparse graphs This manuscript by M. Hahn-Klimroth, G. Maesaka, Y. Mogge, S. Mohr and O. Parczyk is accepted at the *Electronic Journal of Combinatorics*.

While the idea for the paper was brought by OP to a workshop on extremal and probabilistic combinatorics, Theorems 1.1 and 1.2 which establish a sufficient condition for observing a Hamilton cycle and a perfect matching in $\mathcal{G}(n,\beta/n)\cup\mathcal{G}_\alpha$ were discussed intensively by OP and MHK and formally proven and written down by MHK.