Spectral Detection in the Censored Block Model

Alaa Saade
Laboratoire de Physique Statistique
École Normale Supérieure, 24 Rue Lhomond
Paris 75005

Florent Krzakala
Sorbonne Universités, UPMC Univ. Paris 06
Laboratoire de Physique Statistique, CNRS UMR 8550
École Normale Supérieure, 24 Rue Lhomond, Paris

Marc Lelarge
INRIA and École Normale Supérieure
Paris, France

Lenka Zdeborová
Institut de Physique Théorique
CEA Saclay and URA 2306,
CNRS 91191 Gif-sur-Yvette, France.

***Abstract*—We consider the problem of partially recovering hidden binary variables from the observation of (few) censored edge weights, a problem with applications in community detection, correlation clustering and synchronization. We describe two spectral algorithms for this task based on the non-backtracking and the Bethe Hessian operators. These algorithms are shown to be asymptotically optimal for the partial recovery problem, in that they detect the hidden assignment as soon as it is information theoretically possible to do so.**

A. Introduction

In many inference problems, the available data can be represented on a weighted graph. Given the knowledge of the edge weights, the task is to infer latent variables carried by the nodes. Here, we shall consider the problem of recovering binary node labels from censored edge measurements [1], [2]. Specifically, given an Erdős-Rényi random graph $G = (V, E) \in \mathcal{G}(n, \alpha/n)$ with n nodes carrying latent variables $\sigma_i = \pm 1,\ 1 \leq i \leq n$, we draw the edge labels $J_{ij} = \pm 1,\ (ij) \in E$ from the following distribution:

$$P(J_{ij}|\sigma_i, \sigma_j) = (1-\epsilon)\mathbf{1}(J_{ij} = \sigma_i\sigma_j) + \epsilon\mathbf{1}(J_{ij} = -\sigma_i\sigma_j)\,, \quad (1)$$

where ϵ is a noise parameter. In the noiseless case $\epsilon = 0$, we have $\sigma_i\sigma_j = J_{ij}$ and one can easily recover the communities in each connected component along a spanning tree. When $\epsilon = 1/2$, on the other hand, the graph doesn't contain any information about the latent variables σ_i, and recovery is impossible. What happens in between? The problem of *exactly* recovering the latent variables σ_i has been studied in [1]. It turns out that, asymptotically in the large n limit, exact recovery is shown to be possible if and only if

$$\alpha > \alpha_{\text{exact}} = \frac{2\log n}{(1-2\epsilon)^2}\,, \quad (2)$$

where α is the average degree of the graph. Note that the variable of an isolated vertex cannot be recovered so that the average degree has to grow at least like $\log n$, as in the Coupon collector's problem, to ensure that the graph is connected.

We consider in this paper the case where the average degree α will remain fixed as n tends to infinity. In this setting, we cannot ask for exact recovery and we consider here a different question: is it possible to infer an assignment $\hat{\sigma}_i$ of the latent variables that is *positively correlated* with the planted variables σ_i? We call positively correlated an assignment $\hat{\sigma}_i$ such that the following quantity, called *overlap*, is strictly positive:

$$2\left[\max\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}(\hat{\sigma}_i = \sigma_i), \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}(\hat{\sigma}_i = -\sigma_i)\right) - \frac{1}{2}\right]\,. \quad (3)$$

In the limit $n \to \infty$, this overlap vanishes for a random guess $\hat{\sigma}_i$, and is equal to unity if the recovery is exact. We will refer to the task of finding a positively correlated assignment $\hat{\sigma}_i$ as *partial recovery*. This task has been shown [3], [4] to be possible only if

$$\alpha > \alpha_{\text{detect}} = \frac{1}{(1-2\epsilon)^2}. \quad (4)$$

To the best of our knowledge, there is no rigorous proof that this bound is also sufficient. In [3], the same authors also showed that belief propagation (BP) allows to saturate this bound. However, there is no rigorous analysis of BP for this problem and the fact that condition (4) is necessary and sufficient was left as a conjecture in [3] and only the necessary part was proved in [4]. Moreover, from a practical point of view, BP requires the knowledge of the noise parameter ϵ.

In this contribution, we describe two simple spectral algorithms and we show rigorously that they are optimal, in the sense that they can perform partial recovery as soon as $\alpha > \alpha_{\text{detect}}$. Additionally, the output of these algorithms is shown numerically to have an overlap similar to that of BP, without requiring the knowledge of the noise parameter ϵ. This closes the gap from [3], [4], where spectral methods are introduced that succeed only if the connectivity is significantly larger than the threshold (4). The resulting algorithms are thus fast, trivial to implement, and asymptotically optimal.

B. Motivation and Related work

There are various interpretations and models that connect to this problem such as i) Community detection [2]: we try to recover the community membership of the nodes based on noisy (or censored) observations about their relationship; ii) Correlation clustering [5]: we try to cluster the graph G by minimizing the number of "disagreeing edges" ($J_{ij} = -1$) in each cluster. These examples, and others such as synchronisation, are discussed in details in [1].

The inspiration for the present contribution comes from recent developments in the problem of detecting communities

in the (sparse) stochastic block model. The threshold for partial recovery in the stochastic block model was conjectured in [6] and proved in [7]–[9]. Optimal spectral methods, based on the same operators as the algorithms introduced here, were proposed in [10], [11]. These operators were in particular shown to be much better suited to very sparse graphs than the traditional adjacency or Laplacian operators.

Interestingly, this problem first appeared in statistical physics. Indeed, the posterior distribution corresponding to eq. (1) reads, using $\beta_0 = \frac{1}{2}\log\frac{1-\epsilon}{\epsilon}$

$$P(\sigma|J) = \frac{e^{\beta_0 \sum\limits_{(ij)\in E} J_{ij}\sigma_i\sigma_j}}{\mathcal{Z}_J}\,. \tag{5}$$

This is nothing but the spin glass [12] problem where the couplings J_{ij} are correlated with the "planted" configuration σ_i [2], [13]. Such problems can also be shown to be equivalent to spin glasses on the so-called Nishimori line [14], [15]. With these notations, the detection condition (4) corresponds to the well-known spin glass transition [16], [17] at $\sqrt{\alpha_{\rm detect}}\tanh\beta_0 = 1$. In this spin glass context, [18] already conjectured that a spectral algorithm based on the non-backtracking operator (see sec. I-A) was optimal.

C. Outline and main results

In section I, we describe two spectral algorithms that achieve the threshold (4). These algorithms are based on two linear operators: the non-backtracking operator introduced in [10], and the Bethe Hessian introduced in [11]. We further illustrate their properties by showing the results of numerical experiments. In section II, we list the spectral properties of the non-backtracking operator that are relevant to the present context. Finally, we discuss the properties of the Bethe Hessian and its relation with the non-backtracking operator in section III and discuss its connection with the Bethe free energy.

I. Spectral algorithms

A. The non-backtracking operator

The non-backtracking operator acts on the directed edges $i \to j$ of the graph as

$$\mathrm{B}_{i\to j,k\to\ell} = J_{k\ell}\mathbf{1}(j=k)\mathbf{1}(i\neq\ell)\,. \tag{6}$$

It is therefore represented by a $2m\times 2m$ matrix, where m is the number of edges in the graph. As discussed in [10], [18] the motivation for using this operator is that it corresponds to the linear approximation of belief propagation for this problem around the so-called uninformative fixed point of BP.

Similarly to [10], one can show (see Sec. III for details) that the eigenvalues of B that are different from ± 1 form the spectrum of the simpler $2n\times 2n$ matrix

$$\mathrm{B}' = \begin{pmatrix} 0 & D-\mathbb{1} \\ -\mathbb{1} & J \end{pmatrix}, \tag{7}$$

where $\mathbb{1}$ is the $n\times n$ identity matrix, D is the diagonal matrix defined by $D_{ii} = d_i$, where d_i is the degree of node i, and J has entries equal to the edge weights J_{ij}. Furthermore, if $(\lambda\neq\pm1, v\in\mathbb{R}^{2m})$ is an eigenpair of B, then $(\lambda, v'\in\mathbb{R}^{2n})$ is an eigenpair of B′ if

$$v'_{n+i} = \sum_{j\in\partial i} v_{j\to i}, \qquad \forall 1\le i\le n, \tag{8}$$

$$\lambda v'_i = (d_i-1)v'_{n+i}\,, \tag{9}$$

where ∂i and d_i are the set of neighbors and the degree of node i. We will therefore favor using B′. The algorithm is then as follows: given a graph with edge weights J_{ij},

Algorithm 1

1) build the matrix B′
2) compute its leading eigenvalue λ_1 (with largest magnitude), and its corresponding eigenvector $v' = \{v'_i\}$.
3) if $\lambda_1\in\mathbb{R}$ and $\lambda_1 > \sqrt{\alpha}$, where α is the average degree of the graph, set $\hat{x}_i = \mathrm{sign}(v'_{n+i})$. Otherwise, raise an error.

Theorem 1 ensures that whenever (4) holds, this algorithm outputs an assignment $\hat{x}_i$ that is positively correlated with the planted latent variables x_i.

B. The Bethe Hessian

Another operator closely related to the non-backtracking operator was introduced in [11]. This operator, called the Bethe Hessian, is an $n\times n$ real and symmetric matrix defined as

$$\mathrm{H} = (\alpha^2-1)\mathbb{1} - \sqrt{\alpha}J + D\,, \tag{10}$$

where D is the diagonal matrix of vertex degrees. Based on this operator, we propose the following algorithm: given a graph with edge weights J_{ij},

Algorithm 2

1) build the Bethe Hessian H
2) compute its (algebraically) smallest eigenvalue λ, and its corresponding eigenvector v.
3) if $\lambda < 0$, set $\hat{x}_i = \mathrm{sign}(v_i)$. Otherwise, raise an error.

Justifications for this second algorithm, and its relation with the first one, will be provided in section III. Compared to the first algorithm, this second one is based on a smaller, symmetric matrix, which leads to improved numerical performance and stability. Additionally, in the case of more general edge weights $J_{ij}\neq\pm1$, the reduction of B to a smaller matrix B′ fails, and one has to work with a $2m\times 2m$ matrix. The Bethe Hessian, on the other hand, generalizes easily to arbitrary weights without any loss in scalability [11].

C. Numerical results

Before turning to proofs, we show on figure 1 the numerical performance of our two algorithms, and compare them with the performance of belief propagation ([3], [19]) which is believed to be optimal on such locally tree-like graphs in the sense that it gives, arguably, the Bayes optimal value of the overlap asymptotically. As shown in section II, both algorithms 1 and 2 are able to achieve partial recovery as soon as $\alpha > \alpha_{\rm detect}$, and their overlap is similar to that of BP, though of course strictly smaller. Note again that BP requires the knowledge of ϵ while the two spectral algorithms described here do not, are

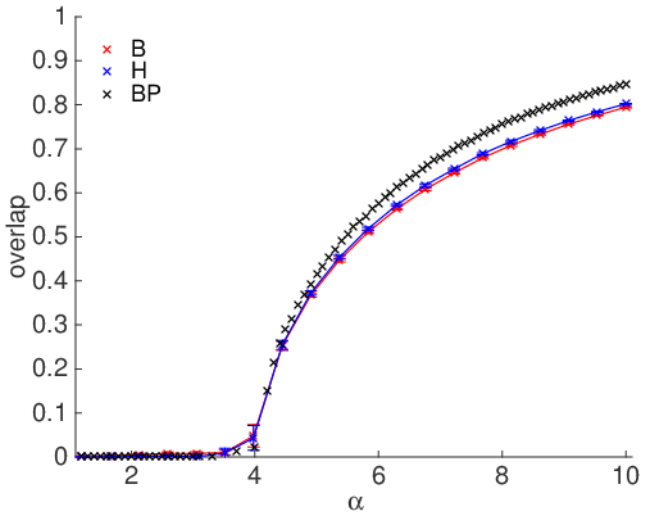

Fig. 1. Overlap as a function of α: comparison between algorithm 1 (based on the non-backtracking operator B), algorithm 2 (based on the Bethe Hessian H), and belief propagation (BP). The noise parameter ϵ is fixed to 0.25 (corresponding to $\alpha_{\rm detect} = 4$), and we vary α. The overlap for B and H is averaged over 20 graphs of size $n = 10^5$. The overlap for BP is estimated asymptotically using the standard method of population dynamics (see for instance [20]), with a population of size 10^4. All three methods output a positively correlated assignment as soon as $\alpha > \alpha_{\rm detect}$. Spectral algorithms 1 and 2 have an overlap similar to that of BP, with the same phase transition, while being simpler and not requiring the knowledge of the parameter ϵ.

trivial to implement, run faster, and avoid the potential non-convergence problem of belief propagation while remaining asymptotically optimal in detecting the hidden assignment. We also observe, empirically, that the overlap given by the Bethe Hessian seems to be always superior to the one provided by the non-backtracking operator.

II. Spectral properties of the non-backtracking operator

In this section, we state results concerning the spectrum of B and show that algorithm 1 outputs an assignment $\hat{\sigma}_i$ that is positively correlated with the planted one, whenever (4) holds.

As already noticed in previous work for the case of an unweighted random graph [10], [21], the superior performance of the non-backtracking operator B is due to the particular shape of its spectrum. In the case of the stochastic block model [22], it decomposes into a bulk of uninformative eigenvalues contained in a disk of radius $\sqrt{\alpha}$ in the complex plane, and a few real and informative eigenvalues outside of the disk. This observation was recently proven in [23], in the case of 2 communities.

The following theorem generalizes this previous result to the present setting and is the main result of this paper.

Theorem 1: Given an Erdős-Rényi random graph with average degree α, variables assigned to vertices $\sigma_i = \pm 1$ uniformly at random independently from the graph and where the edges carry weights sampled from (1), we denote by B the non-backtracking operator defined by (6). and by $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_{2m}|$ the eigenvalues of B in order of decreasing magnitude. Then, with probability tending to 1 as $n \to \infty$, we have:

(i) if $\alpha < \alpha_{\rm detect}$ then $|\lambda_1| \leq \sqrt{\alpha} + o(1)$.

(ii) if $\alpha > \alpha_{\rm detect}$, then $\lambda_1 \in \mathbb{R}$, $\lambda_1 = \alpha(1-2\epsilon)+o(1) > \sqrt{\alpha}$, and $|\lambda_2| \leq \sqrt{\alpha} + o(1)$. Additionally, denoting v the eigenvector associated with λ_1, the following assignment is positively correlated with the planted variables σ_i:

$$\hat{\sigma}_i = \text{sign}\left(\sum_{j \in \partial i} v_{j \to i}\right) .$$

This theorem is illustrated on Fig. 2.

It is then straightforward to show the following:

Corollary 1: The assignment output by Algo. 1 is positively correlated with the planted variables σ_i if and only if

$$\alpha > \alpha_{\rm detect} . \tag{11}$$

We now give a brief sketch of proof for our Theorem 1. The proof relies heavily on the techniques developed in [23]. We try to use notation consistent with [23]: $\vec{E}$ is the set of oriented edges and for any $e = u \to v = (u, v) \in \vec{E}$, we set $e_1 = u$, $e_2 = v$ and $e^{-1} = (v, u)$. For a matrice M, its transpose is denoted by M^*. We start with a simple observation: if t is the vector in $\mathbb{R}^{\vec{E}}$ defined by $t_e = \sigma_{e_2}$ and $\odot$ is the Hadamard product, i.e. $(t \odot x)_e = \sigma_{e_2} x_e$, then we have

$$\mathrm{Bx} = \lambda \mathrm{x} \Leftrightarrow \tilde{\mathrm{B}}(\mathrm{t} \odot \mathrm{x}) = \lambda(\mathrm{t} \odot \mathrm{x}), \tag{12}$$

with $\tilde{\mathrm{B}}$ defined by $\tilde{\mathrm{B}}_{ef} = \mathrm{B}_{\mathrm{ef}} \sigma_{\mathrm{f_1}} \sigma_{\mathrm{f_2}}$. In particular, B an $\tilde{\mathrm{B}}$ have the same spectrum and there is a trivial relation between their eigenvectors. It will be easier to work with $\tilde{\mathrm{B}}$ so to lighten the notation, we will denote (in this section):

$$\mathrm{B}_{ef} = \mathbf{1}(e_2 = f_1)\mathbf{1}(e_1 \neq f_2) P_f,$$

where $P_f = \sigma_{f_1} J_f \sigma_{f_2}$. Note that the random variables P_f are now i.i.d. with $\mathbf{P}(P_f = 1) = 1 - \mathbf{P}(P_f = -1) = 1 - \epsilon$. With this formulation, the problem is said in statistical physics to be "on the Nishimori line" [14], [15].

For the case $(1-2\epsilon)^2\alpha < 1$, the proof is relatively easy. Indeed, from [4], we know that our setting is contiguous to the setting with $\epsilon = 1/2$. In this case, the random variable $P_{i,j}$ are centered and a version of the trace method will allow to upper bound the spectral radius of B. Note however, that one needs to condition on the graph to be ℓ-tangle-free, i.e. such that every neighborhood of radius ℓ contains at most one cycle in order to apply the first moment method.

We now consider the case $(1-2\epsilon)^2\alpha > 1$ and denote by P the linear mapping on $\mathbb{R}^{\vec{E}}$ defined by $(Px)_e = P_e x_{e^{-1}}$ (i.e. the matrix associated to P is $P_{ef} = P_e \mathbf{1}(f = e^{-1})$). Note that $P^* = P$ and since $P_e^2 = 1$, P is an involution so that P is an orthogonal matrix. A simple computation shows that $B^k P = P B^{*k}$, hence $B^k P$ is a symmetric matrix. This symmetry corresponds to the oriented path symmetry in [23] and will be crucial to our analysis.

We also define $\tilde{\alpha} = (1-2\epsilon)\alpha$ and $\chi \in \mathbb{R}^{\vec{E}}$ with $\chi_e = 1$ for all $e \in \vec{E}$. The proof strategy is then similar to Section

5 in [23]. Consider a sequence $\ell \sim \kappa \log_{\tilde{\alpha}} n$ for some small positive κ. Let

$$\varphi = \frac{B^\ell \chi}{\|B^\ell \chi\|}, \quad \theta = \|B^\ell P\varphi\|, \quad \zeta = \frac{B^\ell P\varphi}{\theta}.$$

If $R = B^\ell - \theta\zeta P\varphi^*$ and we can prove that $\|R\|$ is small in comparison with θ, then we can use a theorem on perturbation of eigenvalues and eigenvectors adapted from the Bauer-Fike theorem (see Section 4 in [23]) saying that B^ℓ should have an eigenvalue close to θ.

More precisely, for $y \in \mathbb{R}^{\vec{E}}$ with $\|y\| = 1$, write $y = sP\varphi + x$ with $x \in (P\varphi)^\perp$ and $s \in \mathbb{R}$. Then, we find

$$\|Ry\| = \|B^\ell x + s(B^\ell P\varphi - \theta\zeta)\| \leq \sup_{x:\langle x, P\varphi\rangle = 0, \|x\|=1} \|B^\ell x\|\,.$$

This last quantity can be shown to be upper bounded by $(\log n)^c \alpha^{\ell/2}$ similarly as in Proposition 12 in [23]. Moreover, we can also show that w.h.p.

$$\langle \zeta, P\varphi\rangle \geq c_0, \quad c_0\tilde{\alpha}^\ell \leq \theta \leq c_1\tilde{\alpha}^\ell. \tag{13}$$

These bounds allow to show that B has an eigenvalue $|\lambda_1 - \tilde{\alpha}| = O(1/\ell)$ and that $|\lambda_2| \leq \sqrt{\alpha} + o(1)$.

Note that $\theta = \frac{\|B^\ell B^{*\ell} P\chi\|}{\|B^\ell \chi\|}$, so that we need to compute quantities of the type $\|B^\ell\chi\|$. We now explain the main ideas to compute these quantities. First note that, $(B^\ell\chi)_e$ depends only on the ball of radius ℓ around the edge e. For ℓ not too large, this neighborhood can be coupled with a Galton-Watson branching process with offspring distribution $\mathrm{Poi}(\alpha)$. It is then natural to consider this Poisson Galton-Watson branching process with i.i.d. weights $P_{u,v} \in \{\pm 1\}$ on its edges with mean $1-2\epsilon$. For u in the tree, we denote by $|u|$ its generation and by $Y(u) = \prod_{s=1}^t P_{\gamma_s,\gamma_{s+1}}$ where $\gamma = (\gamma_1, \ldots, \gamma_t)$ is the unique path between the root $o = \gamma_1$ and $u = \gamma_t$. Then $(B^\ell\chi)_e$ is well approximated by:

$$Z_\ell = \sum_{|u|=\ell} Y(u).$$

It is easy to see that $X_t = \frac{Z_t}{\tilde{\alpha}^t}$ is a martingale (with respect to the natural filtration) with zero mean. Moreover we have

$$\begin{aligned}\mathbf{E}\left[Z_t^2\right] &= \mathbf{E}\left[\sum_{u,v:|u|=|v|=t} Y(u)Y(v)\right] \\ &= \sum_{i=0}^t \alpha^{t-i}(1-2\epsilon)^{2i}\alpha^{2i} = O\left(\tilde{\alpha}^{2t}\right),\end{aligned}$$

where the last equality is valid only if $(1-2\epsilon)^2\alpha > 1$. So in this case, we have $\mathbf{E}\left[X_t^2\right] = O(1)$ and the martingale X_t converges a.s. and in L^2 to a limiting random variable $X(\infty)$ with mean one. Following the argument as in [23], this reasoning leads to (13).

We now consider the eigenvector associated with λ_1. It follows from Bauer-Fike theorem (see Section 4 in [23]) that the eigenvector x associated to λ_1 is asymptotically aligned with $\frac{B^\ell B^{*\ell} P\chi}{\|B^\ell B^{*\ell} P\chi\|}$. Thanks to the coupling with the branching process, we can prove that $\|B^\ell B^{*\ell} P\chi\| \approx \tilde{\alpha}^{2\ell}$ and moreover, we have for $e \in \vec{E}$,

$$\frac{(B^\ell B^{*\ell} P\chi)_e}{\tilde{\alpha}^{2\ell}} \approx \frac{\tilde{\alpha}}{\alpha(1-2\epsilon)^2 - 1} X(\infty), \tag{14}$$

where $X(\infty)$ is the limit of the martingale defined above and has mean one. We can now translate this result to the eigenvector of the original non-backtracking operator thanks to (12): $v_e = \sigma_{e_2} x_e$ where x_e is approximated by (14). In particular, we see that $\sum_{e,e_2=v} v_e$ is correlated with σ_v.

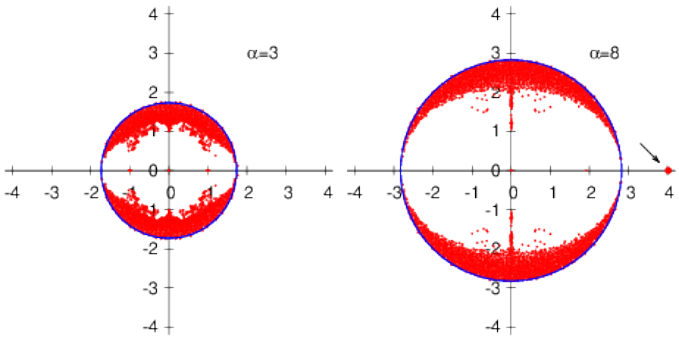

Fig. 2. Spectrum of the non-backtracking matrix in the complex plane for a problem generated with $\epsilon = 0.25$, $n = 2000$. We used $\alpha = 3$ (left side) and $\alpha = 8$ (right side), to be compared with $\alpha_{\mathrm{detect}} = 4$. Each point represents an eigenvalue. In both cases, the bulk of the spectrum is confined in a circle of radius $\sqrt{\alpha}$. However, when $\alpha > \alpha_{\mathrm{detect}}$, a single isolated eigenvalue appears out of the bulk at $(1-2\epsilon)\alpha$ (see the arrow on the right plot) and the corresponding eigenvector is correlated with the planted assignement.

III. From the non-backtracking operator to the Bethe Hessian

In this section, we relate the spectra of H, B and B′ by generalizing some properties discussed in [10], [11]. ($\lambda \neq \pm 1, v \in \mathbb{R}^{2m}$) being an eigenpair of B, we define

$$v_i = \sum_{j\in\partial i} v_{j\to i}, \qquad \forall 1 \leq i \leq n\,. \tag{15}$$

Since $\lambda v_{i\to j} = \sum_{k\in\partial i\setminus j} J_{ki} v_{k\to i}$ it follows that $\lambda v_{i\to j} = v_i - J_{ij}v_{j\to i}$. Closing the equation on the single site elements v_i thus leads to

$$v_i\left(1 + \sum_{k\in\partial i}\frac{J_{ij}^2}{\lambda - J_{ij}^2}\right) - \lambda\sum_{k\in\partial i}\frac{J_{ij}}{\lambda - J_{ij}^2}v_k = 0\,. \tag{16}$$

For convenience, we now define the matrix:

$$\mathrm{H(X)} = (X^2 - 1)\mathbb{1} - XJ + D \tag{17}$$

Note in particular that the Bethe Hessian reads $\mathrm{H} = \mathrm{H}(\sqrt{\alpha})$. Given that the values of J_{ij} are ± 1, all eigenvalues of B different from ± 1 thus must satisfies the following generalization of the Ihara-Bass formula [24] :

$$\det\left[(\lambda^2 - 1)\mathbb{1} - \lambda J + D\right] = \det \mathrm{H}(\lambda) = 0\,. \tag{18}$$

To solve (16) one needs to find an eigenvector $\mathbf{v}$ of $\mathrm{H}(\lambda)$ with a zero eigenvalue. This is a quadratic eigenproblem, which can be turned into a linear one by introducing the matrix B′ of Algo. 1. Indeed, if $\lambda \in \mathbb{R}$ is an eigenvalue of B′ with

eigenvector v', then it follows that $v := \{v'_i\}_{n+1 \leq i \leq 2n}$ is an eigenvector of $\mathrm{H}(\lambda)$ with eigenvalue 0, so that λ is an eigenvalue of B as well (at least if $\lambda \neq \pm 1$), justifying eq. (8,9). Note that since we are interested in values of $\lambda > 1$ (since $\lambda > \alpha$ and we need $\alpha > 1$ from (4)), the limitation of looking at $\lambda \neq \pm 1$ is irrelevant.

Finally, following [11], we can relate the spectra of B and H by the following argument. For X large enough, $\mathrm{H}(X)$ is positive definite. Then as X decreases, $\mathrm{H}(X)$ will gain a new negative eigenvalue whenever X becomes equal to an eigenvalue of B. This justifies the following corollary:

Corollary 2: if the conditions of Theorem 1 apply, then $\mathrm{H} = \mathrm{H}(\sqrt{\alpha})$ has a unique negative eigenvalue if $\alpha > \alpha_{\mathrm{detect}}$, and none otherwise.

Strictly speaking, if we denote by λ_1 the leading eigenvalue of B, we have only shown that the eigenvector with eigenvalue 0 of $\mathrm{H}(\lambda_1)$ is positively correlated with the planted variables if $\alpha > \alpha_{\mathrm{detect}}$. However, we observe numerically (see figure 1) that the eigenvector with negative eigenvalue of H is also positively correlated, and in fact gives a slightly better overlap. This point will have to be clarified in future work.

It is worth noting the Bethe Hessian is also related to the belief propagation algorithm. [25] showed that the fixed points of the BP recursion are stationary points of the so-called Bethe free energy. Direct optimization of the Bethe free energy has then been proposed as an alternative to BP [26]. In this context, [11] showed that the so-called paramagnetic fixed point (corresponding to an uninformative assignment) is a local minimum of the Bethe free energy if and only if H is positive definite. Algo. 2 can therefore be seen as a spectral relaxation of the direct optimization of the Bethe free energy. In the end, both approaches are indeed deeply related to BP.

IV. Conclusion

We have considered the problem of partially recovering binary variables from the observation of censored edge weights, and described two optimal spectral algorithms for this task that can provably perform partial recovery as soon as it is information theoretically possible to do so. Remarkably, these algorithms do not require the knowledge of the noise parameter ϵ and perform almost as well as belief propagation, which is expected (but not proved) to be Bayes optimal for this problem. This allows to close the gap from previous works, both algorithmically, by providing optimal spectral algorithms, and theoretically, by proving that the transition (4) is a necessary *and* sufficient condition for partial recovery.

Acknowledgment

This work has been supported by the ERC under the European Union's FP7 Grant Agreement 307087-SPARCS and by the French Agence Nationale de la Recherche under reference ANR-11-JS02-005-01 (GAP project).

References

[1] E. Abbe, A. S. Bandeira, A. Bracher, and A. Singer, "Decoding binary node labels from censored edge measurements: Phase transition and efficient recovery," *arXiv:1404.4749*, 2014.

[2] E. Abbe and A. Montanari, "Conditional random fields, planted constraint satisfaction and entropy concentration," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Springer, 2013, pp. 332–346.

[3] S. Heimlicher, M. Lelarge, and L. Massoulié, "Community detection in the labelled stochastic block model," 09 2012. [Online]. Available: http://arxiv.org/abs/1209.2910

[4] M. Lelarge, L. Massoulie, and J. Xu, "Reconstruction in the labeled stochastic block model," in *Information Theory Workshop (ITW), 2013 IEEE*, Sept 2013, pp. 1–5.

[5] N. Bansal, A. Blum, and S. Chawla, "Correlation clustering," *Machine Learning*, vol. 56, no. 1-3, pp. 89–113, 2004.

[6] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, "Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications," *Phys. Rev. E*, vol. 84, no. 6, p. 066106, 2011.

[7] E. Mossel, J. Neeman, and A. Sly, "Stochastic block models and reconstruction," *arXiv preprint arXiv:1202.1499*, 2012.

[8] L. Massoulie, "Community detection thresholds and the weak ramanujan property," *arXiv preprint arXiv:1311.3085*, 2013.

[9] E. Mossel, J. Neeman, and A. Sly, "A proof of the block model threshold conjecture," *arXiv preprint arXiv:1311.4115*, 2013.

[10] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang, "Spectral redemption in clustering sparse networks," *Proceedings of the National Academy of Sciences*, vol. 110, no. 52, pp. 20 935–20 940, 2013.

[11] A. Saade, F. Krzakala, and L. Zdeborová, "Spectral clustering of graphs with the bethe hessian," in *Advances in Neural Information Processing Systems*, 2014, pp. 406–414.

[12] M. Mézard, M. A. Virasoro, and G. Parisi, *Spin glass theory and beyond*. World scientific, 1987.

[13] F. Krzakala and L. Zdeborová, "Hiding quiet solutions in random constraint satisfaction problems," *Phys. Rev. Lett.*, vol. 102, p. 238701, 2009.

[14] F. Krzakala, M.-C. Angelini, and F. Caltagirone, "Statistical physics of inference problems," Lecture notes, http://ipht.cea.fr/Docspht/articles/t14/045/public/notes.pdf, 2014.

[15] H. Nishimori, "Internal energy, specific heat and correlation function of the bond-random ising model," *Progress of Theoretical Physics*, vol. 66, no. 4, pp. 1169–1181, 1981.

[16] L. Viana and A. J. Bray, "Phase diagrams for dilute spin glasses," *J. Phys. C: Solid State Physics*, vol. 18, no. 15, p. 3037, 1985.

[17] F. Guerra and F. L. Toninelli, "The high temperature region of the viana–bray diluted spin glass model," *Journal of statistical physics*, vol. 115, no. 1-2, pp. 531–555, 2004.

[18] P. Zhang, "Non-backtracking operator for ising model and its application in attractor neural networks," *arXiv:1409.3264*, 2014.

[19] J. Pearl, "Reverend bayes on inference engines: A distributed hierarchical approach," in *AAAI*, 1982, pp. 133–136.

[20] M. Mezard and A. Montanari, *Information, physics, and computation*. Oxford University Press, 2009.

[21] A. Saade, F. Krzakala, and L. Zdeborová, "Spectral density of the non-backtracking operator on random graphs," *EPL*, vol. 107, no. 5, p. 50005, 2014.

[22] P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," *Social Networks*, vol. 5, no. 2, p. 109, 1983.

[23] C. Bordenave, M. Lelarge, and L. Massoulié, "Non-backtracking spectrum of random graphs: community detection and non-regular ramanujan graphs," *arXiv*, 2015.

[24] H. Bass, "The ihara-selberg zeta function of a tree lattice," *International Journal of Mathematics*, vol. 3, no. 06, pp. 717–797, 1992.

[25] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Bethe free energy, kikuchi approximations, and belief propagation algorithms," *Advances in neural information processing systems*, vol. 13, 2001.

[26] M. Welling and Y. W. Teh, "Belief optimization for binary networks: A stable alternative to loopy belief propagation," in *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 2001, pp. 554–561.