Eur. Phys. J. B **50**, 431–437 (2006)
DOI: 10.1140/epjb/e2006-00155-4

THE EUROPEAN PHYSICAL JOURNAL B

Random field Ising model and community structure in complex networks

S.-W. Son[1], H. Jeong[1], and J.D. Noh[2,a]

[1] Department of Physics, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea
[2] Department of Physics, Chungnam National University, Daejeon 305-764, Korea

Received 27 August 2005 / Received in final form 7 November 2005
Published online 5 May 2006 –

Abstract. We propose a method to determine the community structure of a complex network. In this method the ground state problem of a ferromagnetic random field Ising model is considered on the network with the magnetic field $B_s = +\infty$, $B_t = -\infty$, and $B_{i\neq s,t} = 0$ for a node pair s and t. The ground state problem is equivalent to the so-called maximum flow problem, which can be solved exactly numerically with the help of a combinatorial optimization algorithm. The community structure is then identified from the ground state Ising spin domains for all pairs of s and t. Our method provides a criterion for the existence of the community structure, and is applicable equally well to unweighted and weighted networks. We demonstrate the performance of the method by applying it to the Barabási-Albert network, Zachary karate club network, the scientific collaboration network, and the stock price correlation network.

PACS. 89.75.Hc Networks and genealogical trees – 89.65.-s Social and economic systems – 05.10.-a Computational methods in statistical physics and nonlinear dynamics – 05.50.+q Lattice theory and statistics (Ising, Potts, etc.)

1 Introduction

Network theory is a useful tool for the study of complex systems. Universal features of some biological, social, and technological systems have been studied through their network structure [1–3]. Recent studies have revealed that some complex networks have the community structure, which means that highly interconnected nodes are clustered in distinct parts. The community may represent functional modules in biological networks [4–7], industrial sectors in economic networks [8,9], and coteries of intimate individuals in social networks [10].

Recently various methods have been suggested for finding out the community structure in a given network [11]. Girvan and Newman proposed an algorithm based on iterative removal of links with the highest betweenness centrality [10–12]. The betweenness centrality of an edge is given by the number of the pathways passing through it among shortest paths between all node pairs [13]. Nodes in different communities, if any, would be connected through rare inter-community links. Hence one could isolate communities by repeatedly removing links with the highest betweenness centrality. Similar methods were also considered in references [14–16]. Optimization techniques were also considered to find out the community structure. In those approaches, the community structure is found by optimizing an auxiliary quantity, such as the modularity [17,18].

Some physical problems turned out to be useful in detecting the community structure. Various spin systems embedded in a network have been used to study the community structure [19–23]. Fu and Anderson considered an Ising spin glass model for a graph partitioning [19]. Blatt et al. proposed the idea that multivariate data can be partitioned into clusters from correlation properties of a ferromagnetic Potts model [20]. Reichardt and Bornholdt suggested that community structures can be detected from low energy spin configurations of a Potts model with mixed ferromagnetic and antiferromagnetic couplings [21]. Also studied were random walks [24] and resistor network problems [25].

In this paper we propose a method for finding the community structure. Our approach is motivated by the observation on the Zachary network, a classical example of social networks with the community structure [10]. It is an acquaintance network of 34 members in a karate club. Once there arose a conflict between two influential members, which resulted in the breakup of the club into two. It is reasonable to think that the members would tend to minimize the number of broken ties, which can be accomplished by the breakup in accordance with the community structure. In fact, the resulting shape after the breakup coincides with the community structure of the original karate club network [10]. It suggests that the community

[a] e-mail: noh@cnu.ac.kr

structure of a given network may be found by simulating the breakup caused by an enforced frustration among nodes.

We simulate the breakup by studying the ferromagnetic random field Ising model (FRFIM): The Ising spins $\sigma_i = \pm 1$ are assigned to all nodes $i = 1, \ldots, N$, they interact ferromagnetically through links, and the quenched random magnetic field B_i is applied to each spin. The ferromagnetic interaction represents the cost of broken ties, and the random field introduces the frustration. In particular, we consider the case where the positive infinite magnetic field is applied to one spin and the negative infinite magnetic field to another. It amounts to imposing the boundary condition that the two spins are in opposite states. It simulates the conflict raised by the two members in the Zachary network. From this, we will identify the community structure from the ground state spin domain pattern of the FRFIM.

This paper is organized as follows. In Section 2 we introduce the FRFIM in general weighted networks. The ground state problem of the FRFIM can be solved exactly with a numerical algorithm, which will be explained in Appendix. Then the method for finding out the community structure is presented. In Section 3, we apply the method to several networks and present the results. We conclude the paper with summary and discussion in Section 4.

2 Method

Consider a weighted network G of N nodes. The connectivity of G can be represented with the weight matrix $\{J_{ij} | i,j = 1, \cdots, N\}$, where J_{ij} is a prescribed weight or strength of a link between nodes i and j if they are connected or $J_{ij} = 0$ otherwise. We assume that the weights are non-negative, $J_{ij} \geq 0$, and that the weights are symmetric, $J_{ij} = J_{ji}$. For an unweighted network, the matrix elements take the binary value 0 or 1, and the weight matrix reduces to the usual adjacency matrix.

The FRFIM on the network is defined with the Hamiltonian

$$H = -\frac{1}{2} \sum_{i,j} J_{ij} \sigma_i \sigma_j - \sum_i B_i \sigma_i, \quad (1)$$

where $\sigma_i = \pm 1$ is the Ising spin variable at each node i. The spins interact ferromagnetically with the coupling strength $\{J_{ij}\}$. They are also coupled with the quenched random magnetic field $\{B_i\}$.

The FRFIM model has been studied extensively in d dimensional regular lattices in order to investigate the nature of the glass phase transition (see Ref. [26] and references therein). It has also been used to investigate the disorder-driven roughening transition of interfaces in disordered media [27]. The phase transition in the FRFIM on complex networks, which would also be interesting, has not been studied so far. The issue will be studied elsewhere [28].

The specific feature of the FRFIM depends on the distribution of the random field $\{B_i\}$. In this work, we consider the simple yet informative magnetic filed distribution

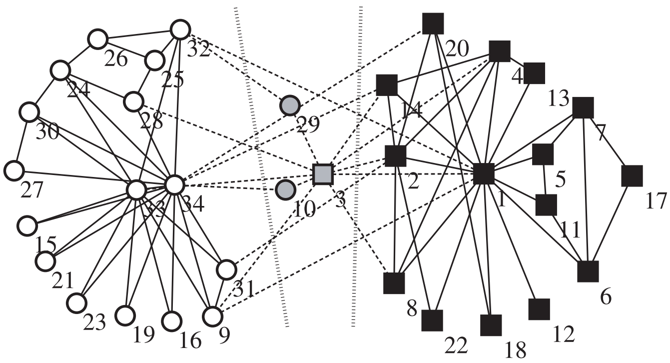

Fig. 1. Zachary karate club network. The links connecting nodes that are (not) in the same community are represented with solid (dashed) lines. The dotted lines separate the communities.

given by

$$B_i = \begin{cases} +\infty \ , \text{ for } i = s \\ -\infty \ , \text{ for } i = t \\ \ \ 0 \quad , \text{ for } i \neq s, t \end{cases} \quad (2)$$

for two nodes s and t. It amounts to imposing the boundary condition that $\sigma_s = +1$ and $\sigma_t = -1$, which induces frustration among the nodes. This specific random field distribution is adopted in order to mimic the conflict in the Zachary network. In the ground state, nodes are separated into different spin domains, which will be related to the community structure of the underlying network.

As an explicit example, we consider the Zachary karate club network which is illustrated in Figure 1. The node labeled as 1 (34) corresponds to the club instructor (administrator). They had a conflict, which resulted in the breakup. Nodes on the side of the administrator and the instructor after the breakup are denoted with circular and rectangular symbols, respectively. With $J_{ij} = 1$ for all links and the magnetic field given by equation (2) with $s = 1$ and $t = 34$, one can study the FRFIM on the network. Solving the ground state problem, we found that it has degenerate ground states: The black (white) nodes belong to the + (−) spin domain in all ground states, while the gray nodes (3, 10, 29) may belong to either domain. Note that the spin domains almost coincide with the actual shape of the network after the breakup; all black (white) nodes are on the side of the administrator (instructor). The gray nodes are in a marginal state. It is reasonable to think that they do not belong to either community. In the previous work [10], the node 3 was misclassified. Our result hints that it is due to the marginality.

The example clearly shows that the FRFIM is useful for finding the community structure. For a general application, (i) one needs to know the ground state(s) of the FRFIM of equation (1) with the quenched random magnetic field given in equation (2) for any node pair of s and t. Then, one needs to identify the set of all nodes that belong to the same spin domain as s and t in *all* ground states. Those sets will be called the *coteries* and denoted by $\mathcal{C}_s$ and $\mathcal{C}_t$, respectively. The number of nodes in the coterie $\mathcal{C}$ will be called the coterie size and denoted by $|\mathcal{C}|$. (ii) More importantly, one needs to specify the node pair s and t which is relevant to the community structure. An arbitrary choice of s and t will not provide any information on the community structure. For example, if we take $s = 12$ and $t = 15$ in the Zachary network in Figure 1, we obtain that $\mathcal{C}_s = \{12\}$ and all other nodes are in $\mathcal{C}_t$. This merely means that the node 12 is a peripheral node.

For (i), the ground state problem of the FRFIM can be solved exactly with the help of a numerical combinatorial optimization algorithm (see Appendix). This is achieved by mapping the ground state problem onto the minimum cut problem or the maximum flow problem [29]. The algorithm allows us to find all ground states, with which we can find the coteries $\mathcal{C}_s$ and $\mathcal{C}_t$ for any pair of s and t. We explain the detailed procedure in Appendix.

For (ii), the community structure can be found from the distribution of the coterie sizes for all pairs of s and t. For a certain pair of s and t, it may be that $|\mathcal{C}_s| \sim |\mathcal{C}_t| \sim \mathcal{O}(1) \ll N$. This happens when s and t are peripheral nodes of the network; most nodes are not influenced by them. Such a pair does not provide any information on the community structure. When $|\mathcal{C}_s| \sim \mathcal{O}(1) \ll |\mathcal{C}_t| \sim \mathcal{O}(N)$, s is a peripheral node while t is inside the bulk. The coteries $\mathcal{C}_s$ and $\mathcal{C}_t$ do not correspond to a community either. On the contrary, $\mathcal{O}(1) \ll |\mathcal{C}_s| \sim |\mathcal{C}_t| \sim \mathcal{O}(N)$, can occur only when there exist communities whose sizes are of the order of N, where s and t are chosen among "influential" nodes in different communities. In this case, we will regard the coteries $\mathcal{C}_s$ and $\mathcal{C}_t$ as the communities in the network.

In order to distinguish the different cases, we define the "separability" D_{st} for a node pair s and t as the product of the coterie sizes,

$$D_{st} = |\mathcal{C}_s| \cdot |\mathcal{C}_t| \; . \tag{3}$$

It ranges in the interval $1 \le D_{st} \le N^2/4$. We propose that the community structure be detected with the distribution of the separability D_{st} for all pairs of s and t. If $D_{st} \lesssim O(N)$ for all pairs of s and t, then we conclude that the network has no community structure. On the other hand, if $D_{st} \sim O(N^2)$ for a certain pair of s and t, then we conclude that the network consists of communities that can be identified from the coteries $\mathcal{C}_s$ and $\mathcal{C}_t$. Moreover, the nodes s and t may be regarded as the influential nodes of the communities. Therefore, in our method, the existence of the community structure is verified with the scaling behavior of the maximum value of the separability with the network size.

For a given network size N, the scaling can be examined with the quantity $\ln D_{st}/\ln N$. Without the community structure, it would be close to or much less than 1 for all node pairs. A node pair with $\ln D_{st}/\ln N > 1$ indicates the presence of the community structure.

3 Results

We have tested the method by applying it to the Barabási-Albert (BA) network [30], the Zachary karate club network [10], the scientific collaboration network [10], and the stock price correlation network [31]. In each network, the separability was calculated for all node pairs, and the separability distribution was examined with a so-called rank plot, where $[\ln D/\ln N]$ is plotted against a normalized rank of each node pair. The rank is assigned to each node pair in the ascending order of the separability. It is then normalized so that the rank of the pair with the maximum value of the separability is equal to 1.

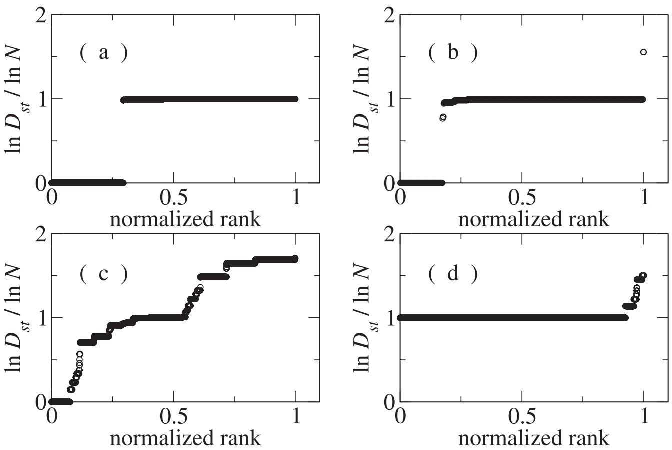

Fig. 2. The rank plot for the separability distribution for the BA network (a), the Zachary karate club network (b), the scientific collaboration network (c), and the stock price correlation network (d).

The BA network is an unweighted network. It is known that the BA network does not have a community structure. We grew a BA network of $N = 100$ nodes, and calculated the separability D_{st} for all node pairs. The separability distribution is presented as the rank plot in Figure 2a. We find that the separability is clustered at $D_{st} = 1$ and near $D_{st} \simeq N$ for all pairs of s and t, hence $\ln D_{st}/\ln N \lesssim 1$. This confirms that the BA network does not have the community structure (see Fig. 3a), and demonstrates the form of the separability distribution for networks without the community structure.

Next we studied the separability distribution of the Zachary karate club network of $N = 34$ nodes, which is presented in Figure 2b. We found that $D_{st} \lesssim N$ for all node pairs except $(1, 34)$ and $(1, 33)$. For the pairs $(s, t) = (1, 33)$ and $(1, 34)$, we obtained the same coteries, $\mathcal{C}_s$ of 15 nodes and $\mathcal{C}_t$ of 16 nodes, which are marked with the black and the white symbols in Figure 1, respectively (see also Fig. 3b). Therefore, we can conclude that there exist two communities in the network and that the node 1 is the influential node of one community and the nodes 33 and 34 are in the other community. In fact, nodes 1 and 34 correspond to the club instructor and the administrator, respectively. The detected communities are in good agreement with the network shape after the breakup.

We also investigated the community structure of a larger and more complex network. We examined the unweighted collaboration network of $N = 118$ scientists in the Santa Fe Institute [10]. In this network, two nodes (scientists) are linked if they coauthored at least one article. The rank plot is presented in Figure 2c. One can see

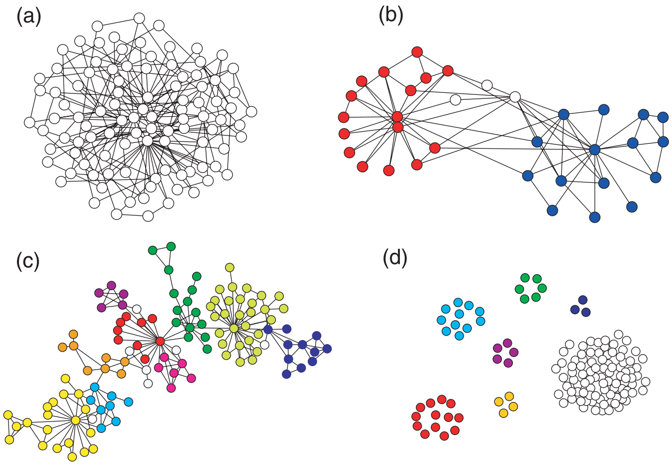

Fig. 3. (Color online) The community structure of (a) the BA model network ($N = 100$), (b) the Zachary karate club network ($N = 34$), (c) the scientific collaboration network ($N = 118$), and (d) the stock correlation network ($N = 137$). Nodes in different communities are distinguished with color. The white symbols represent the marginal nodes [33].

that the separability is distributed broadly, which indicates that the network has multiple (more than two) communities.

In such a case, the communities can be identified by applying our method hierarchically: first of all, one can find the node pair (s_0, t_0) with the largest separability, and the corresponding coteries $\mathcal{C}_{s_0}$ and $\mathcal{C}_{t_0}$. The coterie may consist of a single community or be the union of several sub-communities. In order to investigate the sub-structure, one constructs the sub-network which consists of all nodes and links within each coterie. Then, one can apply the method to the sub-networks. This can be performed hierarchically until a sub-network no longer has the community structure. Or one may proceed with the iteration only when the subnetwork size is equal to or larger than a threshold value m. The resulting coteries can then be identified as communities up to a resolution m.

With the hierarchical application of our method, we found the community structure of the scientific collaboration network as shown in Figure 3c. Here, we identify all communities whose size are equal to or larger than $m = 5$. The community structure is in good agreement with that found in reference [10].

Our method is also applicable to weighted networks. As an example of weighted networks, we studied the economic network of 137 companies in the New York Stock Exchange (NYSE) market. The network is constructed through the stock price return correlation between the companies for the 21 year period from 1983 to 2003 [31]. With the stock price $P_i(t)$ of a company i at time t, the return is given by $R_i(t) = \ln P_i(t+\Delta t) - \ln P_i(t)$ with the unit time interval Δt taken to be one day. Then, the stock price correlation is given by

$$C_{ij} = \frac{\langle (R_i - \langle R_i\rangle)(R_j - \langle R_j\rangle)\rangle}{\sqrt{(\langle R_i^2\rangle - \langle R_i\rangle^2)(\langle R_j^2\rangle - \langle R_j\rangle^2)}},$$

where the angular bracket indicates the time average over the period. Its value ranges in the interval $-1 \le C_{ij} \le 1$, and is large for strongly correlated company pairs. In fact, only 0.83% among all pairs of i and j have negative values of C_{ij} with negligible absolute values. Hence, one can neglect the effect of the negative correlation. It has been shown that the structural information of the economic system is encoded in the correlation matrix $\{C_{ij}\}$ [8,32].

In order to apply our method, all weights are required to be non-negative. Hence, we assume that the weight is given by $J_{ij} = e^{aC_{ij}}$ with a positive constant a taken to be 20. The weights are positive for all pairs of nodes, and the economic network is fully connected. The separability distribution is shown in the rank plot in Figure 2d. As in the collaboration network, there are several nontrivial separability levels. We identified all communities whose size is equal to or larger than 3 with the same hierarchical method as in the collaboration network. The resulting shape of the network is illustrated in Figure 3d. We confirmed that the communities are formed by companies in the same industrial sector categorized by Yahoo finance. For example, the largest community consists of 13 companies in the energy sector (red), which contains the following energy companies; HAL, KMG, NBL, COP, SLB, CVX, VLO, XOM, BP, RD, OXY, MRO, APA [34]. The second largest cluster corresponds to the group of 11 companies related to the electric utilities. And the others include health care (6 companies), basic material (5), rail road (4), and airline (3) [35]. This study shows that our method works well for weighted networks. We note that many nodes (white symbols) remain unclassified. We attribute it to the fully-connectedness of the network. Our method works better if there exist strong centers of communities. In a fully-connected network, a community could be isolated if there are much stronger centers than in a sparse network. This explains why our method found only a few communities.

4 Summary and discussion

In this paper we have proposed a method for finding the community structure of general networks. It is achieved by studying the ground state problem of the FRFIM on the networks with a magnetic field distribution given in equation (2) for two arbitrary nodes s and t. The coteries $\mathcal{C}_s$ and $\mathcal{C}_t$ are defined as the sets of all nodes that belong to the same spin domains as s and t in all possible degenerate ground states, respectively. The community structure is then manifested in the coterie pattern for the pair with the maximum value of the separability D_{st} defined in equation (3). Our method is motivated from the observation of the Zachary karate club network, which shows that the resulting shape of the network after breakup is determined by the underlying community structure. In our method, the response of the networks subject to schism is simulated with the FRFIM.

Our method is applicable to both unweighted and weight networks. Usually the existence of the community structure is tested with the modularity [17]. In our method we introduce the separability for that purpose: If the separability scales as $D_{st} \lesssim \mathcal{O}(N)$ for all node pairs as in the BA network, the network does not have a community structure. On the other hand, if $D_{st} \sim \mathcal{O}(N^2)$ for a certain pair of s and t, one can conclude that the network has a community structure and that the nodes s and t are influential nodes in each community. Figure 3 shows the performance of the method in real-world networks.

Our method has a different feature, which can be compared to others in terms of the betweenness centrality [10]. This can be illustrated by the following example network. Consider two m-regular graphs $G1$ and $G2$ with all nodes having m links. And then add l links between nodes in $G1$ and $G2$ to make a graph G. Obviously the new links have large values of the betweenness centrality for finite values of l. Hence, the methods based on the betweenness centrality would divide G into the sum of $G1$ and $G2$. On the contrary, our method divides the graph G only when $m \geq l$, that is to say, when there are nodes whose degree is larger than the the number of the interlinks. This shows that our method has a stricter condition for the community structure than others. In the context of the FRFIM, for $m < l$, all other nodes are not influenced by the random field applied to any node pair. Our method claims that the entire graph G constitutes a single community itself rather than being decomposed into two communities.

One of the weak points of our method is the time complexity. Practically, the ground state problem of the FRFIM in sparse networks of N nodes has the time complexity of $\mathcal{O}(N^{\theta})$ with $\theta \simeq 1.2$ [29]. Since one has to solve the ground state problems for all magnetic field distributions, the total time complexity scales as $\mathcal{O}(N^{2+\theta})$. Hence, in the practical sense, our method is limited to networks of up to a few thousands of nodes. One may avoid the time complexity problem if the important nodes are known a priori. In network theory, the importance of nodes can be measured by, e.g., the degree or the betweenness centrality. Hopefully the community structure of large networks can be studied if one incorporates such important measures into our method.

In the present work, we adopted the specific form of the magnetic field as in equation (2). One may consider the FRFIM with a more general magnetic field distribution. We expect that the ground state property depends on the community structure of an underlying network, which can then be used to detect the community structure. We leave the extension for a future work.

This work was supported by Korea Research Foundation Grand(KRF-2003-003-C00091). JDN would like to thank KIAS for the hospitality during the visit.

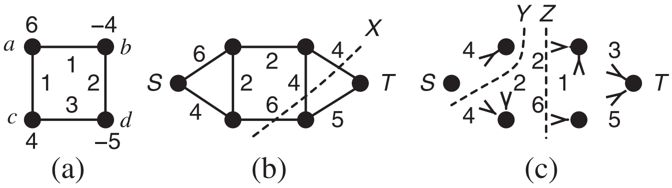

Fig. 4. (a) A network G with 4 nodes (filled circles) and 4 links (lines) for the FRFIM. Figures represent the magnetic fields and the interaction strengths, respectively. (b) The corresponding capacitated network G' with the link capacities. (c) The maximum-flow configuration with $v^\star = 8$. The dotted lines represent saturated links with $x^\star_{\alpha\beta} = c_{\alpha\beta}$. The dashed lines X, Y, and Z represent boundaries associated with ST-cuts.

Appendix: Minimum cut and maximum flow problem

This Appendix is intended to introduce the combinatorial optimization algorithm for solving the ground state problem of the FRFIM. For a more rigorous description, we refer the readers to reference [29].

Consider a network G of N nodes with the symmetric weight matrix $\{J_{ij} \geq 0\}$ $(i, j = 1, \cdots, N)$. The ferromagnetic random field Ising model on G is defined by the Hamiltonian in equation (1) with the quenched random magnetic field $\{B_i\}$. The ground state is the spin configuration that has the minimum energy among all 2^N configurations. One might find the ground state by enumerating all spin configurations, which is obviously time consuming and inefficient. We will explain the efficient way for solving the ground state problem.

It is useful to introduce a capacitated network denoted by G': having all the nodes and links of G, G' contains two additional nodes S, called the source, and T, called the sink, and additional links between the source (sink) and the nodes with the positive (negative) magnetic field. G' is also a weighted network with the symmetric weight matrix $\{c_{\alpha\beta}\}$ $(\alpha, \beta = S, T, 1, \cdots, N)$. For a link (ij) from the original network G, the weight is given by $c_{ij} = 2J_{ij}$. For the additional link, the weight is given by $c_{Si} = B_i$ for all i with $B_i > 0$ and $c_{iT} = |B_i|$ for all i with $B_i < 0$. The weight of the network G' is usually called the capacity. Figure 4 illustrates the relation between a network G of four nodes $\{a, b, c, d\}$ and the corresponding capacitated network G'.

In the capacitated network G' we define a ST-cut as a decomposition of all nodes into two disjoint sets $\mathcal{S}$ and $\mathcal{T}$ with $S \in \mathcal{S}$ and $T \in \mathcal{T}$. It will be denoted by $[\mathcal{S}, \mathcal{T}]$. For a given $[\mathcal{S}, \mathcal{T}]$, some links connect nodes in the different sets. The set of such links forms the boundary of the cut, which is denoted by $(\mathcal{S}, \mathcal{T}) = \{(\alpha\beta) | \alpha \in \mathcal{S}, \beta \in \mathcal{T}\}$. The cut capacity $C[\mathcal{S}, \mathcal{T}]$ is then defined as the total sum of the capacity of the boundary links, that is,

$$C[\mathcal{S}, \mathcal{T}] = \sum_{(\alpha\beta) \in (\mathcal{S}, \mathcal{T})} c_{\alpha\beta} \ . \quad (4)$$

Figure 4 shows some examples of the cut. The boundary denoted by X is associated with a cut $[\{S,a,b,c\},\{T,d\}]$, whose cut capacity is 14.

There exists an one-to-one correspondence between the Ising spin configuration on the weighted network G and the cut $[\mathcal{S},\mathcal{T}]$ of the capacitated network G'. It is achieved by assigning $\sigma_i = +1(-1)$ for all nodes i in $\mathcal{S}(\mathcal{T})$ and vice versa. Hence, the sets $\mathcal{S}$ and $\mathcal{T}$ correspond to up and down spin domains, respectively, and the boundary $(\mathcal{S},\mathcal{T})$ corresponds to the spin domain wall. Furthermore, one can easily verify that the energy E of the FRFIM of a spin configuration $\{\sigma_i\}$ and the cut capacity $C[\mathcal{S},\mathcal{T}]$ satisfy the relation

$$E(\{\sigma_i\}) = C[\mathcal{S},\mathcal{T}] + E_0 \qquad (5)$$

where $E_0 = -\sum_{i,j} J_{ij}/2 - \sum_i |B_i|/2$. Therefore, solving the ground state of the FRFIM on G is equivalent to finding the optimal ST-cut on G' whose cut capacity is minimum. It is called the *minimum cut problem*.

The minimum cut problem can be further mapped on to the *maximum flow problem*: On the capacitated network G', a flow denotes a set of flow variables $\{x_{\alpha\beta}\}$ defined for all links in G' which are subject to a capacity constraint

$$0 \le x_{\alpha\beta} \le c_{\alpha\beta} \qquad (6)$$

and a mass balance constraint

$$\sum_\beta{}' x_{\alpha\beta} - \sum_\beta{}' x_{\beta\alpha} = v\delta(\alpha,S) - v\delta(\alpha,T). \qquad (7)$$

Here $\sum'$ means a sum over all adjacent nodes of α, $\delta()$ denotes the Kronecker δ symbol, and v is a non-negative parameter. The mass balance constraint allows us to interpret the flow $\{x_{\alpha\beta}\}$ as a conserved flux configuration of, e.g., a fluid which originated from the source S by the amount of v and targeted to the sink T through the network G'.

Due to the capacity constraint, there exists the upper bound in v, beyond which a flow satisfying equations (6) and (7) does not exist. Then, the question that arises naturally is to find the maximum value $v^\star$ and the corresponding flow $\{x^\star_{\alpha\beta}\}$ that can be delivered. This is the maximum flow problem.

The celebrated max-flow/min-cut theorem of Ford and Fulkerson [36] states that for a given capacitated network G', the maximum flow $v^\star$ is equal to the minimum cut capacity, that is to say,

$$v^\star = \min_{[\mathcal{S},\mathcal{T}]} C[\mathcal{S},\mathcal{T}] \ . \qquad (8)$$

The rigorous proof of the theorem can be found elsewhere [29]. Intuitively the theorem states that the maximum flow is limited by the bottleneck in the network whose capacity is given by the minimum cut capacity.

The maximum flow problem can be solved numerically in a polynomial time with the augmenting path algorithm or the preflow-push/relabel algorithm [29,36]. In the augmenting path algorithm, one repeatedly searches for a path from S to T via *unsaturated* $(x_{\alpha\beta} < c_{\alpha\beta})$ links and updates $\{x_{\alpha\beta}\}$ by augmenting flows along the path. When the augmenting path does not exist any more, the resulting flow corresponds to the maximum flow configuration. The preflow-push/relabel algorithm is a more sophisticated and efficient algorithm.

Once the maximum flow configuration $\{x^\star_{\alpha\beta}\}$ is found, the minimum cut is constructed easily. Let $\mathcal{S}_S$ be the set of all nodes of G' that can be *reachable* from the source S only through *unsaturated* $(x^\star_{\alpha\beta} < c_{\alpha\beta})$ links. Trivially, $\mathcal{S}_S$ does not include the sink T, since there does not exist any augmenting path in the maximum flow configuration. Hence, the set $\mathcal{S}_S$ and its complement $\overline{\mathcal{S}_S}$ defines a cut $[\mathcal{S}_S,\overline{\mathcal{S}_S}]$, which is indeed a minimum cut of G'.

One may find the minimum cut in another way. Let $\mathcal{T}_T$ be the set of all nodes of G' that can be *reachable* from the sink T only through *unsaturated* links. Then, $\mathcal{T}_T$ and its complement $\overline{\mathcal{T}_T}$ defines a cut $[\overline{\mathcal{T}_T},\mathcal{T}_T]$, which is also the minimum cut.

The two cuts $[\mathcal{S}_S,\overline{\mathcal{S}_S}]$ and $[\overline{\mathcal{T}_T},\mathcal{T}_T]$ may be different, which implies that the corresponding FRFIM has degenerate ground states. In that case, all degenerate ground states can be found systematically [29]. In this work, we are interested in the spins that are fixed in all ground states. One can easily verify that all nodes $i \in \mathcal{S}_S$ $(\mathcal{T}_T)$ except for S (T) are in the spin state $\sigma_i = +1$ (-1) in all ground states. The other nodes $j \notin \mathcal{S}_S$ and $\mathcal{T}_T$ may be in either state $\sigma_j = \pm 1$.

We provide an example illustrating the mapping between the FRFIM and the maximum flow or the minimum cut problem in Figure 4. The maximum flow configuration is depicted in Figure 4c with the maximum flow $v^\star = 8$. The links drawn with dotted lines are saturated $(x^\star_{\alpha\beta} = c_{\alpha\beta})$. The sets of all nodes that are reachable from S and T through unsaturated links are given by $\mathcal{S}_S = \{S,a\}$ and $\mathcal{T}_T = \{T,b,d\}$. They yield the minimum cuts $[\mathcal{S}_S,\overline{\mathcal{S}_S}]$ and $[\overline{\mathcal{T}_T},\mathcal{T}_T]$ whose boundaries are Y and Z, respectively. Hence, one finds that $\sigma_a = +1$ and $\sigma_b = \sigma_d = -1$ in all degenerate ground states. The node c does not belong to either $\mathcal{S}_S$ or $\mathcal{T}_T$. Hence σ_c may be either $+1$ or -1.

In the present work, we considered the FRFIM on a weighted network G with the specific magnetic field distribution given in equation (2) for a certain node pair s and t. Then, we need to find the coterie $\mathcal{C}_s$ $(\mathcal{C}_t)$ of s (t) which is the set of all nodes that are in the same spin state as s (t) in the ground state. We can summarize the method to find the coteries:

1. Construct the capacitated network G'.
2. Find the maximum flow configuration $\{x^\star_{\alpha\beta}\}$ using the numerical algorithms.
3. Find the set $\mathcal{S}_S$ $(\mathcal{T}_T)$ of all nodes that are reachable from S (T) through unsaturated links with $x^\star_{\alpha\beta} < c_{\alpha\beta}$.
4. Then, the coteries are given by $\mathcal{C}_s = \mathcal{S}_S - \{S\}$ and $\mathcal{C}_t = \mathcal{T}_T - \{T\}$.

After finding the coteries, the community structure can be investigated with the method explained in Section 2.

References

1. R. Albert, A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002)
2. S.N. Dorogovtsev, J.F.F. Mendes, Adv. Phys. **51**, 1079 (2002)
3. M.E.J. Newman, SIAM Rev. **45**, 167 (2003)
4. H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, A.-L. Barabási, Nature (London) **407**, 651 (2000)
5. P. Holme, M. Huss, H. Jeong, Bioinformatics **19**, 532 (2003)
6. D. Wilkinson, B.A. Huberman, Proc. Natl. Acad. Sci. **101**, 5241 (2004)
7. E. Ravasz, A.L. Somera, D.A. Mongru, Z.N. Oltvai, A.-L. Barabási, Science **297**, 1551 (2002); E. Ravasz, A.-L. Barabási, Phys. Rev. E **67**, 026112 (2003)
8. R.N. Mantegna, Eur. Phys. J. B **11**, 193 (1999); G. Bonanno, G. Caldarelli, F. Lillo, R.N. Mantegna, Phys. Rev. E **68**, 046130 (2003)
9. J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertesz, A. Kanto, Phys. Rev. E **68**, 056110 (2003)
10. M. Girvan, M.E.J. Newman, Proc. Natl. Acad. Sci. **99**, 7821 (2002)
11. M.E.J. Newman, Eur. Phys. J. B **38**, 321 (2004)
12. M.E.J. Newman, M. Girvan, Phys. Rev. E **69**, 026113 (2004)
13. M.E.J. Newman, Phys. Rev. E **64**, 016131 (2001); M.E.J. Newman, Phys. Rev. E **64**, 016132 (2001)
14. J.R. Tyler, D.M. Wilkinson, B.A. Huberman, e-print `arXiv:comd-mat/0303264` (2003)
15. F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, D. Parisi, Proc. Natl. Acad. Sci. **101**, 2658 (2004)
16. S. Fortunato, V. Latora, M. Marchiori, Phys. Rev. E **70** 056104 (2004)
17. M.E.J. Newman, Phys. Rev. E **69**, 066133 (2004)
18. A. Clauset, M.E.J. Newman, C. Moore, Phys. Rev. E **70**, 066111 (2004)
19. Y. Fu, P.W. Anderson, J. Phys. A **19**, 1605 (1986)
20. M. Blatt, S. Wiseman, E. Domany Phys. Rev. Lett. **76**, 3251 (1996)
21. J. Reichardt, S. Bornholdt, Phys. Rev. Lett. **93**, 218701 (2004)
22. A. Vazquez, A. Flammini, A. Maritan, A. Vespignani, Nat. Biotechnol. **21**, 697 (2003)
23. R. Guimerá, M. Sales-Pardo, L.A.N. Amaral, Phys. Rev. E **7-**, 025101(R) (2004)
24. H. Zhou, Phys. Rev. E **67**, 061901 (2003)
25. F. Wu, B.A. Huberman, Eur. Phys. J. B **38**, 331 (2004)
26. A.A. Middleton, D.S. Fisher, Phys. Rev. B **65**, 134411 (2002)
27. J.D. Noh, H. Rieger, Phys. Rev. Lett. **87**, 176102 (2001); J.D. Noh, H. Rieger, Phys. Rev. E **66**, 036117 (2002)
28. S.-W. Son, H. Jeong, J.D. Noh, unpublished
29. M. Alava, P.M. Duxbury, C. Moukarzel, H. Rieger, in *Phase Transitions and Critical Phenomena*, edited by C. Domb, J.L. Lebowitz (Academic, Cambridge, 2000) Vol. **18**, pp. 141–317; A. Hartmann, H. Rieger, *Optimization Algorithms in Physics* (Wiley VCH, Berlin, 2002)
30. A.-L. Barabási, R. Albert, Science **286**, 509 (1999); A.-L. Barabási, R. Albert, H. Jeong, Physica A **272**, 173 (1999)
31. D.-H. Kim, H. Jeong, Phys. Rev. E **72**, 046133 (2005)
32. J.D. Noh, Phys. Rev. E **61**, 5981 (2000)
33. For Figure 3c, yellow and skyblue correspond to a group of scientists working primarily on the structure of RNA, orange, red, violet, pink, and green correspond to a group working primarily in statistical physics, lightgreen represents a group working on mathematical models in ecology, blue represents a group of scientists using agent-based models to study problems in economics and traffic flow
34. The tickers correspond to the following companies in the NYSE: Halliburton Co. (HAL), Kerr-Mc-Gee Corp. (KMG), Noble Energy Inc. (NBL), ConocoPhillips (COP), Schlumberger Ltd. (SLB), Chevron Texaco Corp. (CVX), Valero Energy Corp. (VLO), Exxon Mobil Corp. (XOM), BP PLC(BP), Royal Dutch Petroleum Co. (RD), Occidental Petroleum Corp. (OXY), Marathon Oil Corp. (MRO), and Apache Corp. (APA)
35. Each industrial cluster contains the following companies. Utilities (11 companies, skyblue): Southern Company Inc. (SO), Public Service Enterprise Group Inc. (PEG), PG&E Corp. (PCG), Exelon Corp. (EXC), Entergy Corp. (ETR), Edison International (EIX), American Electric Power Co. Inc. (AEP), Consolidated Edison Inc. (ED), DTE Energy Co. (DTE), CenterPoint Energy Inc. (CNP), and People Energy Corp. (PGL). Health care (6, green): Merck&Co. Inc. (MRK), Wyeth (WYE), Bristol-Myers Squibb Co. (BMY), Johnson&Johnson Inc. (JNJ), Eli Lilly and Co. (LLY), and Pfizer Inc. (PFE). Basic material (5, violet): Boise Cascade Corp. (BCC), Georgia-Pacific Corp. (GP), Louisiana-Pacific Corp. (LPX), Weyerhaeuser Co. (WY), and International Paper Co. (IP). Rail road (4, orange): Union Pacific Corp. (UNP), CSX Corp. (CSX), Burlington Northern Santa Fe Corp. (BNI), and Norfolk Southern Corp. (NSC). Airline (3, blue): AMR Corp. (AMR), Delta Air Lines Inc. (DAL), and Southwest Airlines Inc. (LUV).
36. L.R. Ford, D.R. Fulkerson *Flows in Networks* (Princeton University Press, 1962)