Nonbacktracking operator for the Ising model and its applications in systems with multiple states

Pan Zhang

Phys. Rev. E **91**, 042120 — Published 17 April 2015

DOI: 10.1103/PhysRevE.91.042120

Non-backtracking operator for the Ising model and its applications in systems with multiple states

Pan Zhang[1]

[1]*Santa Fe Institute, Santa Fe, New Mexico 87501, USA**

The non-backtracking operator for a graph is the adjacency matrix defined on directed edges of the graph. The operator was recently shown to perform optimally in spectral clustering in sparse synthetic graphs, and have a deep connection to Belief Propagation (BP) algorithm. In this paper we consider non-backtracking operator for Ising model on a general graph with a general coupling distribution, and study the spectrum of this operator analytically. We show that spectral algorithms based on this operator is equivalent to Belief Propagation algorithm linearized at the paramagnetic fixed-point, and recovers replica-symmetry results on phase boundaries obtained by replica methods. This operator can be applied directly to systems with multiple states like Hopfield model. We show that spectrum of the operator can be used to determine number of patterns that stored successfully in the network, and the associated eigenvectors can be used to retrieve all the patterns simultaneously. We also give an example on how to control the Hopfield model, i.e. making network more sparse while keeping patterns stable, using the non-backtracking operator and matrix perturbation theory.

I. INTRODUCTION

Ising model, which consists of spins arranged in a graph, is a fundamental model in statistical physics. It also has direct applications in many fields of science including e.g. physics [1, 2], neuron science [3], computer science [4, 5] and social science [6]. With different distribution of couplings, Ising model can act as a model of different behaviors like ferromagnets, spin glasses and neural networks. It can be seen as a simple model for collaborative behaviors of iterating systems and gives a simple example of phase transitions, a phenomenon appearing in different fields of science.

Exact solutions of Ising model only exist on special cases of topologies, e.g. one dimensional and two dimensional lattices. On general graphs that many real-world applications meet, one usually needs approximations, especially with a general distribution of couplings. Commonly used approximations include naïve mean-field approximation, TAP approximation and Kikuchi expansions. Among those approximations, one of the most popular approximation is Bethe approximation which assumes independence of conditional probabilities. Bethe approximation is exact on trees, and is found to be a good approximation on (random) sparse systems. Message passing algorithms based on Bethe approximation, so called Belief Propagation (BP) algorithms and its variations, have been used widely in many fields [7]. BP is also named replica-symmetric cavity method in statistical physics [11], it gives good estimate of free energy and marginals of the system, the stabilities of BP fixed-points have been used widely to estimate the phase transitions of the system. In a recent paper, it was shown that in the graph clustering problems, there is a connection between Belief Propagation and so-called "non-backtracking" operator [9], that belief propagation linearized at a symmetric fixed-point corresponds to spectral clustering algorithms using non-backtracking operator B, with

$$B_{i\to j,k\to l} = \delta_{l,i}(1-\delta_{k,j}) \tag{1}$$

defined on directed edges of the graph. In last equation $i \to j$ denotes a directed edge from node i to node j. It has been shown that spectral algorithms using B works all the way down to the detectability transitions in the stochastic block model, and improves significantly the performance of clustering over other matrices like random walk matrix and Laplacians. The reason for the improvement is that non-backtracking matrix of a graph as a good property that almost all its eigenvalues are confined by a disk in the complex plane. So the eigenvalues corresponding to community structure lay outside the disk and are easily distinguishable.

Note that Ising model with a given topology, e.g. on a two-dimensional lattice or on a random network, can be seen as a graphical model defined on a graph weighted by couplings J_{ij}. And we know that without external fields, Ising model has a symmetric solution, so-called "paramagnetic solution", which is analogous to the symmetric fixed point in graph clustering problem. Actually this analogy motivated this work in the first place, and the motivations are based on following questions:

*Electronic address: pan@santafe.edu

- Given the analogy between Ising model and network clustering problem on the paramagnetic solution, can we generalize the non-backtracking operator from a graph to Ising model? In another words, how to define the non-backtracking operator from a un-weighted graph to a graph weighted by couplings?
- Does non-backtracking operator in Ising model still has good properties like sharp edge of spectral band?
- How do spectral algorithms based on non-backtracking operator work in sparse Ising models? Do they work as well as BP or even better?

In the following text we will first derive the non-backtracking operator of Ising model (we will call it C in this paper) on a graph with a given distribution of couplings at a certain temperature, by linearizing belief propagation algorithm at paramagnetic fixed-point. Then we compute some properties on its spectrum, and show that using spectrum of C we can recover replica symmetric results on phase transitions of ferromagnets and spin glasses.

The second part of this paper is focused on the applications of this operator when Ising model has multiple stable states, and acts as an associative memory like Hopfield model [4]. The Hopfield network is an Ising model with couplings generated using Hebb's rule. In the network many binary patterns are remembered as memory, and they can be retrieved using Markov Chain Monte-Carlo simulation (MCMC) or using belief propagation. However memory retrieval using MCMC is slow, and it is hard for both MCMC and BP to estimate how many patterns are remembered in the network. In this paper we show how to use spectrum of C to detect phases and phase transitions of Hopfield model in the thermodynamic limit. In single instances of Hopfield network, we show that number of real-eigenvalues outside the bulk in spectrum can be used to detect number of patters remembered in network, and the associated eigenvectors are correlated with those remembered patterns. Thus we can retrieval all the patterns simultaneously using sign of the eigenvectors, as opposed to MCMC and BP where one pattern is retrieved in each run. We also propose a spectral method based on C that removes edges of the network that influence least the real eigenvalues of C. By doing this, we can make the network more sparse while keep patterns remembered in the network stable.

The main contributions of this paper are:

- We show that spectrum of non-backtracking operator for the Ising model is analogous to spectrum of non-backtracking operator of graphs.
- We apply the non-backtracking operator to Hopfield model and propose spectral algorithms based on this operator for retrieving patterns and for controlling the networks.

The paper is organized as follows. Section II includes definitions of the non-backtracking matrix for Ising model, some properties of its spectrum are also computed. In Section III we give several examples of Ising model with specific coupling distributions, including ferromagnet, Sherrington-Kirkpatrick model and Viana-Bray model. In Section IV we consider the operator for Hopfield model and controling neural networks using eigenvectors of the operator. We conclude in Section V.

II. NON-BACKTRACKING OPERATOR FOR ISING MODEL

We consider Ising model on a graph with n spins, $\{\sigma\}$ is used to denote a configuration, with $\sigma_i \in \{-1, +1\}$ and $i \in \{1, ..., n\}$. Energy (Hamiltonian) of system with zero external field is pairwise:

$$E\left(\{\sigma\}\right) = -\sum_{\langle ij\rangle} J_{ij}\sigma_i\sigma_j. \tag{2}$$

For each configuration $\{\sigma\}$, at inverse temperature β, there is a Gibbs measure $P\{\sigma\} = \frac{1}{Z}e^{-\beta E(\{\sigma\})}$. On a single instance of the graph, one can use belief propagation algorithm to study the marginals of the Boltzmann distribution. BP equation reads [7]

$$\psi_{i\to j} = \frac{1}{Z_{i\to j}} \prod_{k\in\partial i\backslash j} \left(2\sinh(\beta J_{ij})\psi_{j\to i} + e^{-\beta J_{ij}}\right), \tag{3}$$

where $\psi_{i\to j}$ denotes "cavity" probability of σ_i taking value $+1$ with spin j removed from the graph, ∂i denotes neighbors of i and

$$Z_{i\to j} = \prod_{k\in\partial i\backslash j} \left(2\sinh(\beta J_{ij})\psi_{j\to i} + e^{-\beta J_{ij}}\right) + \prod_{k\in\partial i\backslash j} \left(2\sinh(-\beta J_{ij})\psi_{j\to i} + e^{\beta J_{ij}}\right), \tag{4}$$

is the normalization. When BP equation converges on a graph, marginals of a spin taking value +1 can be computed using

$$\psi_i = \frac{1}{Z_i} \prod_{k \in \partial i} \left(2 \sinh(\beta J_{ij}) \psi_{j \to i} + e^{-\beta J_{ij}} \right), \tag{5}$$

where again Z_i is the normalization. It is easy to check that $\psi_{i \to j} = \psi_i = \frac{1}{2}$ is always a solution to BP equation (3) (though it could be not stable). We call this fixed point "paramagnetic" fixed-point.

As we introduced in the introduction, non-backtracking operator appears naturally in linearizing BP of inference of stochastic block model [9], also in linearizing BP for modularity [12], because in both cases BP has a factorized fixed point: $\psi_{i \to j} = \frac{1}{q}$ for every directed edge $i \to j$, where q denotes the number of groups one variable could take. So we can do similar things as we did for clustering problem before: deriving non-backtracking operator by linearizing BP equation at paramagnetic fixed point.

First we write BP messages as deviations from the paramagnetic fixed point: $\psi_{i \to j} = \frac{1}{2} + \Delta_{i \to j}$. At the first order we have

$$\Delta_{i \to j} = \sum_{k \in \partial i \backslash j} \left. \frac{\partial \psi_{i \to j}}{\partial \psi_{k \to i}} \right|_{\psi = \frac{1}{2}} \Delta_{k \to i}, \tag{6}$$

With derivatives of BP messages evaluated at paramagnetic fixed-point read

$$\left. \frac{\partial \psi_{i \to j}}{\partial \psi_{k \to i}} \right|_{\psi = \frac{1}{2}} = \tanh(\beta J_{ik}). \tag{7}$$

Equation (6) is a linearized version of BP equation (3), and it is equivalent to eigenvector problem of an operator that we call *non-backtracking operator for Ising model* C, with:

$$C_{i \to j, k \to l} = \delta_{l,i} (1 - \delta_{k,j}) \tanh(\beta J_{lk}). \tag{8}$$

If we use m to denote number of edges in the graph, then the size of C is $2m \times 2m$. Note that in right hand side of last equation, $\delta_{i,l}(1 - \delta_{k,j})$ is the non-backtracking operator for a graph B (Eq. (1)), and has been used in [9] for graph clustering problem. The same as B, matrix C is also defined on the graph, however now on each edge there is a weight $\tanh(\beta J_{ij})$. Note that in discussion of [9] authors suggested a weighted non-backtracking matrix, and our matrix C can be seen as such weighted non-backtracking matrix, with weight $\tanh(\beta J_{ij})$ on each edge.

Using some techniques from tree-reconstruction theory [9], we can compute some properties of C's spectrum. Since the matrix is not symmetric $C_{i \to j, k \to l} \neq C_{k \to l, i \to j}$, the eigenvalues of C are complex. We can show (see Appendix A for details of computations) that in the complex plan, almost all C's eigenvalues are confined by a disk of radius R, with

$$R = \sqrt{\hat{c} \left\langle \tanh^2(\beta J_{ij}) \right\rangle}. \tag{9}$$

In another words, it means the density of eigenvalues, $\rho(\lambda) = \frac{1}{2m} \sum_{i \to j} \delta(\lambda - \lambda_{i \to j})$, is non-zero inside the disk, and is zero outside the bulk. However keep in mind that there could be several eigenvalues outside the disk. There is another way to understand this edge of bulk, in the sense of noise propagation on trees: consider some random noise with mean 0 and unit variance are put on the leaves of a tree; by iterating C, the noise may be propagated to the root of the tree with non-vanishing variance, if $\hat{c} \left\langle \tanh^2(\beta J_{i \to j}) \right\rangle >= 1$. Thus radius of disk $R > 1$ tells us that random noise could make Edwards-Anderson order parameter [13] of the system finite, which is a sign showing that the system is in the "spin glass" phase, and replica symmetry is breaking. So R tell us where is the spin glass phase and its relative strength to other possible phases. $R = 1$ point is also known as the de Almeida-Thouless local stability condition [14], the Kesten-Stigum bound [15], or the threshold for census or robust reconstruction [7, 16].

If there are real-eigenvalues outside the bulk, they may tell us other possible phases of the system, which correspond to BP solutions other than spin glass phase. For example, if all the couplings are positive, it is easy to see that there is a positive eigenvalue associated with an eigenvector with most of its entries positive that we call *ferromagnetic eigenvector*. Actually if not all edges in the system are positive, the eigenvector with most entries positive could be still there with a positive real eigenvalue. Assume the graph is a tree with edge weighted by $\tanh(\beta J_{ij})$, and all its nodes associated with a spin with value 1. If we iterate the non-backtracking operator for many times, we can approximately recover this eigenvector and compute the eigenvalue (see Appendix B for details of computations) as

$$\mu = \hat{c} \left\langle \tanh(\beta J_{ij}) \right\rangle. \tag{10}$$

Besides the ferromagnetic eigenvalue, there could be other real-eigenvalues representing other phases of the system. So the competition of real-eigenvalues and R gives us the information of phase transitions of the system.

Note that those eigenvalues of C correspond to parallel iterating of the matrix, not to the sequential iterating, as people usually do in converging BP. However we can link the spectrum to the stability of sequentially-updated BP: each iterating in sequential updating corresponds to additionally adding a constant to diagonal elements of C. So sequential updating effectively increases all the eigenvalues simultaneously. This effect actually enhances the positive eigenvalues and suppresses the negative eigenvalues. Note that damping gives similar effects. Thus we can summarize the relation between spectral of C to the stability of sequentially-updated BP:

- If the eigenvalue outside the bulk is positive, it represents a ferromagnetic (or retrieval) solution to both sequentially-updated BP and parallel-updated BP.
- If the eigenvalue outside the bulk is negative, then it represents a anti-ferromagnetic solution in parallel-updated BP, but does **not** represent a fixed-point in sequentially-updated BP, because negative eigenvalues will be suppressed down by effect of sequential updating. Actually this is the reason why in for example (planted) graph coloring problem, sequential updating (or damping) in BP can help avoid the ferromagnetic solution.

We note that R and μ of C for ferromagnetic, anti-ferromagnetic and spin glasses have been already shown in [37] (authors called C stability matrix of BP) , using a different technique. However in [37], authors focused on the stability of Belief Propagation, not the non-backtracking matrix itself. And in this paper, using techniques from tree-reconstruction theory, we displayed detailed picture of C's spectrum: we claim that spectrum of C is essentially analogous to that of B, the non-backtracking matrix of a graph, with a sharp edge of band. And we also made it clear the relation between spectrum of C and stability of paramagnetic solutions in both parallel and sequential BP. Another reason that we keep the detailed computation of μ and R in this paper is that we need to generalize this computation to compute the spectrum of non-backtracking matrix of Hopfield neural networks in Sec. IV.

Also note that the system studied in this paper corresponds to sequential dynamics. If parallel dynamics of the system is considered, the energy for one configuration would be

$$\tilde{H}(\{\sigma\}) = -\frac{1}{\beta}\sum_i \log 2\cosh(\beta \sum_j J_{ij}\sigma_j),$$

which is different from that of sequential dynamics Eq. (2). So the BP equations for parallel dynamics are different from sequential dynamics. As a consequence, the non-backtracking operators, which can be obtained by linearizing BP equations at the paramagnetic fixed-point, are different.

III. NON-BACKTRACKING OPERATOR FOR FERROMAGNETS AND SPIN GLASSES

Now we take several examples of Ising model with different couplings distributions. For the first example we consider ferromagnetic Ising model, which has $J_{ij} = 1$ for every edge $\langle ij\rangle$, on Erdős-Rényi (ER) random graphs. ER graphs is sparse with $\hat{c} = c$, so results given by non-backtracking matrix would be quite accurate, due to the locally tree-like structure of the topology. By setting $J_{ij} = 1$ in Eq. (9) and (10) we simply have $R_F = \sqrt{c}\tanh\beta$ and $\mu_F = c\tanh\beta$. Thus μ_F is always larger than R_F with finite β and $\hat{c} > 1$. It means this system can never be in spin glass phase: it has to be in either ferromagnetic phase when $\mu_F > 1$ or paramagnetic when $\mu_F < 1$. And at $c\tanh\beta = 1$, i.e. $\beta = \tanh^{-1}\frac{1}{c}$, system undergoes a transition from paramagnetic to ferromagnetic phase. Note that in ferromagnetic, C can be expressed as

$$C = B\tanh\beta, \tag{11}$$

where B is the non-backtracking matrix on the graph. From [9] we know that B has the largest eigenvalue $\lambda_1^B = \hat{c}$, so as a sanity check, one can verify that $\mu = \tanh(\beta) * \lambda_1^B$. It has been shown in [9] that non-trivial part of B's spectrum can be obtained from a $2n \times 2n$ matrix

$$B' = \begin{pmatrix} 0 & D-I \\ -I & A \end{pmatrix},$$

where D is diagonal matrix with degree of nodes on the diagonal entries, I is the identity matrix and A is the adjacency matrix. Thus non-trivial part of C's spectrum for ferromagnet can actually be computed with shorter computational time.

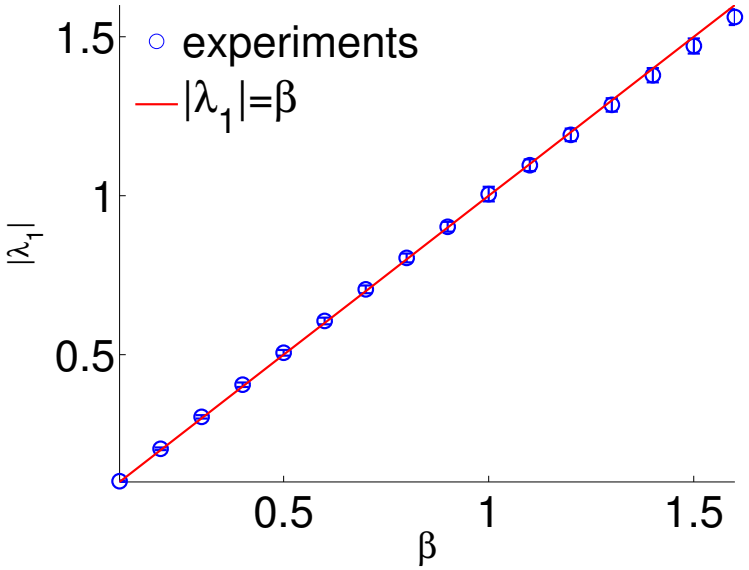


FIG. 1: (Color online) Absolute value of the largest eigenvalue of non-backtracking operator, $|\lambda_1|$, as a function of β for SK model with $n = 60$. Each point is averaged over 10 instances.

If the graph of the model is not a purely random graph, but has a community structure, spectral methods based on non-backtracking matrix of Ising ferromagnets can be used to detect communities in the graph. We refer readers that interested in this part to Appendix C for more details.

If the couplings of Ising model are not identical, but drawn randomly from a distribution, then the model behaves as spin glass models, which were proposed to explain magnetic systems with frozen structural disorder. Theory of spin glasses e.g. replica symmetry breaking theory and cavity method find many application is many fields. As an example we consider non-backtracking matrix of two spin glass models, namely Sherrington-Kirkpatrick model [20] and Viana-Bray model [21]. It is well known that those spin glass models have paramagnetic phase at high temperature and spin-glass phase at low-temperature where both magnetization and Edwards-Anderson parameter are non-zero, and replica symmetry is broken [2]. First we consider Sherrington Kirkpatrick model which is defined on fully connected graph with couplings following Gaussian distribution:

$$P(J_{ij}) = \frac{1}{\sqrt{n}}\mathcal{N}(0,1).$$

From Eq. (9) and Eq. (10), we have $\mu_{\rm SK} = 0$ and

$$R_{\rm SK} = \sqrt{n\left\langle \tanh^2(\beta J_{ij})\right\rangle} \approx \sqrt{n\left\langle \beta^2 J_{ij}^2\right\rangle} = \beta. \tag{12}$$

So the paramagnetic solution is stable as long as $\beta \leq 1$, and system enters spin glass phase with $\beta > 1$. Actually these are the well known replica symmetry results.

To test theory (Eq. (12)), in Fig. 1 we plot the edge of bulk (absolute value of the largest eigenvalues) as a function of β for networks with $n = 60$. We can see that it consistent well with $|\lambda_1| = \beta$.

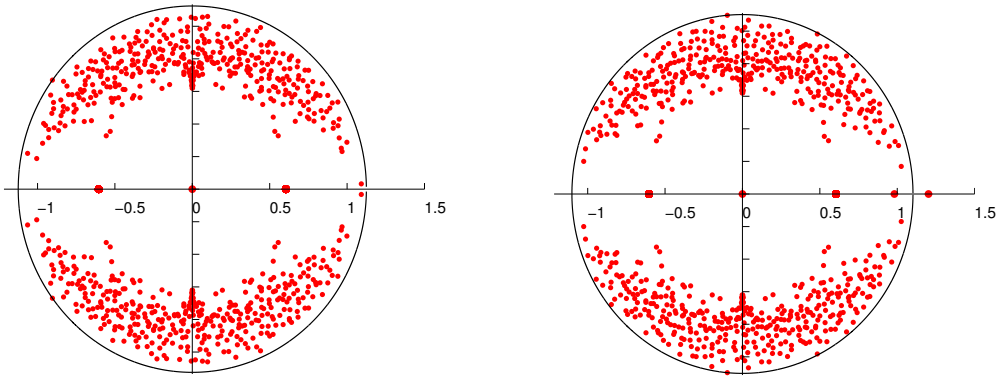
Note that though (see appendix) Hessian matrix of Bethe free energy works as well as non-backtracking matrix in detecting paramagnetic to ferromagnetic transition in ferromagnets, it does not work well in SK model. Our numerical result (as well as experimental results in [37]) show that Hessian matrix for SK model is positive definite, hence it does not tell us where system encounters the spin glass transition.

Then we consider Viana-Bray model [21], also called $\pm J$ model, which is a spin glass model defined on random graphs. It is similar to the ferromagnets on random graphs but with couplings taking $+1$ and -1 randomly which give frustration to the system.

$$p(J_{ij}) = p^+\delta_{J_{ij}} + (1-p^+)(1-\delta_{j_{ij}})$$

By varying p^+ and β, the system can be in paramagnetic phase, spin glass phase or ferromagnetic phase. From Eq.(9)(10) we have

$$\mu = 2p^+\tanh\beta - 1, \qquad R = \sqrt{c}\tanh\beta.$$



(a) (b)

FIG. 2: (Color online) Spectrum (in the complex plane) of non-backtracking matrix of Viana-Bray model on a graph with $n = 500$, $c = 3$. In (a) $\beta = 0.7 > \tanh^{-1}\frac{1}{\sqrt{c}}$, $p^+ = 0.5$, system is in spin glass phase and in (b) $p^+ = 0.7$, system is in ferromagnetic phase

Obviously system undergoes paramagnetic to spin glass transition at $\beta = \tanh^{-1}\frac{1}{\sqrt{c}}$ independent of p^+. And may undergo paramagnetic to ferromagnetic transition if $\mu > R$. So our result recovers the phase boundaries derived using replica methods in [21]. In Fig. 2 we plot the spectrum of non-backtracking matrix for Viana-Bray model in the complex plane for system in two phases, one in spin glass phase where there is no real eigenvalues outside the bulk, and the other one in ferromagnetic phase where there is one real eigenvalue outside the bulk and is greater than 1.

IV. NON-BACKTRACKING OPERATOR FOR HOPFIELD MODEL AND CONTROLLING OF THE NETWORKS

A. Spectrum and pattern retrieving

Hopfield model is a classic model for associative memory, it stores patterns by couplings of Ising model. We assume P random binary patterns $\xi_i^\mu = [-1, +1]$ with $i = [1, ..., n]$, $\mu = [1, ..., P]$ stored in network by Hebb's rule [22]:

$$J_{ij} = \frac{1}{P}\sum_{\mu=1}^{P}\xi_i^\mu\xi_j^\mu.$$

If P patterns are memorized successfully, they can be retrieved by Glauber dynamics [23]: at time t one neuron i is randomly selected and its value σ_i^t is updated according to the probabilistic rule

$$p(\sigma_i^t = \sigma) = \frac{e^{\beta\sum_{j\in\partial i}J_{ij}\sigma\sigma_j^{t-1}}}{2\cosh\beta\sum_{j\in\partial i}J_{ij}\sigma_j^{t-1}}.$$

From a random configuration, Glauber dynamics could converge to set of configurations that correlated with one pattern. If all patterns are memorized successfully in the network, each pattern will behave like an attractor to the Glauber dynamics, with different basin of attractions. That is why Hopfield model is categorized into attractor neural networks [24]. The reason that patterns behave as attractors to the dynamics is that using Hebb's rule, each pattern corresponds to one pure state (a minima of free energy), and Glauber dynamics, which a spin-flip algorithm that converges to equilibrium, will trap into one of those local minimums of free energy.

The original Hopfield network was defined on fully-connected network, and its statistical mechanics has been studied [25] using replica method. However fully-connected model is not biologically realistic. In this paper we will focus on its more biologically-realistic version, which is Hopfield model on a random graph. The model is also called finite-connectivity Hopfield model and its statistical mechanics has also studied using replicas [26], and phase boundaries were analyzed. The phase diagram of Hopfield model (both fully-connected and on a random graph) is composed of

three parts: at high temperature every neuron behaves the same, magnetization is zero, system is in the paramagnetic phase; at a low temperature and with a low number of patterns, patterns are attractive to Glauber dynamics and system is in "retrieval" or "memory" phase; at a low temperature and a high number of patterns, network is confused by so many patterns hence none of the pattern is memorized successfully, Glauber dynamics will go away even started from one of the patterns, system is in spin glass phase.

Note that replica results were derived for the model in the thermodynamic limit averaged over disorder (realizations of patterns), hence it is hard to apply the replica result to a single instance of network, to do tasks such as retrieving patterns and determining number of patterns. The classic way to do such tasks is Glauber dynamics. However on large networks it is time consuming, and the result depends on the initial configuration of the dynamics. And it is hard for Glauber dynamics to determine how many patterns stored in the network, because one can not explore all the initial configurations.

A faster way to retrieval patterns is running BP on the instance. But there is still a drawback that it is difficult to answer how many patterns there are in the network, since exploring all possible initial messages for BP is hard.

In the rest part of this section we will study how to do the pattern retrieval task using non-backtracking matrix. It recovers replica result at thermodynamics limit easily, and on single instances it has advantages over Glauber dynamics and BP for speed as well as for determining number of patterns.

First we study the spectrum of non-backtracking operator in the thermodynamic limit. From Eq. (9), the edge of bulk of Hopfield model, which tells us where is the spin glass phase, is expressed as

$$R_{\mathrm{H}} = \sqrt{\hat{c}\left\langle \tanh^2\left(\frac{\beta}{P}\sum_{\mu=1}^{P}\xi^{\mu}\right)\right\rangle}, \tag{13}$$

the averaging is taken over random patterns.

Obviously there is no stable "ferromagnetic" phase since if we iterate C with all-one configuration $\{1, 1, ..., 1\}$ on leaves of the tree, we get

$$\hat{c}\left\langle \tanh\left(\frac{\beta}{P}\sum_{\mu=1}^{P}\xi^{\mu}\right)\right\rangle = 0.$$

However, note that if instead of all-one configuration, we put a pattern on the leaves of the tree and do iterating, the information of the pattern could be preserved during iterating. It is very similar to initialize belief propagation by messages correlated with one pattern, or running Glauber dynamics from a pattern. Analysing iterating with a random pattern on leaves is difficult, but we can show it in an equivalent but easier way, by focusing on one of the stored patterns, let us say ξ^1 without loss of generality. Since patterns are chosen randomly and independently, we can do a gauge transformation to all the patterns which transforms ξ^1 to all-one configuration, and transforms other patterns to $\hat{\xi}_i^{\mu} = \xi_i^1\xi_i^{\mu}$. It is easy to see that this transformation does not change distance between patterns, hence does not change the property of the system. Original patterns can be recovered easily by product ξ_i^1 again: $\xi_i^{\mu} = \xi_i^1\hat{\xi}_i^{\mu}$. Under this transformation, couplings become

$$\hat{J}_{ij} = \frac{1}{P}\left[1 + \sum_{\mu=2}^{P}\hat{\xi}_i^{\mu}\hat{\xi}_j^{\mu}\right].$$

Obviously now couplings are biased to positive, and there could be a ferromagnetic phase in the transformed system. Then we can treat $\frac{1}{p}$ term as a "signal" term that tells us information about all-one configuration, and term $\sum_{\mu=2}^{P}\hat{\xi}_i^{\mu}\hat{\xi}_j^{\mu}$ as a "noise" term that gives some cross-talk noise to signal term. If there is only one pattern, cross-talk noise is 0, the problem essentially goes back to the ferromagnets. When number of pattern is small, noisy term has small fluctuation thus signal remains clear during iterating.

It is easy to see that the eigenvector correlated to the all-one vector in the transformed system is related to the eigenvector correlated with the first pattern in the original system. And the eigenvalue associated with the eigenvector correlated with all-one vector in the transformed system is the same as the eigenvalue associated with eigenvector correlated with first pattern in original system. With $n \to \infty$, this eigenvalue can be written as

$$\mu_{\mathrm{H}} = \hat{c}\left\langle \tanh\frac{1}{P}\left(1 + \sum_{\mu=2}^{P}\hat{\xi}^{\mu}\right)\right\rangle. \tag{14}$$

So setting $\mu_{\mathrm{H}} = 1$ gives paramagnetic to retrieval transition and $R_{\mathrm{H}} = 1$ gives paramagnetic to spin glass transitions.

If we write last equation in the form of

$$\mu_{\rm H} \;=\; \hat{c}\left\langle \tanh\frac{1}{P}\left(\xi^1\xi^1+\sum_{\mu=2}^{P}\hat{\xi}^\mu\right)\right\rangle = \hat{c}\left\langle \xi^1\tanh\frac{1}{P}\left(\xi^1+\sum_{\mu=2}^{P}\hat{\xi}^\mu\right)\right\rangle, \tag{15}$$

we can see that the expression for phase boundaries are in agreement with result obtained by replica method [26], by considering a little difference in definition of the models. Moreover, with patterns chosen randomly in the thermodynamic limit, we have

$$\begin{aligned}\mu_{\rm H} &= \frac{\hat{c}}{2^P}\sum_{s=0}^{P-1}\binom{P-1}{s}\tanh\left(\beta\frac{P-2s}{P-1}\right)\\ R_{\rm H} &= \sqrt{\frac{\hat{c}}{2^P}\sum_{s=0}^{P}\binom{P}{s}\tanh^2\left(\beta\frac{P-2s}{P}\right)}.\end{aligned} \tag{16}$$

If we do the gauge transform on every pattern, we would have P such eigenvalues. But note that though gauge transform was used in computing the $\mu_{\rm H}$ and $R_{\rm H}$, it is only a trick to study analytically in an easier way the eigenvalues. In a single instance, we do not need to do such transform, there should be P eigenvalues corresponding to P patterns. And on the given network, once we find real eigenvalues outside the bulk, we can use eigenvectors associated with those eigenvalues to retrieve patterns.

For more detail, when we have an eigenvector $v_{j\to i}$, first we obtain the vector S containing summation of incoming values from edges for nodes, from an eigenvector v,

$$S_i = \sum_{j\in\partial i}\tanh(\beta J_{ij})v_{j\to i}.$$

Then we set positive entries of S to 1 and negative entries to -1, and zero entries to 1 or -1 randomly. The obtained configuration should be correlated with the pattern, i.e. the overlap parameter which is inner product of obtained configuration and a pattern

$$O = \frac{1}{n}\sum_{i=1}^{n}S_i\xi_i^\mu, \tag{17}$$

is positive.

In Fig. 3 we plot one memorized binary pattern compared with one eigenvector of non-backtracking matrix, we can see that the sign of components of eigenvector are correlated with sign of the binary pattern.

In Fig. 4 we plot the spectrum of a Hopfield network with different parameters in the complex plane. In left panel, $P = 3$, three patterns are memorized, system is in retrieval phase. There are three eigenvectors outside the edge of bulk, and the associated eigenvectors are correlated with three stored patterns, with the overlap (Eq.(17)) for three patterns much larger than 0. In right panel, $P = 12$, none of the patterns is memorized successfully. In spectrum, the edge of bulk is larger than 1, there is no real eigenvalues outside the bulk, so system is in spin glass phase.

In the thermodynamic limit, self-averaging property will make those P eigenvalues identical. But in the finite-size networks, we could obtain different eigenvalues, telling us that P patterns are not memorized equally strong. One example is shown in left panel of Fig. 4 where three real-eigenvalues outside the bulk are not identical. When P real-eigenvalues are all out of bulk but not equal, though from random initial vectors, iterating C will always leads to the eigenvector associated with the largest eigenvalue, other patterns could be retrievable by other methods, E.g. by running Glauber dynamics starting from a configuration or running BP from a random messages. We have tested that by running Glauber dynamics we can reach other patterns corresponding to the eigenvalues outside the bulk from random initial configurations. We think the pattern with the largest eigenvalue are indeed stronger than other two patterns, and has larger basin of attraction than other two patterns; however its basin attraction does not dominate the entire configuration space, so other two patterns are still able to attract Glauber dynamics.

Another possible situation is that not all P eigenvalues are outside the bulk and larger than 1. This means not all the patterns are memorized stably. We have tested that in this case only patterns corresponding to eigenvalues outside the bulk are attractive to Glauber dynamics. For those patterns corresponding to eigenvalues inside bulk or smaller than 1, Glauber dynamics starting from exactly the pattern will still run away and converge to configurations around other patterns. So those patterns are indeed not retrievable.

So as described above, we can find all real-eigenvalues and use all real-eigenvectors associated with those eigenvalues to retrieve pattern simultaneously, as opposed to BP or Glauber dynamics, where only one pattern is retrieved at one time, without knowing neither total number of patterns, nor which patterns are memorized stronger than the others.

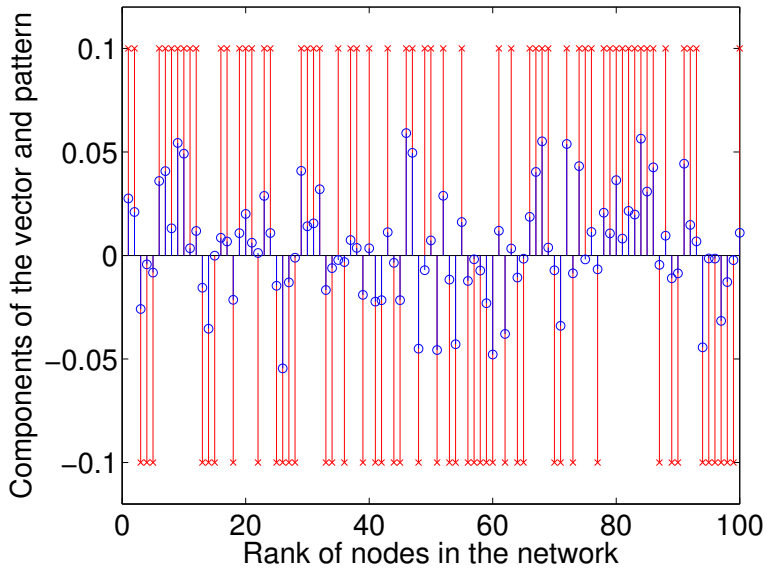


FIG. 3: (Color online) Comparison of a binary pattern and one eigenvector of non-backtracking matrix. Red stars are components of the binary pattern and blue circles denote components of the eigenvector. Parameters are $n = 100, \beta = 0.02, P = 3, c = 28$.

B. Controlling the Hopfield networks using the non-backtracking operator

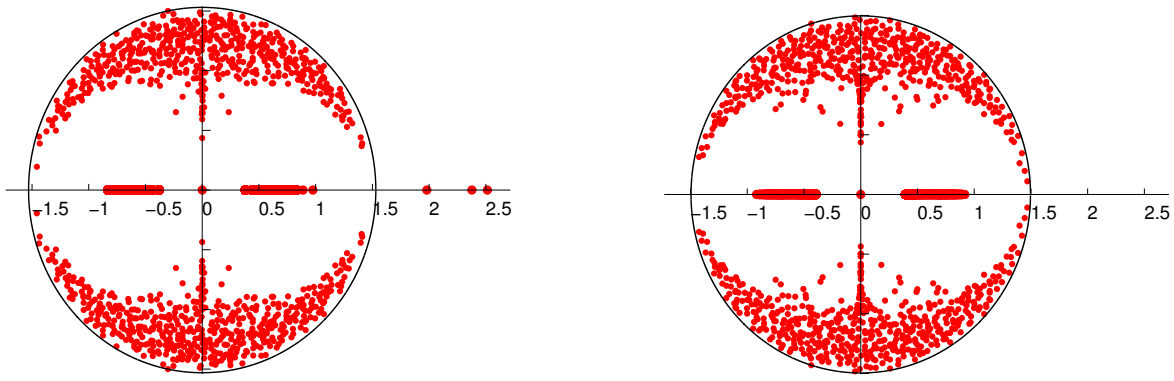
Besides the phase diagram, another property that has been studied extensively in neural networks is the capacity of the network that we denote by $\mathcal{C}$. Capacity is maximum number of patterns (divided by number of neurons) that can be stored successfully. This property is closely related to the strength of the memorized patterns. In our picture of spectrum of non-backtracking operator, by increasing number of patterns, number of real eigenvalues outside of bulk increases, but each of eigenvalue becomes closer to the bulk. When P exceeds $\mathcal{C}$, real-eigenvalues disappear and become complex eigenvalues, then system enters spin glass phase and losses all the memory.

A large amount of studies have been devoted to increase capacity and retrieval abilities of neural networks by optimizing the learning rule or by optimizing the topology of the network [27–33]. In the following text of the paper we are going to discuss along the line of optimizing the topology of the network. We know that fully-connected Hopfield network can store number of patterns proportional to number of neurons, but the fully-connected topology is not biological realistic. On the other hand, when we dilute the network, i.e. by making the network more and more sparse by removing edges, number of patterns stored is decreasing but the number of patterns per edge is increasing [26]. Thus one interesting question would be, given a network, how to make the network more sparse while keeping information of pattern stored by the network unchanged. This problem could be studied easily by non-backtracking matrix, since real eigenvalues outside the bulk tell us the stability of patterns, if we can keep those eigenvalues unchanged in manipulating topology of the network, we can keep those patterns stable in the network. A nature idea to keep eigenvalues could be by converting the problem to an optimization problem that at each time we remove the edge that decrease least the gap of eigenvalues,

$$\Delta = \sum_i (\lambda_i - |\lambda_e|),$$

where λ_i is the eigenvalue corresponding to the pattern we want to keep, and λ_e is the edge of bulk of non-backtracking matrix. This process is easy to implement but the computation is time consuming because selecting the edge that decreases least the spectral gap we have to remove every edge from the graph one by one, compute eigenvalues, then add the edge back. One way to make the process faster is to remove the edge that changes least the spectral gap among small randomly selected subset of edges. Actually if we are interested in keeping one eigenvalue λ instead of keeping the gap, we can do things much easier.

Assume after one edge removed, matrix C becomes $C_{\text{new}} = C + \Delta C$, If we look ΔC as a perturbation to the matrix C and assume that after the perturbation, an eigenvalue changes from λ to $\Delta\lambda$. Then we can show that change of eigenvalues can be expressed a function of left eigenvectors $\{u_{i\to j}\}$ and right eigenvectors $\{v_{j\to i}\}$ (see Appendix. D



(a) (b)

FIG. 4: (Color online) Spectrum (in the complex plane) of non-backtracking operator for a Hopfield network with $n = 500, c = 8$ and 3 patterns with $\beta = 1.2$ (a) and 12 patterns with $\beta = 2.5$ (b). In (a), three patterns are all memorized, overlap (Eq.(17)) for three patterns obtained using first three eigenvectors are 0.628, 0.620 and 0.448. In (b), none of the patterns is memorized, system is in the spin glass phase.

for the derivations):

$$\Delta\lambda \;=\; \frac{2\lambda}{uv}(u_{i\to j}v_{i\to j} + u_{j\to i}v_{j\to i}). \tag{18}$$

By computing once the eigenvectors, we know exactly by deleting which edge the eigenvalue is changed the least. So we can make network more sparse while keeping information of a pattern stable, by removing edges iteratively that give least $u_{i\to j}v_{i\to j} + u_{j\to i}v_{j\to i}$. For more detail, at each step, left and right eigenvectors associated with a real eigenvalue are computed, one (or several edges) that minimizes $u_{i\to j}v_{i\to j} + u_{j\to i}v_{j\to i}$ are removed, then the above process are repeated until network is sparse enough. Note this procedure is similar to the decimation algorithm using marginals of a message passing algorithm in solving constraint satisfaction problems [34], where nodes having most biased marginals are removed (fixed) at each iterating and decimation method in inverse Ising problem [35].

We did some numerical experiments on our edge-removing approach and compared its performance with random removing in Fig 5, where the real eigenvalue and edge of bulk are plotted as a function of fraction of edges removed. We can see from the figure that random removing makes the real eigenvalue decreasing very fast while our process keeps the eigenvalue unchanged up to more than half edges removed. Though edge of bulk in our process decreases also slower, the real eigenvalue joins bulk later than that in random removing.

Note that the problem of removing edges is just a simple example of optimizing topology of a neural network using the non-backtracking operator. Since information of stability of patterns can be obtained at same time by non-backtracking matrix we believe other kinds of topology optimization on neural networks can also be done easily using non-backtracking matrix.

V. CONCLUSION AND DISCUSSIONS

As a conclusion, we have studied non-backtracking operator for Ising model with general coupling distributions. It can be seen as a generalization of non-backtracking operator from a unweighted graph to weighted graph with edges weighted by $\tanh(\beta J_{ij})$. Using tree-reconstruction theory, we computed edge of bulk and possible ferromagnetic eigenvalue in spectrum of the operator, and we showed that in the thermodynamic limit the spectrum tells us phase transitions of the model, and the results recover replica symmetry results given by replica method in various of couplings distributions. In the Hopfield model, we showed that the number of real-eigenvalues outside the bulk determines number of patterns that have been remembered in the network, and the associated eigenvectors can be used to retrieval these patterns. Finally we showed how to use eigenvectors of the operator to make network sparser while keeping patterns stable.

In the field of neural networks, another important problem is the learning of neural networks, e.g. how to learn couplings of the network that remember random patterns. Authors in [33] proposed a methods for this task that

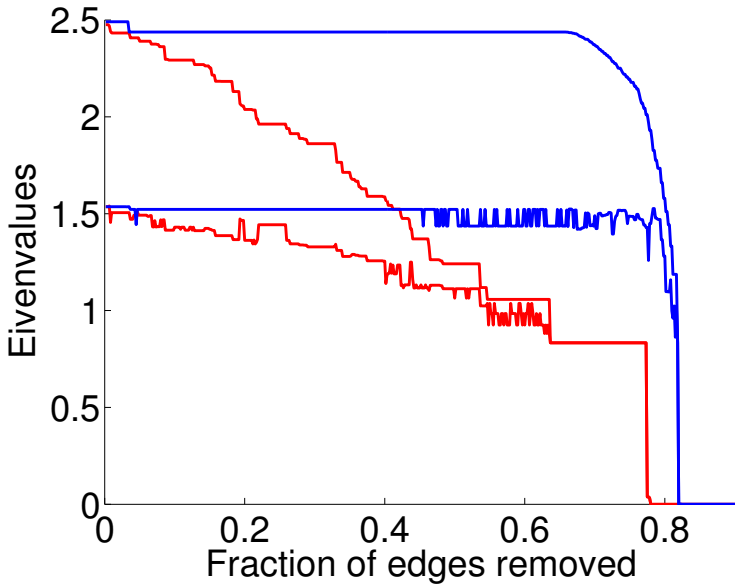


FIG. 5: (Color online) The largest eigenvalue and edge of bulk of a Hopfield network as a function of fraction of edge removed from the network. Red lines denote random removing and blue lines denote our approach. In both random removing and our approach, upper line represents the largest eigenvalue and the lower line represents edge of bulk, that is the absolute value of the first complex eigenvalue. In this network $n = 200, c = 6, P = 2, \beta = 1.2$.

works much better than Hebb's rule on sparse networks. Their method uses belief propagation to estimate correlation of samplings (patterns) and for learning of couplings. We see that it is possible to extend our result on Hopfield networks to the BP method in that learning method.

All above studies were made for Ising model without external fields. With external fields present, it has been shown [38] that paramagnetic solution still exists but with finite magnetization. It would be interesting to extend current study to that case.

If one compares the spectrum of non-backtracking operator of a Hopfield model with P patterns and that of a network generated by stochastic block model with P communities [9], one would find that they are quite similar. However the stochastic block model corresponds to a Potts model with P groups and one BP fixed-point, as opposed to Hopfield model which corresponds to Ising model with 2 groups and P BP fixed-points (states). So it would be interesting to study the connection in detail. Note that spectrum density of non-backtracking matrix of a graph has been computed in [36] using Belief Propagation. It would be interesting to compute non-backtracking matrix of Ising model using the same technique.

In Hopfield model, in the language of dynamics, our results correspond to standard Glauber dynamics. If one uses a different, e.g. time-delayed, dynamics [39], the capacity of the neural networks could be enhanced. However time-delayed dynamics is more complicated than parallel dynamics, so far we are not clear how it changes the spectrum of C. We think this is a interesting topic how to extend the non-backtracking matrix to the neural networks with different dynamics. We leave this for future work.

Acknowledgments

P.Z. was supported by AFOSR and DARPA under grant FA9550-12-1-0432. We are grateful to Ruben Andrist, Florent Krzakala, Cristopher Moore, Alaa Saade and Lenka Zdeborová for helpful conversations, and Federico Ricci-Tersenghi for discussing and pointing out references [37, 38].

[1] B. Simon, *The Statistical Mechanics of Lattice Gases, Volume I* (Princeton University Press Princeton, New Jersey, 1993).
[2] M. Mézard, G. Parisi, and M. Virasoro, *Spin glass theorey and beyond* (World Scientific, 1987).
[3] E. Schneidman, M. J. Berry, R. Segev, and W. Bialek, Nature **440**, 1007 (2006).

[4] J. J. Hopfield, Proc. Natl. Acad. Sci. USA **79**, 2554 (1982).
[5] D. H. Ackley, G. E. Hinton, and T. J. Sejnowski, Cognitive Science **9**, 147 (1985), ISSN 0364-0213.
[6] C. Castellano, S. Fortunato, and V. Loreto, Reviews of modern physics **81**, 591 (2009).
[7] M. Mezard and A. Montanari, *Information, Physics and Computation* (Oxford University press, 2009).
[8] J. Yedidia, W. Freeman, and Y. Weiss, Exploring artificial intelligence in the new millennium **8**, 236 (2003).
[9] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang, Proc. Natl. Acad. Sci. USA **110**, 20935 (2013), http://www.pnas.org/content/110/52/20935.full.pdf+html, URL `http://www.pnas.org/content/110/52/20935.abstract`.
[10] J. Yedidia, W. Freeman, and Y. Weiss, in *International Joint Conference on Artificial Intelligence (IJCAI)* (2001).
[11] M. Mézard and G. Parisi, Eur. Phys. J. B **20**, 217 (2001).
[12] P. Zhang and C. Moore, Proceedings of the National Academy of Sciences **111**, 18144 (2014). http://www.pnas.org/content/111/51/18144.full.pdf+html, URL `http://www.pnas.org/content/111/51/18144.abstract`.
[13] S. Edwards and P. Anderson, J. Phys. F **5**, 965 (1975).
[14] J. R. L. de Almeida and D. J. Thouless, J. Phys. A **11**, 983 (1978).
[15] H. Kesten and B. P. Stigum, The Annals of Mathematical Statistics **37**, 1211 (1966); H. Kesten and B. P. Stigum, The Annals of Mathematical Statistics **37**, 1463 (1966).
[16] S. Janson and E. Mossel, Annals of probability pp. 2630–2649 (2004).
[17] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, Phys. Rev. E **84**, 066106 (2011); A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, Phys. Rev. Lett. **107**, 065701 (2011).
[18] A. Saade, F. Krzakala, and L. Zdeborová, arXiv preprint arXiv:1406.1880 (2014).
[19] F. Ricci-Tersenghi, J. Stat. Mech. **2012**, P08015 (2012), URL `http://stacks.iop.org/1742-5468/2012/i=08/a=P08015`.
[20] D. Sherrington and S. Kirkpatrick, Phys. Rev. Lett **35**, 1792 (1975).
[21] L. Viana and A. Bray, J. Phys. C **18**, 3037 (1985).
[22] D. O. Hebb, *The organization of behavior: A neuropsychological approach* (John Wiley & Sons, 1949).
[23] R. J. Glauber, Journal of Mathematical Physics **4**, 294 (1963).
[24] D. J. Amit, *Modeling brain function: The world of attractor neural networks* (Cambridge University Press, 1992).
[25] D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. A **32**, 1007 (1985); D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. Lett. **55**, 1530 (1985).
[26] B. Wemmenhove and A. Coolen, J. Phys. A: Math Gen. **36**, 9617 (2003).
[27] M. Anthony and P. L. Bartlett, *Neural network learning: Theoretical foundations* (cambridge university press, 2009).
[28] P. N. McGraw and M. Menzinger, Phys. Rev. E **68**, 047102 (2003).
[29] B. J. Kim, Phys. Rev. E **69**, 045101 (2004).
[30] L. G. Morelli, G. Abramson, and M. N. Kuperman, Eur. Phys. J. B **38**, 495 (2004).
[31] P. Zhang and Y. Chen, Physica A **387**, 4411 (2008).
[32] P. Zhang and Y. Chen, Physica A **387**, 1009 (2008).
[33] A. Braunstein, A. Ramezanpour, R. Zecchina, and P. Zhang, Phys. Rev. E **83**, 056114 (2011).
[34] M. Mézard, G. Parisi, and R. Zecchina, Science **297**, 812 (2002).
[35] A. Decelle and P. Zhang, arXiv:1502.01660 (2015). URL `http://arxiv.org/abs/1502.01660`.
[36] A. Saade, F. Krzakala, and L. Zdeborová, Europhysics Letters **107**, 50005 (2014).
[37] J. M. Mooij and H.J. Kappen, Journal of Statistical Mechanics: Theory and Experiment **2005**, P11012 (2005).
[38] G. Parisi, F. Ricci-Tersenghi, and T. Rizzo, Journal of Statistical Mechanics: Theory and Experiment **2014**, P04013 (2014).
[39] P. Sen and B.K. Chakrabarti, Physics Letters A **162**, 327 (1992); P.K. Maiti, P.K. Dasgupta and B.K. Chakrabarti, International Journal of Modern Physics B **9**, 3025 (1995).

Appendix A: Computing edge of bulk in spectrum of C

Following the technique used in [9], the edge of bulk of eigenvalues of C can be computed by the following inequality,

$$\sum_{a=1}^{2m} |\lambda_a|^{2r} \leq \operatorname{Tr} C^r (C^r)^T, \tag{A1}$$

where λ_a denotes one eigenvalue, and there are $2m$ such eigenvalues. Note that $\left[C^r(C^r)^T\right]_{i\to j, i\to j}$ is contributed by paths connecting edge $i \to j$ with distance r-step away:

$$\left[C^r(C^r)^T\right]_{i\to j, i\to j} = \sum_U \prod_{(x\to y)\in U} \tanh^2(\beta J_{x\to j}),$$

where U denotes one such path and $(x \to y) \in U$ denotes edges belonging to path U. When n is large, and couplings $\{J_{ij}\}$ are i.i.d. distributed, for each path U we have

$$\prod_{(x\to y)\in U} \tanh^2(\beta J_{x\to j}) = \left\langle \tanh^2(\beta J_{i\to j})\right\rangle^r,$$

where average $\langle\cdot\rangle$ is taken over realizations of couplings, and there are $\hat{c}^r$ such paths with $\hat{c}$ denoting excess degree

$$\hat{c} = \sum_k \frac{p(k)k(k-1)}{\sum_k p(k)k} = \frac{< k^2 >}{c} - 1.$$

By taking expectation we have

$$\mathbb{E}\,\mathrm{Tr}\, C^r (C^r)^T = 2m\hat{c}^r \left\langle \tanh^2(\beta J_{i\to j})\right\rangle^r.$$

Then relation (A1) can be written as

$$\mathbb{E}\left(|\lambda|^{2r}\right) \le \left(\hat{c}\left\langle \tanh^2(\beta J_{i\to j})\right\rangle\right)^r, \tag{A2}$$

and it holds for any r.

Appendix B: computing eigenvalue associated with the ferromagnetic eigenvector

We can approximately recover this eigenvector by the following scheme: assume the graph is a tree with all its nodes associated with spin $S_k = 1$, or equivalently on the edge associated with leaves of the tree $S_{k\to l} = 1$.

By r steps iterating of C, we define an ferromagnetic "approximate" eigenvector $g^{(r)}$ as

$$g^{(r)} = \frac{1}{\mu^r} C^r S. \tag{B1}$$

If with $r \to \infty$ the vector $(g^{(r)})_{i\to j}$ is still correlated with $S_{k\to l}$, the all-one vector on the leave, system has a ferromagnetic phase where information is preserved during iterating. On a tree, last equation can be written as contribution of paths U connecting edge $i \to j$ and edge $k \to l$ distance r away:

$$(g^{(r)})_{i\to j} = \frac{1}{\mu^r} \sum_U \prod_{(x\to y)\in U} \tanh(\beta J_{xy}), \tag{B2}$$

where $(x \to y) \in U$ denotes edges belong to path U, and there are $\hat{c}^r$ such paths.

When n and r are large, if we assume the self-average property,

$$\prod_{\langle x,y\rangle} \tanh(\beta J_{xy}) = \langle \tanh(\beta J_{xy})\rangle^r, \tag{B3}$$

Eq. (B2) can be written as

$$(g_{i\to j})^{(r)} = \frac{1}{\mu^r} \sum_U \langle \tanh(\beta J_{xy})\rangle^r, \tag{B4}$$

where $\langle\cdot\rangle$ is again taking average over realizations of couplings.

Iterating C once gives

$$\left(Cg^{(r)}\right)_{i\to j} = \frac{1}{\mu^r} c^{r+1} \langle \tanh(\beta J_{xy})\rangle^{r+1} = \mu g^{(r)}_{i\to j}, \tag{B5}$$

$g_{i\to j}$ is indeed an eigenvector of C if $g^r_{i\to j} - g^{r+1}_{i\to j}$ goes to zero when r is large. In another words, $g_{i\to j}$ is an eigenvector if the noise propagation is slower than propagation of information $\{1, 1, .., 1\}$, that is $\mu > R$, which asks the informative eigenvalue of ferromagnetic phase out of the bulk.

Appendix C: Non-backtracking operator for Ising model on networks with community structures and its relation with Hessian of Bethe free energy

If the graph has community structures, e.g. the number of edges connecting same group is much larger than number of edges connecting to different groups, besides the eigenvector pointing in direction to all-one vector, there could be other eigenvectors with some positive and some negative entries, with signs representing group memberships. In this work we consider networks that generated by a generative model named Stochastic Block Model (SBM) or planted partition model [17]. SBM is one of the widely used model to generate benchmark networks containing community structure. In the model, there are q groups of nodes, each node i has a pre-assigned group label t_i^*, edges are generated independently according to a $q \times q$ matrix p by connecting each pair of nodes $\langle ij \rangle$ with probability $p_{t_i^*,t_j^*}$. Here for simplicity we discuss the commonly studied case that matrix p has only two distinct entries, $p_{\rm ab} = p_{\rm in}$ if $a = b$ and $p_{\rm out}$ if $a \neq b$, and we use $\epsilon = p_{\rm out}/p_{\rm in}$ to denote the ratio between these two entries. A phase transition at ϵ^* [17] has been discovered in this model that finite amount of information about planted partition $\{t^*\}$ can be inferred with $\epsilon < \epsilon^*$, and one can do the optimal inference of SBM to recover information of $\{t^*\}$ all the way down to the transition. So it is a good benchmark for testing performance of community detection algorithms. Here we consider the simple two-group case with equal group sizes. From [9] we know that in the detectable regime, edge of bulk of matrix B is $\sqrt{c}$, it has two real eigenvalues outside the bulk, the first one $\lambda_1 = 1$ corresponds to the eigenvector with entries having same sign. And the second one $\lambda_2 = \frac{1-\epsilon}{1+\epsilon}c$ corresponds to the eigenvector correlated with community structure. According to Eq. (11), edge of bulk of matrix C on the network is

$$R = \sqrt{c}\tanh\beta,$$

first eigenvalue is $\mu_1 = c\tanh\beta$, and the eigenvalue corresponding to the community structure is

$$\mu_2 = \frac{1-\epsilon}{1+\epsilon}c\tanh\beta.$$

So if $\mu_2 > 1$, the community structure in the network is detectable by Ising model (by excluding the ferromagnetic phase in some way).

Recall that Belief Propagation we used to derive the non-backtracking matrix at the beginning of the paper is nothing but an iterative method to decrease Bethe free energy; and a fixed point of Belief Propagation is a local minimum of Bethe free energy [10]. Obviously paramagnetic solution is always one local minimum of Bethe free energy. As we discussed above, eigenvalues larger than one in non-backtracking matrix tells us the instability of this paramagnetic fixed-point, which is also the instability of this local minimum of Bethe free energy. There is another way to study the instability of a minimum of free energy, which is Hessian matrix of Bethe free energy (we call it Hessian for simplicity) evaluated at paramagnetic fixed-point [18, 19], which reads

$$H_{ij} = \left.\frac{\partial^2 F_{\rm Bethe}}{\partial\psi_i\partial\psi_j}\right|_{\psi=\frac{1}{2}} = \left[1 + \sum_k \frac{\tanh^2(\beta J_{ik})}{1-\tanh^2(\beta J_{ik})}\right]\delta_{ij} - \frac{\tanh(\beta J_{ij})}{1-\tanh^2(\beta J_{ij})}. \tag{C1}$$

Detailed derivations can be found at e.g. [19]. For ferromagnets last equation reduces to

$$H = I + \frac{\tanh^2\beta}{1-\tanh^2\beta}D - \frac{\tanh\beta}{1-\tanh^2\beta}A. \tag{C2}$$

The paramagnetic phase is stable if this matrix is positive semidefinite. If it has negative eigenvalues, there must be other minimums of free energy at some point in the phase space, in direction that pointed by components of eigenvectors corresponding to negative eigenvalues. This scheme has been used in [9] to do spectral clustering on graphs where the Hessian is expressed as a function of r, with $r = \frac{1}{\tanh\beta}$ in our settings.

Given β and the graph, we have the following picture on the non-backtracking matrix and Hessian matrix: if $\mu_1 < 1$, only paramagnetic fixed-point is stable, Hessian is positive semi-definite. If $\mu_1 > 1$ and $\mu_2 < 1$, paramagnetic fixed-point is unstable towards the ferromagnetic phase, Hessian has one negative eigenvalues. If $\mu_2 > 1$, paramagnetic fixed-point is unstable towards both the ferromagnetic phase and the planted configuration (the community structure), Hessian has two negative eigenvalues. Then we see that non-backtracking matrix works equivalently to Hessian in determine the phases of ferromagnets. But we note that computing spectrum of Hessian is much faster than non-backtracking matrix, because Hessian is symmetric and size of Hessian is smaller ($n \times n$ instead of $2m \times 2m$ in non-backtracking matrix).

If we are not interested in the phase of Ising model but only the problem of node clustering in a given network, we can choose β as we want. As we shown above, β does not influence the shape of spectrum of non-backtracking

matrix, if $\mu_2 > R$, the second eigenvalue is always outside the bulk, thus we can always check whether there is a second eigenvalue outside the bulk to determine number of groups and use the eigenvector associated with it to detect the community structure. This is the idea explored in [9] which uses matrix B in doing spectral clustering.

The situation is different for Hessian, if β is too small, Hessian is positive semidefinite, it looses its ability of determining number of groups in network. So it is important to select a good β value that gives negative eigenvalues. The authors of [18] suggested to chose $r = \sqrt{c}$, equivalent to set $\tanh\beta = \frac{1}{\sqrt{c}}$ in our settings. As we discussed above, $r = \sqrt{c}$ is the point that edge of bulk reaches 1, thus if there is a second real-eigenvalue of C outside the bulk and is greater than 1, or equivalently a stable state correlated with community structure in Ising model, one should be able to find a second negative eigenvalue of H with $r = \sqrt{c}$.

On the one hand, detecting number of groups using real eigenvalues of B outside the bulk is easier than tunning β value to look for negative eigenvalues of H, especially on real-world networks. But on the other hand, tunning β gives H ability to optimize accuracy of detected communities, at least in synthetic networks.

Appendix D: Deriving Eq.(18)

After removing one edge, matrix C becomes $C_{\text{new}} = C + \Delta C$, an eigenvalue changes from λ to $\Delta\lambda$, a left eigenvalue changes from u to $u + \Delta u$, and a right eigenvector changes from v to $v + \Delta v$. Then we have the following relation

$$(C + \Delta C)(v + \Delta v) = (\lambda + \Delta\lambda)(v + \Delta v).$$

Ignoring second order terms, we have

$$\Delta C v + C\Delta v = \lambda\Delta v + \Delta\lambda v.$$

By multiplying left eigenvector u in both sides of last equation, we will finally arrive at the expression of change of eigenvalue:

$$\Delta\lambda = \frac{u\Delta C v}{uv}. \tag{D1}$$

In our problem ΔC is actually a matrix with entries zero except those with edges $i \to j$ and $j \to i$ involved. Then the last equation can be written as

$$\begin{aligned}
\Delta\lambda &= \frac{1}{uv}\sum_{i\to j,k\to l} u_{i\to j}\Delta C_{i\to j,k\to l}v_{k\to l} \\
&= \frac{1}{uv}\Bigg[u_{i\to j}\sum_{k\in\partial i\setminus j}\tanh(\beta J_{ik})v_{k\to i} + v_{i\to j}\sum_{l\in\partial j\setminus i}\tanh(\beta J_{jl})u_{j\to l} \\
&\quad + u_{j\to i}\sum_{l\in\partial j\setminus i}\tanh(\beta J_{jl})v_{l\to j} + v_{j\to i}\sum_{k\in\partial i\setminus j}\tanh(\beta J_{ik})u_{i\to k}\Bigg] \\
&= \frac{2\lambda}{uv}(u_{i\to j}v_{i\to j} + u_{j\to i}v_{j\to i}).
\end{aligned} \tag{D2}$$