

Stochastic block model and exploratory analysis in signed networks

Jonathan Q. Jiang^{*}*Department of Computer Science, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong, China*

(Received 2 January 2015; published 15 June 2015)

We propose a generalized stochastic block model to explore the mesoscopic structures in signed networks by grouping vertices that exhibit similar positive and negative connection profiles into the same cluster. In this model, the group memberships are viewed as hidden or unobserved quantities, and the connection patterns between groups are explicitly characterized by two block matrices, one for positive links and the other for negative links. By fitting the model to the observed network, we can not only extract various structural patterns existing in the network without prior knowledge, but also recognize what specific structures we obtained. Furthermore, the model parameters provide vital clues about the probabilities that each vertex belongs to different groups and the centrality of each vertex in its corresponding group. This information sheds light on the discovery of the networks' overlapping structures and the identification of two types of important vertices, which serve as the cores of each group and the bridges between different groups, respectively. Experiments on a series of synthetic and real-life networks show the effectiveness as well as the superiority of our model.

DOI: [10.1103/PhysRevE.91.062805](https://doi.org/10.1103/PhysRevE.91.062805)

PACS number(s): 89.75.Fb, 05.10.-a

I. INTRODUCTION

The study of networks has received considerable attention in recent literature [1–3]. This is mainly attributed to the fact that a network provides a concise mathematical representation for social [4,5], technological [6], biological [7–9], and other complex systems [1–3] in the real world, which paves the way for executing proper analysis of such systems' organizations, functions, and dynamics.

Many networks are found to possess a multitude of mesoscopic structural patterns, which can be coarsely divided into “assortative” or “community” structure and “disassortative” or “bipartite or multipartite” structure [10,11]. In addition, other types of mesoscopic structures, such as the “core-periphery” motif, have been observed in real-life networks as well. Along with these discoveries, a large number of techniques have been proposed for mesoscopic structure extraction, in particular for community detection (see, e.g., [8,10–14] and recent reviews [2,3,15]). Most, if not all, existing techniques require us to know which specific structure we are looking for before we study it. Unfortunately, we often know little about a given network and have no idea what specific structures can be expected and subsequently detected by what specific methods. Biased results will be obtained if an inappropriate method is chosen. Even if we know something beforehand, it is still difficult for a method that is exclusively designed for a certain type of mesoscopic structure to uncover the aforementioned miscellaneous structures that may simultaneously coexist in a network or may even overlap with each other [8,16–20].

To overcome these difficulties, a mixture model [11], a stochastic block model [21], and their various extensions and combinations [22–27] have been recently introduced to enable an “exploratory” analysis of networks, allowing us to extract unspecified structural patterns even if some edges in the networks are missing [28,29]. By fitting the model to the

observed network structure, vertices with the same connection profiles are categorized into a predefined number of groups. The philosophy of these approaches is quite similar to that of the “role model” in sociology [30]—individuals having locally or globally analogous relationships with others play the same “role” or take up the same “position” [31]. It is clear to see that the possible topologies of the groups include community structure and multipartite structure, but they can be much, much wider.

One common assumption shared by these models is that the target networks contain positive links only. However, we frequently encounter the signed networks, which have both positive and negative edges, in biology [19,32], computer science [33], and last but definitely not least, social science [34–37]. The negative connections usually represent hostility, conflict, opposition, disagreement, and distrust between individuals or organizations, as well as the anticorrelation among objectives, whose coupled relation with positive links has been empirically shown to play a crucial role in the function and evolution of the whole network [32,37].

Several works have been conducted to detect community structure in these kinds of networks. Yang *et al.* [34] proposed an agent-based method that performs a random walk from one specific vertex for a few steps to mine the communities in positive and signed networks. Gómez *et al.* [35] presented a generalization of the widely used modularity [10,14] to allow for negative links. Traag and Bruggeman [36] extended the Potts model to incorporate negative edges, resulting in a method similar to the clustering of signed graphs. These approaches focus on the problem of community detection and thus they inevitably suffer a devastating failure if the signed networks comprise other structural patterns, for example the disassortative structure, as shown in Sec. IV A. To make matters worse, they simply give a “hard” partition of signed networks in which a specific vertex could belong to one and only one cluster. Similar to the positive networks, we have good reason to believe that the signed networks also simultaneously include all kinds of mesoscopic structures that might overlap with each other.

^{*}Present address: Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong, China; csai.jiang@gmail.com

In this paper, we aim to capture and extract the intrinsic mesoscopic structure of networks with both positive and negative links. This goal is achieved by dividing the vertices into groups such that the vertices within each group have similar positive and negative connection patterns to other groups. We propose a generalized stochastic block model, referred to as the signed stochastic block model (SSBM), in which the group memberships of each vertex are represented by unobserved or hidden quantities, and the relationship among groups is explicitly characterized by two block matrices, one for the positive links and the other for the negative links. By using the expectation-maximization algorithm, we fit the model to the observed network structure and reveal the structural patterns without prior knowledge of what specific structures exist in the network. As a result, not only can various unspecific structures be successfully found, but also their types can be immediately elucidated by the block matrices. In addition, the model parameters tell us the fuzzy group memberships and the centrality of each vertex, which enable us to discover the networks' overlapping structures and to identify two kinds of important vertices, i.e., group core and bridge. Experiments on a number of synthetic and real-world networks validate the effectiveness and the advantage of our model.

The rest of this paper is organized as follows. We begin with the depictions of the mesoscopic structures, especially the definitions of the community structure and disassortative structure, in signed networks in Sec. II. Then we introduce an extension of the stochastic block model in Sec. III, and we show how to employ it to perform an exploratory analysis of a given network with both positive and negative links. Experimental results on a series of synthetic networks with various designed structures and three social networks are given in Sec. IV, followed by the conclusions in Sec. V.

II. MESOSCOPIC STRUCTURES IN SIGNED NETWORKS

It is well known that the mesoscopic structural patterns in positive networks can be roughly classified into the following two different types: "Assortative structure," usually called "community structure" in most cases, refers to groups of vertices within which connections are relatively dense and between which they are sparser [10–12]. In contrast, "disassortative structure," also named "bipartite structure" or more generally "multipartite structure," means that network vertices have most of their connections outside their group [10, 11, 13].

For a signed network, its mesoscopic structure is quite different from and much more complicated than that in a positive network since both the density and the sign of the links should be taken into account at the same time. The intuitive descriptions of assortative structure and disassortative structure given in Refs. [10, 11] are no longer suitable. A natural question arises: How can we characterize the mesoscopic structures in a network that has both positive and negative edges? Guidance can be provided by the social balance theory [38], which states that the attitudes of two individuals toward a third person should match if they are positively related. In this situation, the triad is said to be socially balanced. A network is called balanced provided that all its triads are balanced. This concept can be further generalized

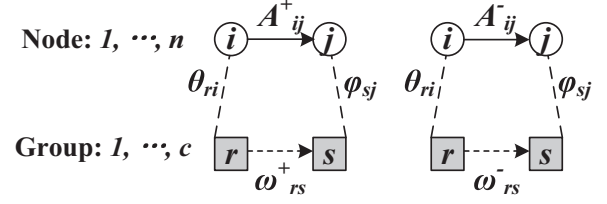


FIG. 1. Stochastic block model for signed networks. Unfilled circles represent observed network structure and filled ones correspond to hidden memberships. The solid line between vertices i and j indicates the existence of one positive or negative edge connecting them. The dashed line indicates that the relation between the corresponding quantities is unobserved and requires being learned from the observed network data.

to k -balance [39, 40] when the network can be divided into k clusters, each having only positive links within itself and negative links with others.

Following this principle, we can reasonably describe the community structure in a signed network as a set of groups of vertices within which positive links are comparatively dense and negative links are sparser, and on the contrary between which positive links are much looser and negative links are thicker [34–36]. Obviously, this is an extension of the standard community structure in networks with positive edges. In contrast, disassortative structure can be defined as a collection of vertices that have most of their negative links within the group to which they belong while the majority of their positive connections are outside their group.

III. METHODS

A. The SSBM model

Given a directed network $G = (V, E)$, we can represent it by an adjacency matrix A . The entries of the matrix are defined as follows: $A_{ij} = 1$ if a positive link is present from vertex i to vertex j , $A_{ij} = -1$ if a negative link is present from vertex i to vertex j , and $A_{ij} = 0$ otherwise. For weighted networks, A_{ij} can be generalized to represent the weight of the link. We further separate the positive component from the negative one by setting $A_{ij}^+ = A_{ij}$ if $A_{ij} > 0$ and 0 otherwise, and $A_{ij}^- = -A_{ij}$ if $A_{ij} < 0$ and 0 otherwise, so $A = A^+ - A^-$.

Suppose that the vertices fall into c groups whose memberships are "hidden" or "missing" for the moment and will be inferred from the observed network structure. The number of groups c can also be inferred from the data, which will be discussed in Sec. III C, but we take it as a given here. The standard solution for such an inference problem is to give a generative model for the observed network structure and then to determine the parameters of the model by finding its best fit [11, 22–25].

The model we use is a kind of stochastic block model that parametrizes the probability of each possible configuration of group assignments and edges as follows (see Fig. 1 for a schematic illustration). Given an edge e_{ij} , we choose a pair of groups r and s for its tail and head with probability ω_{rs}^+ if e_{ij} is positive, or with probability ω_{rs}^- if e_{ij} is negative. The two scalars ω_{rs}^+ and ω_{rs}^- giving the probability that a

randomly selected positive and negative edge from group r to s , respectively, explicitly characterize various types of connecting patterns among groups, as we will see later. Then, we draw the tail vertex i from group r with probability θ_{ri} and the head vertex j from group s with probability ϕ_{sj} . Intuitively, the parameter θ_{ri} captures the centrality of vertex i in the group r from the perspective of outgoing edges, while ϕ_{sj} describes the centrality of vertex j in the group s from the perspective of incoming edges. The parameters ω_{rs}^+ , ω_{rs}^- , θ_{ri} , and ϕ_{sj} satisfy the normalization condition

$$\sum_{r=1}^c \sum_{s=1}^c \omega_{rs}^+ = 1, \quad \sum_{r=1}^c \sum_{s=1}^c \omega_{rs}^- = 1, \\ \sum_{i=1}^n \theta_{ri} = 1, \quad \sum_{j=1}^n \phi_{sj} = 1.$$

Let \overleftarrow{g}_{ij} and \overrightarrow{g}_{ij} be the group membership of the tail and head of the edge e_{ij} , respectively. So far, we have introduced all the quantities in our model: observed quantities $\{A_{ij}\}$, hidden quantities $\{\overleftarrow{g}_{ij}, \overrightarrow{g}_{ij}\}$, and model parameters $\{\omega_{rs}^+, \omega_{rs}^-, \theta_{ri}, \phi_{sj}\}$. To simplify the notations, we shall henceforth denote by ω^+ the entire set $\{\omega_{rs}^+\}$ and similarly ω^- , θ , ϕ , \overleftarrow{g} , and \overrightarrow{g} for $\{\omega_{rs}^-\}$, $\{\theta_{ri}\}$, $\{\phi_{sj}\}$, $\{\overleftarrow{g}_{ij}\}$, and $\{\overrightarrow{g}_{ij}\}$. The probability that we observe a positive edge e_{ij}^+ can be written as

$$\Pr(e_{ij}^+ | \omega^+, \theta, \phi) = \sum_{rs} \omega_{rs}^+ \theta_{ri} \phi_{sj}, \quad (1)$$

and the probability of observing a negative edge e_{ij}^- is

$$\Pr(e_{ij}^- | \omega^-, \theta, \phi) = \sum_{rs} \omega_{rs}^- \theta_{ri} \phi_{sj}. \quad (2)$$

The marginal likelihood of the signed network, therefore, can be represented by

$$\Pr(A | \omega^+, \omega^-, \theta, \phi) = \prod_{ij} \left(\sum_{rs} \omega_{rs}^+ \theta_{ri} \phi_{sj} \right)^{A_{ij}^+} \left(\sum_{rs} \omega_{rs}^- \theta_{ri} \phi_{sj} \right)^{A_{ij}^-}. \quad (3)$$

Note that the self-loop links are allowed, and the weights A_{ij}^+ and A_{ij}^- , respectively, are viewed as the number of positive and negative multiple links from vertex i to vertex j , as is done in many existing models [23–25].

To infer the missing group memberships \overleftarrow{g} and \overrightarrow{g} , we need to maximize the likelihood in Eq. (3) with respect to the model parameters ω^+ , ω^- , θ , and ϕ . For convenience, one usually works not directly with the likelihood itself but with its logarithm,

$$\mathcal{L} = \ln \Pr(A | \omega^+, \omega^-, \theta, \phi) \\ = \sum_{ij} A_{ij}^+ \ln \left(\sum_{rs} \omega_{rs}^+ \theta_{ri} \phi_{sj} \right) \\ + \sum_{ij} A_{ij}^- \ln \left(\sum_{rs} \omega_{rs}^- \theta_{ri} \phi_{sj} \right). \quad (4)$$

The maximum of the likelihood and its logarithm occur in the same place because the logarithm is a monotonically increasing function.

Considering that the group memberships \overleftarrow{g} and \overrightarrow{g} are unknown, it is intractable to optimize the log-likelihood \mathcal{L} directly again. We can, however, make a good guess of the hidden variables \overleftarrow{g} and \overrightarrow{g} according to the network structure and the model parameters, and seek the maximization of the following expected log-likelihood:

$$\overline{\mathcal{L}} = \sum_{\overleftarrow{g}, \overrightarrow{g}} \Pr(\overleftarrow{g}, \overrightarrow{g} | A^+, \omega^+, \theta, \phi) \ln \Pr(A^+ | \overleftarrow{g}, \overrightarrow{g}, \omega^+, \theta, \phi) + \sum_{\overleftarrow{g}, \overrightarrow{g}} \Pr(\overleftarrow{g}, \overrightarrow{g} | A^-, \omega^-, \theta, \phi) \ln \Pr(A^- | \overleftarrow{g}, \overrightarrow{g}, \omega^-, \theta, \phi) \\ = \sum_{ijrs} \Pr(r, s | e_{ij}^+, \omega^+, \theta, \phi) [A_{ij}^+ (\ln \omega_{rs}^+ + \ln \theta_{ri} + \ln \phi_{sj})] + \sum_{ijrs} \Pr(r, s | e_{ij}^-, \omega^-, \theta, \phi) [A_{ij}^- (\ln \omega_{rs}^- + \ln \theta_{ri} + \ln \phi_{sj})] \\ = \sum_{ijrs} q_{ijrs}^+ A_{ij}^+ (\ln \omega_{rs}^+ + \ln \theta_{ri} + \ln \phi_{sj}) + \sum_{ijrs} q_{ijrs}^- A_{ij}^- (\ln \omega_{rs}^- + \ln \theta_{ri} + \ln \phi_{sj}), \quad (5)$$

where $q_{ijrs}^+ = \Pr(\overleftarrow{g}_{ij} = r, \overrightarrow{g}_{ij} = s | e_{ij}^+, \omega^+, \theta, \phi)$ is the probability that one will find a positive edge e_{ij}^+ with its tail vertex i from group r and its head vertex j from group s given the network and the model parameters. An analogous interpretation can also be made for $q_{ijrs}^- = \Pr(\overleftarrow{g}_{ij} = r, \overrightarrow{g}_{ij} = s | e_{ij}^-, \omega^-, \theta, \phi)$.

With the expected log-likelihood, the best estimate of the value of \mathcal{L} together with the position of its maximum give the most likely values of the model parameters. Finding the maximum still presents a problem, however, since the calculation of q_{ijrs}^+ and q_{ijrs}^- requires the values of ω^+ , ω^- , θ , and ϕ , and vice versa. One possible solution is to adopt an iterative self-consistent approach that evaluates both simultaneously. Like many previous works [11, 23–25], we

utilize the expectation-maximization (EM) algorithm, which first computes the posterior probabilities of hidden variables using estimated model parameters and observed data (the E-step), and then reestimates the model parameters (the M-step).

In the E-step, we calculate the expected probabilities q_{ijrs}^+ and q_{ijrs}^- given the observed network A and parameters ω^+ , ω^- , θ , and ϕ ,

$$q_{ijrs}^+ = \frac{\Pr(\overleftarrow{g}_{ij} = r, \overrightarrow{g}_{ij} = s, e_{ij}^+ | \omega^+, \theta, \phi)}{\Pr(e_{ij}^+ | \omega^+, \theta, \phi)} \\ = \frac{\omega_{rs}^+ \theta_{ri} \phi_{sj}}{\sum_{rs} \omega_{rs}^+ \theta_{ri} \phi_{sj}},$$

$$q_{ijrs}^- = \frac{\Pr(\overleftarrow{g}_{ij} = r, \overrightarrow{g}_{ij} = s, e_{ij}^- | \omega^-, \theta, \phi)}{\Pr(e_{ij}^- | \omega^-, \theta, \phi)} = \frac{\omega_{rs}^- \theta_{ri} \phi_{sj}}{\sum_{rs} \omega_{rs}^- \theta_{ri} \phi_{sj}}. \quad (6)$$

In the M-step, we use the values of q_{ijrs}^+ and q_{ijrs}^- estimated in the E-step to evaluate the expected log-likelihood and to find the values of the parameters that maximize it. Introducing the Lagrange multipliers ρ^+ , ρ^- , γ_r , and λ_s to incorporate the normalization conditions, the expected log-likelihood expression to be maximized becomes

$$\begin{aligned} \tilde{\mathcal{L}} = & \overline{\mathcal{L}} + \rho^+ \left(1 - \sum_{rs} \omega_{rs}^+\right) + \rho^- \left(1 - \sum_{rs} \omega_{rs}^-\right) \\ & + \sum_r \gamma_r \left(1 - \sum_i \theta_{ri}\right) + \sum_s \lambda_s \left(1 - \sum_j \phi_{sj}\right). \end{aligned} \quad (7)$$

By letting the derivative of $\tilde{\mathcal{L}}$ be 0, the maximum of the expected log-likelihood appears at the places where

$$\begin{aligned} \omega_{rs}^+ &= \frac{\sum_{ij} A_{ij}^+ q_{ijrs}^+}{\sum_{ijrs} A_{ij}^+ q_{ijrs}^+}, \\ \omega_{rs}^- &= \frac{\sum_{ij} A_{ij}^- q_{ijrs}^-}{\sum_{ijrs} A_{ij}^- q_{ijrs}^-}, \\ \theta_{ri} &= \frac{\sum_{js} A_{ij}^+ q_{ijrs}^+ + \sum_{js} A_{ij}^- q_{ijrs}^-}{\sum_{ijs} A_{ij}^+ q_{ijrs}^+ + \sum_{ijs} A_{ij}^- q_{ijrs}^-}, \\ \phi_{sj} &= \frac{\sum_{ir} A_{ij}^+ q_{ijrs}^+ + \sum_{ir} A_{ij}^- q_{ijrs}^-}{\sum_{ijr} A_{ij}^+ q_{ijrs}^+ + \sum_{ijr} A_{ij}^- q_{ijrs}^-}. \end{aligned} \quad (8)$$

Equations (6) and (8) constitute our EM algorithm for exploratory analysis of signed networks. When the algorithm converges, we obtain a set of values for hidden quantities q_{ijrs}^+ , q_{ijrs}^- and model parameters ω^+ , ω^- , θ , and ϕ .

It is worthwhile to note that the EM algorithm is known to converge to local maxima of the likelihood but not always to global maxima. With different starting values, the algorithm may give rise to different solutions. To obtain a satisfactory solution, we perform several runs with different initial conditions and return the solution giving the highest log-likelihood over all the runs.

Now we consider the computational complexity of the EM algorithm. For each iteration, the cost consists of two parts. The first part is from the calculation of q_{ijrs}^+ and q_{ijrs}^- using Eq. (6), whose time complexity is $O(mc^2)$. Here m is the edges in the network and c is the number of groups. The second part is from the estimation of the model parameters using Eq. (8), whose time complexity is also $O(mc^2)$. We use T to denote the number of iterations before the iteration process converges. Then, the total cost of the EM algorithm for our model is $O(Tmc^2)$. It is difficult to give a theoretical estimation for the number T of iterations. Generally speaking, T is determined by the network structure and the initial condition.

B. Soft partition and overlapping structures

The parameters, obtained by fitting the model to the observed network structure with the EM algorithm, provide us with useful information for the mesoscopic structure in a given network. Specifically, the matrices ω^+ and ω^- , in analogy with the image graph in the role model [41], characterize the connecting patterns among different groups, which determine the type of structural patterns. Furthermore, θ and ϕ indicate the centrality of a vertex in its groups from the perspective of outgoing edges and incoming edges, respectively. Consequently, the probability of vertex i drawn from group r when it is the tail of edges can be defined as

$$\alpha_{ir} = \frac{\sum_s (\omega_{rs}^+ + \omega_{rs}^-) \theta_{ri}}{\sum_{rs} (\omega_{rs}^+ + \omega_{rs}^-) \theta_{ri}}, \quad (9)$$

and vertex i can be simply assigned to the group r^* to which it most likely belongs, i.e., $r^* = \arg \max_r \{\alpha_{ir}, r = 1, 2, \dots, c\}$. The result yields a hard partition of the signed network.

In fact, the set of scalars $\{\alpha_{ir}\}_{r=1}^c$ supply us with the probabilities that vertex i belongs to different groups, which can be referred to as soft or fuzzy memberships. Assigning vertices to more than one group has attracted by far the most interest, particularly in overlapping community detection [8, 16–18]. The vertices belonging to several groups are found to take a special role in networks, e.g., signal transduction in biological networks. Furthermore, some vertices, considered as “instable” [16], are located on the border between two groups and thus are difficult to classify into any group. It is of great importance to reveal the global organization of a signed network in terms of overlapping mesoscopic structures and to find the instable vertices. We employ here bridgeness [17] and group entropy [20] to capture the vertices’ instabilities and to extract the overlapping mesoscopic structure. These two measures of vertex i are computed as

$$b_i = 1 - \sqrt{\frac{c}{c-1} \sum_{r=1}^c \left(\alpha_{ir} - \frac{1}{c}\right)^2}, \quad (10)$$

$$\xi_i = - \sum_{r=1}^c \alpha_{ir} \log_c \alpha_{ir}. \quad (11)$$

Note that vertex i has a large bridgeness b_i and entropy ξ_i when it most likely participates in more than one group simultaneously, and vice versa. From the perspective of incoming edges, we can represent the probability of vertex j belonging to group s by

$$\beta_{js} = \frac{\sum_r (\omega_{rs}^+ + \omega_{rs}^-) \phi_{sj}}{\sum_{rs} (\omega_{rs}^+ + \omega_{rs}^-) \phi_{sj}}. \quad (12)$$

These statements for α_{ir} also apply to β_{js} , so we do not need to repeat them.

The model described above focuses on directed networks. Actually, the model could be easily generalized to undirected networks by letting the parameter θ be identical to ϕ . The derivation follows the case of directed networks, and the results are the same as Eqs. (6) and (8).

C. Model selection

Up to now, our model has assumed that the number of groups c is known *a priori*. This information, however, is unavailable in many cases. It is necessary to provide a criterion to determine an appropriate group number for a given network. Several methods have been proposed to deal with this model selection issue. We adopt the minimum description length (MDL) principle, which was also utilized in the previous generative models for network structure exploration [25].

According to the MDL principle, the required length to describe the network data comprises two components. The first one describes the coding length of the network, which is $-L$ for a directed network and $-L/2$ for an undirected network. The other gives the length for coding model parameters, i.e., $-\sum_{rs} \ln \omega_{rs}^+ - \sum_{rs} \ln \omega_{rs}^- - \sum_{ri} \ln \theta_{ri} - \sum_{sj} \ln \phi_{sj}$ for a directed network and $-\sum_{rs} \ln \omega_{rs}^+ - \sum_{rs} \ln \omega_{rs}^- - \sum_{ri} \ln \theta_{ri}$ for an undirected network. The optimal c is the one that minimizes the total description length.

IV. EXPERIMENTAL RESULTS

In this section, we extensively test our SSBM model on a series of synthetic signed networks with various known structures, including community structure and disassortative structure. After that, the method is also applied to three real-life social networks.

A. Synthetic networks

The ad hoc networks, designed by Girvan and Newman [12], have been broadly used to validate and compare community detection algorithms [14–16,20]. By contrast, there exists no such benchmark for community detection in networks with both positive and negative links. We generate the signed ad hoc networks with controlled community structure using the method developed in Refs. [34,42]. The networks have 128 vertices, which are divided into four groups with 32 vertices each. Edges are placed randomly such that they are positive within groups and negative between groups, and the average degree of a vertex is 16. The community structure is controlled by three parameters, p_{in} indicating the probability of each vertex connecting to other vertices in the same group, p_+ the probability of positive links appearing between groups, and p_- the probability of negative links arising within groups. Thus, the parameter p_{in} regulates the cohesiveness of the communities, and the remaining parameters p_+ and p_- add noise to the community structure when p_{in} is fixed.

For the synthetic networks, we simply consider their hard partition as defined in Sec. III B. The results are evaluated by the normalized mutual information (NMI) [43], which can be formulated as

$$\text{NMI}(C_1, C_2) = \frac{\sum_{i=1}^c \sum_{j=1}^c n_{ij} \ln \frac{n_{ij} n}{n_i^{(1)} n_j^{(2)}}}{\sqrt{(\sum_{i=1}^c n_i^{(1)} \ln \frac{n_i^{(1)}}{n})(\sum_{i=1}^c n_i^{(2)} \ln \frac{n_i^{(2)}}{n})}},$$

where C_1 and C_2 are the true group assignment and the assignment found by the algorithms, respectively, n is the number of vertices, n_{ij} is the number of vertices in the known group i that are assigned to the inferred group j , $n_i^{(1)}$ is the

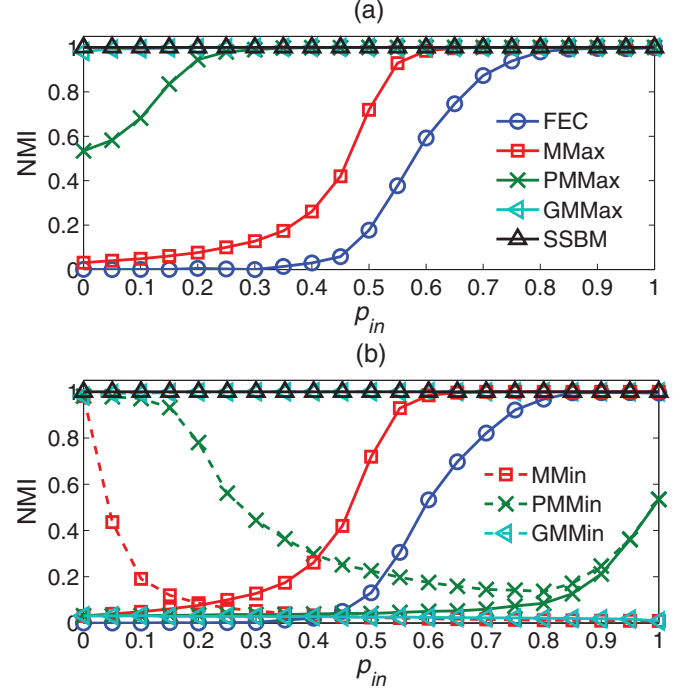


FIG. 2. (Color online) NMI of our method and other approaches on balanced ad hoc networks with controlled community structure (a) and disassortative structure (b). Each point is an average over 50 realizations of the networks.

number of vertices in the true group i , and $n_j^{(2)}$ is the number of vertices in the inferred group j . The larger the NMI value, the better the partition obtained by the algorithms.

We conduct two different experiments. First, we set the two parameters p_+ and p_- to be zero and gradually change p_{in} from 1 to 0. In this situation, all the generated synthetic networks are 4-balanced. Figure 2(a) reports the experimental results obtained by our method and two state-of-the-art approaches, namely generalized modularity maximization through simulated annealing (denoted by GMMMax) [35,36] and the finding and extracting community (FEC) method [34]. In addition, we also implement the simulated annealing algorithm to maximize the standard modularity by ignoring the sign of the links (denoted by MMax) and removing the negative edges (denoted by PMMax), respectively. Each point in the curves is an average over 50 realizations of the synthetic random networks. Bear in mind that the community structure becomes less cohesive as the parameter p_{in} decreases from 1 to 0. We can see that both the SSBM model and the GMMMax method perform fairly well and are almost able to perfectly recover the communities in the synthetic networks for all cases. When $0 \leq p_{in} \leq 0.1$, our model is even slightly superior to the GMMMax approach. The remaining three methods, however, can only achieve promising results when p_{in} is sufficiently large. They all show a fast deterioration as p_{in} becomes smaller and smaller. For example, the NMI of the FEC algorithm begins to drop once p_{in} exceeds 0.8, and then quickly reduces to less than 0.2 when $p_{in} = 0.5$ and even to approximately 0 when p_{in} is smaller than 0.3. Similar performances can be observed for the MMax and PMMax approaches as well. These results are quite understandable since both the SSBM model

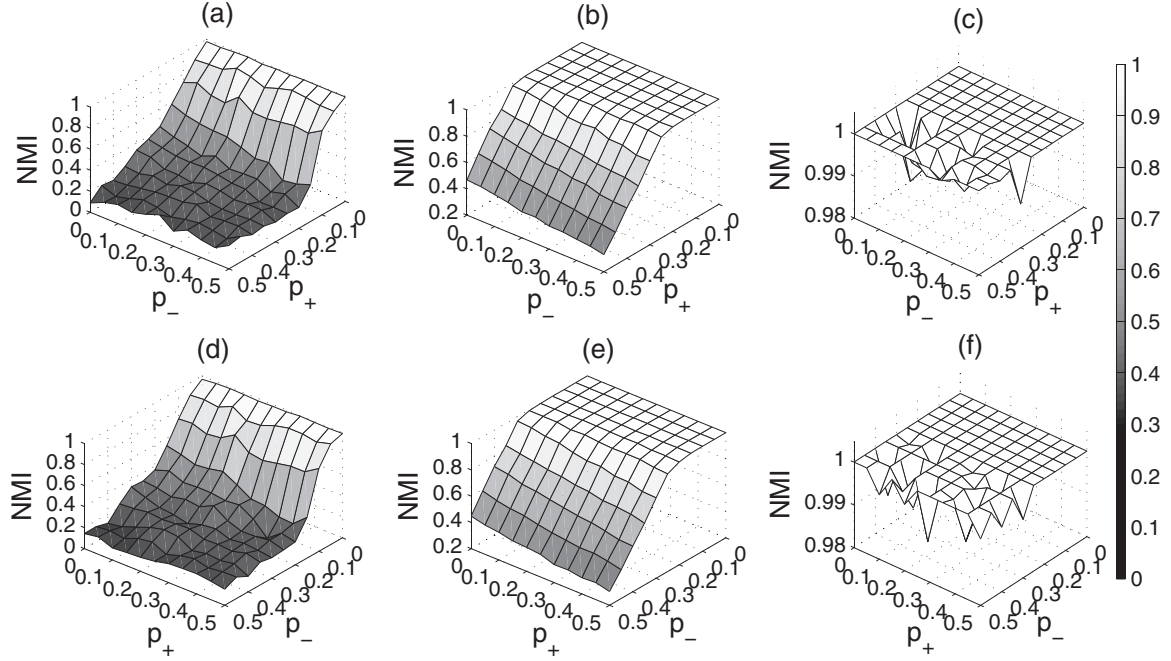


FIG. 3. NMI on unbalanced ad hoc networks with controlled community structure (first row) for (a) FEC, (b) GMMax, and (c) SSBM, and with controlled disassortative structure (second row) for (d) FEC, (e) GMMin, and (f) SSBM. Each point is an average over 50 realizations of the networks.

and the GMMax method consider the contribution made by the negative links in signed networks, which is either neglected or removed in the remaining three approaches. This highlights the importance of the negative edges for community detection in the signed networks. Moreover, the PMMax method always outshines the MMax method, especially when p_{in} in the range $0 \leq p_{in} \leq 0.5$, which is in agreement with the results reported in Ref. [42], indicating that the positive links in signed networks have a significant impact on community detection.

Then, we fix the parameter $p_{in} = 0.8$ and gradually change the other two parameters p_+ and p_- from 0 to 0.5, respectively. Clearly, all the synthetic networks are not balanced in this setting. The results obtained by our model and two updated algorithms are give in the upper row of Fig. 3. As we can see, the SSBM model consistently, and sometimes significantly, outperforms the other two approaches. More specifically, its NMI is always 1 except for a few negligible perturbations. By contrast, the FEC algorithm cannot offer a satisfactory partition of the signed networks when $0 \leq p_+ \leq 0.3$ and $0 \leq p_- \leq 0.5$, whose NMI is less than 0.4 at all times. When $0.3 \leq p_+ \leq 0.5$ and $0 \leq p_- \leq 0.5$, the GMMax approach exhibits a competitive performance, but its NMI suddenly collapses and continuously decreases once p_+ is larger than 0.3.

We turn now to the second experiment in which the synthetic networks have controlled disassortative structure. The signed networks are generated in the same way, except that we randomly place negative links within groups and positive links between groups. Similarly, the disassortative structure in these networks is controlled by three parameters again. p_{in} indicates the probability of each vertex connecting to other vertices in the same group, p_+ is the probability of positive links appearing within groups, and p_- is the probability of negative links arising between groups.

We first study the balanced networks by setting p_+ and p_- to zero and changing p_{in} from 1 to 0 once again. As shown in Fig. 2(b), the FEC algorithm, the MMax method, and our model achieve performances that are very similar to those in the first experiment. That is, our model always successfully finds the clusters in the synthetic networks in all cases, while the FEC algorithm and the MMax method perform fairly well when p_{in} is large enough, but quickly degrade as p_{in} approaches 0. The PMMax and the GMMax methods, however, perform rather badly. The NMI of the PMMax method seems to be no greater than 0.5 even if $p_{in} = 1$, while the NMI of the GMMax approach nearly vanishes in all cases. This is because the two methods, which seek standard and generalized modularity maximization, respectively, are suitable only for community detection. As a consequence, they deserve to suffer a serious failure in this experiment. Instead, one should minimize the modularity to uncover the multipartite structure in networks, as indicated in Ref. [10]. Therefore, we apply the simulated annealing algorithm to minimize the generalized modularity (denoted by GMMin) and the standard modularity by ignoring the sign of the links (denoted by MMin) and excluding the negative connections (denoted by PMMin), respectively. We see from Fig. 2(b) that the GMMin method can compete with our SSBM model except for a slight inferiority when $0 \leq p_{in} \leq 0.1$. However, the MMin and PMMin approaches perform unsatisfactorily due to the fact that they do not consider the contributions derived from the negative links.

We investigate next the disassortative structure in unbalanced synthetic networks by fixing $p_{in} = 0.8$ and changing p_+ and p_- from 0 to 0.5 step by step. The lower row of Fig. 3 gives the results obtained by the FEC method, the GMMin approach, and our SSBM model, which are quite similar to

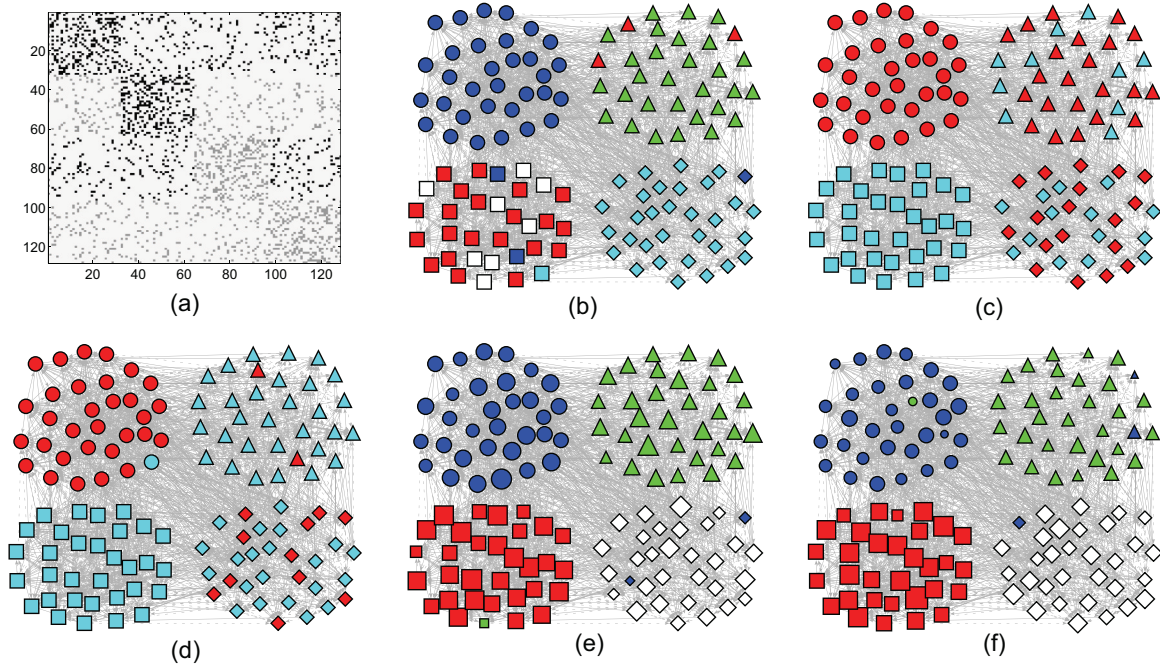


FIG. 4. (Color online) Detecting the mesoscopic structure of a synthetic network. (a) The adjacency matrix of the signed network where the black dots denote the positive links and the gray dots represent the negative edges. The partitioning results for different methods (b) EFC, (c) GMMax, (d) GMMin and SSBM from the perspective of outgoing edges (e) and incoming edges (f), where the solid edges denote the positive links and the dashed edges represent negative links. The sizes of the vertices in (e) and (f) indicate their centrality degree in the corresponding groups according to the parameters θ and ϕ , respectively.

those in the first experiment. In particular, although the SSBM does not perform perfectly in some cases, its NMF is still rather high, say, more than 0.98. When $0 \leq p_- \leq 0.3$, the GMMin approach yields sufficiently good results, but its NMF reduces at a very fast speed along with p_- toward 0.5. The FEC algorithm achieves the worst performance in all cases.

Finally, we focus on a synthetic network containing a multitude of mesoscopic structures, whose adjacency matrix is given in Fig. 4(a). Intuitively, according to the outgoing edges in this network, the second group is the community structure and the third group belongs to the disassortative structure. The first group, with positive outgoing links only, can be viewed as an example of the standard community structure in positive networks, while the last group, which includes only negative outgoing links, can be referred to as an extreme example of disassortative structure in signed networks. Meanwhile, from the perspective of incoming edges, the four groups exhibit different types of structural patterns, which cannot be categorized simply as community structure or disassortative structure. We apply the FEC algorithm, the GMMax method, the GMMin method, and our model to this signed network. Limited by their intrinsic assumptions, the FEC algorithm, the GMMax method, and the GMMin method fail to uncover the structural patterns, as shown in Figs. 4(b)–4(d). In particular, the generalized modularity proposed in Refs. [35,36], regardless of whether it is maximum or minimum, misleads us into receiving an improper partition of the network in which the four groups merge with each other. But by dividing vertices with the same connection profiles into groups, our model could accurately detect all types of mesoscopic structures, both from the perspective of outgoing

links [Fig. 4(e)] and incoming edges [Fig. 4(f)]. Furthermore, the obtained parameters θ and ϕ reveal the centrality of each vertex in its corresponding group from the two perspectives.

B. Real-life networks

We test our method further by applying it to several real networks containing both positive and negative links. The first network is a relation graph of 10 parties of the Slovene Parliamentary in 1994 [44]. The weights of links in the network were estimated via 72 questionnaires among 90 members of the Slovene National Parliament. The questionnaires were designed to estimate the distance between the ten parties on a scale from -3 to 3 , and the final weights were the averaged values multiplied by 100.

Applying our model to this signed network, we find that the MDL achieves its minima when $c = 2$, as shown in Fig. 5(a), indicating that there are exactly two communities in the network. Figure 6(a) gives the partition obtained by our method, which divides the network into two groups of equal size and produces a completely consistent split with the true communities in the network. As expected, vertices within the same community are mostly connected by positive links, while vertices from different communities are mainly connected by negative links. We shade each vertex proportional to the parameters $\{\alpha_{ir}\}_{r=1}^c$, the magnitude of which supplies us with the probabilities of each vertex belonging to different groups.¹

¹Both this network and the Gahuku-Gama Subtribes network are undirected graphs, and therefore the parameter α is identical to β , and θ is identical to ϕ .

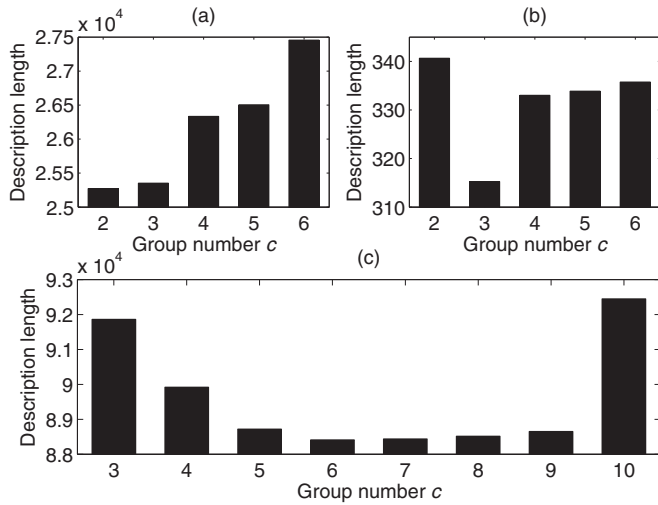


FIG. 5. Model selection results for (a) the Slovene Parliamentary network, (b) the Gahuku-Gama Subtribes network and (c) the international conflict and alliance network.

From Table I, we see that all the vertices can be exclusively separated into two communities, except for the vertex “SNS,” which belongs to the circle group with probability 0.0186 and to the square group with probability 0.9814. In other words, the two communities overlap with each other at this vertex, resulting in its high bridgeness of 0.0372 and group entropy of 0.1334. This is validated by the observation that the vertex has two negative links with vertices “ZS-ESS” and “DS” in the same community. We also visualize the learned parameters ω^+ and ω^- in Fig. 6(b), which indeed provide a coarse-grained description of the signed network and reveal that this network actually has two communities.

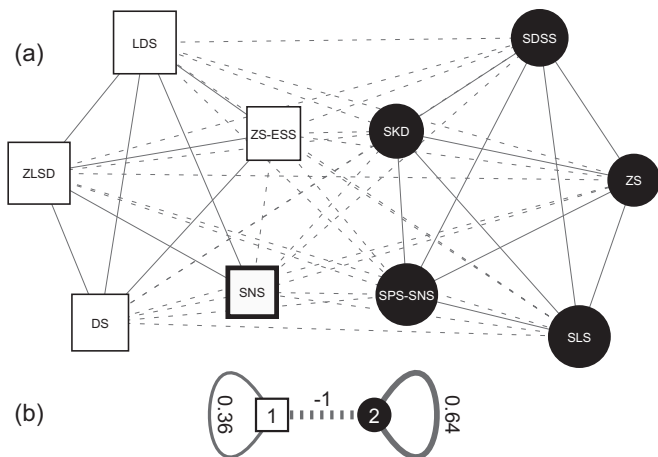


FIG. 6. Exploratory analysis of the Slovene Parliamentary network [44]. The solid edges denote positive links and the dashed edges represent negative links. The true community structure in this network is represented by two different shapes, namely a circle and a square. The shades of nodes indicate the membership α obtained by fitting our model to this network. The sizes of the vertices, proportional to θ , indicate their centrality degree with respect to their corresponding group.

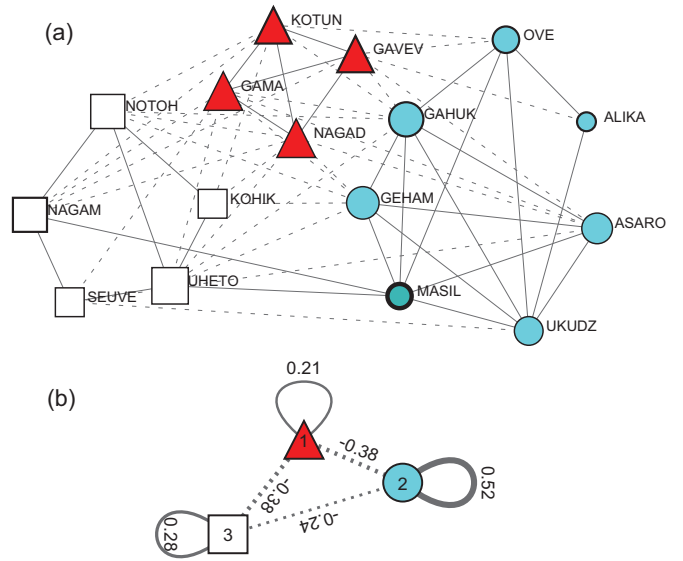


FIG. 7. (Color online) Exploratory analysis of the Gahuku-Gama Subtribes network [45]. The solid edges denote positive links and the dashed edges represent negative links. The true community structure in this network is represented by three different shapes, while the inferred groups are denoted by different colors. The sizes of the vertices are proportional to the parameters θ .

The second network is the Gahuku-Gama Subtribes network, which was created based on Read’s study on the cultures of the Eastern Central Highlands of New Guinea [45]. This network describes the political alliance and enmities among the 16 Gahuku-Gama subtribes, which were distributed in a particular area and were engaged in warfare with one another in 1954. The positive and negative links of the network correspond to political arrangements with positive and negative ties, respectively. Figure 5(b) tells us that this signed network consists of three groups, because the MDL of the SSBM model is at a minimum when $c = 3$. The three groups categorized by our model are given in Fig. 7(a), and they match perfectly with the true communities in the signed network. As shown in Table II, the vertex “MASIL” participates in the circle group with probability 0.7143 and in the square group with probability 0.2857. As a result, it has a large value of bridgeness 0.3773 and group entropy 0.5446. This implies that these two groups overlap with each other at this vertex, which is approved by the fact that the vertex “MASIL” has two positive links connected to “NAGAM” and “UHETO,” respectively. The learned parameters ω^+ and ω^- supply us with a thumbnail of the signed network again in Fig. 7(b).

Finally, we test our model on the network of international relations taken from the Correlates of War data set over the period 1993–2001 [36]. In this network, positive links represent military alliances and negative links denote military disputes. The disputes are associated with three hostility levels, from “no militarized action” to “interstate war.” For each pair of countries, we chose the mean level of hostility between them over the given time interval as the weight of their negative link. The positive links denote the alliances: 1 for entente, 2 for nonaggression pact, and 3 for defense

TABLE I. The soft group membership α , bridgeness b_i [17], and group entropy ξ_i [20] of each vertex in the Slovene Parliamentary network [42]. Larger bridgeness or entropy means that the corresponding nodes are more “instable.”

| Vertex | SKD | ZLSD | SDSS | LDS | ZS-ESS | ZS | DS | SLS | SPS-SNS | SNS |
|---------------|-------|-------|-------|-------|--------|-------|-------|-------|---------|--------|
| α_{i1} | 1.000 | 0 | 1.000 | 0 | 0 | 1.000 | 0 | 1.000 | 1.000 | 0.0186 |
| α_{i2} | 0 | 1.000 | 0 | 1.000 | 1.000 | 0 | 1.000 | 0 | 0 | 0.9814 |
| b_i | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.0372 |
| ξ_i | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0.1334 |

pact. Finally, we normalized both the negative and positive links into the interval $[0, 1]$, and the final weight of the link among each pair of countries is the remainder of the weight of the normalized positive links minus the weight of the normalized negative links. The obtained network contains a giant component consisting of 161 vertices (countries) and 2517 links (conflicts or alliances). Here, we only investigate the structure of the giant component.

The structure of this network was investigated in several existing studies. These studies indicated that there are six main power blocs, each consisting of a set of countries with similar actions of alliances or disputes. In Ref. [36], the authors labeled these power blocs as (i) the West, (ii) Latin America, (iii) the Muslim World, (iv) Asia, (v) West Africa, and (vi) Central Africa. Applying the SSBM model to this network, we find that the MDL arrives at its minimum when $c = 6$, as illustrated in Fig. 5(c). By partitioning the network into six groups, we summarize the results in Fig. 8. From the rearranged adjacency matrix [Fig. 8(c)], we can conclude that the first, second, third, and fifth groups, from bottom left to top right, distinctly belong to the community structure, while the sixth group can be viewed as the disassortative structure. However, the fourth group cannot be simply categorized as either community structure or disassortative structure. In agreement with the assumption of the SSBM model, vertices in the six groups exhibit similar connection profiles, although miscellaneous structural patterns coexist in this network.

From the perspective of the outgoing edges, we obtain a split of the network that is similar to the one obtained in Ref. [36], as shown in Fig. 8(a). However, several notable differences exist between the two results. Specifically, “Pakistan” is grouped with the West and “South Korea” is grouped with the Muslim World in Ref. [36]. These false categorizations

can be correctly amended, which is consistent with the configuration depicted in Huntington’s renowned book *The Clash of Civilizations* [46]. In addition, we categorized “Australia,” which is grouped with the West in Ref. [36], into the group Asia for understandable reasons. Figure 8(b) gives a quite different structure of this network from the perspective of incoming edges. Three groups, namely the West, Latin America, and the Muslim World, stay almost the same. But “Russia,” together with some countries of the former Soviet Union, are isolated from the Asia group and form another independent power bloc. Meanwhile, the remaining countries in the Asia group join with the West Africa countries to constitute a bigger cluster. It is not difficult to see that all the changes appear to be in accordance with the history and evolution of the international relations.

Recall that the parameters θ and ϕ provide us with the centrality degrees of each vertex in its corresponding group from the perspective of outgoing edges and incoming edges, respectively. In other words, the parameters measure the importance of each vertex in its group. For a better visualization, the sizes of vertices in Figs. 8(d) and 8(e) are proportional to the magnitude of the scalars θ and ϕ . Coincidentally, we discover that the big vertices, marked by the red bold border, usually stand for the dominant countries in their corresponding groups. For example, the largest vertex of the West is “USA” in Fig. 8(d). In fact, this state often serves as a leader in its power bloc. A similar interpretation can be given for the vertex “Russia” in the Asia group. We further check the bridgeness and group entropy for each vertex in the network (data not shown), and we mark the vertices, which have large values of these two measures, with the black bold border. As anticipated, these kinds of vertices are particularly prone to residing on the boundaries of different groups. That is to say,

TABLE II. The soft group membership α , bridgeness b_i [17] and group entropy ξ_i [20] of each vertex in the Gahuku-Gama Subtribes network [45]. Larger bridgeness or entropy means that the corresponding nodes are more “instable.”

| Vertex | GAVEV | KOTUN | OVE | ALIKA | NAGAM | GAHUK | MASIL | UKUDZ | NOTOH | KOHIK |
|---------------|-------|-------|-------|-------|-------|-------|--------|-------|-------|-------|
| α_{i1} | 1.000 | 1.000 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| α_{i2} | 0 | 0 | 1.000 | 1.000 | 0 | 1.000 | 0.7143 | 1.000 | 0 | 0 |
| α_{i3} | 0 | 0 | 0 | 0 | 1.000 | 0 | 0.2857 | 0 | 1.000 | 1.000 |
| b_i | 0 | 0 | 0 | 0 | 0 | 0 | 0.3773 | 0 | 0 | 0 |
| ξ_i | 0 | 0 | 0 | 0 | 0 | 0 | 0.5446 | 0 | 0 | 0 |
| Vertex | GEHAM | ASARO | UHETO | SEUVE | NAGAD | GAMA | | | | |
| α_{i1} | 0 | 0 | 0 | 0 | 1.000 | 1.000 | | | | |
| α_{i2} | 1.000 | 1.000 | 0 | 0 | 0 | 0 | | | | |
| α_{i3} | 0 | 0 | 1.000 | 1.000 | 0 | 0 | | | | |
| b_i | 0 | 0 | 0 | 0 | 0 | 0 | | | | |
| ξ_i | 0 | 0 | 0 | 0 | 0 | 0 | | | | |

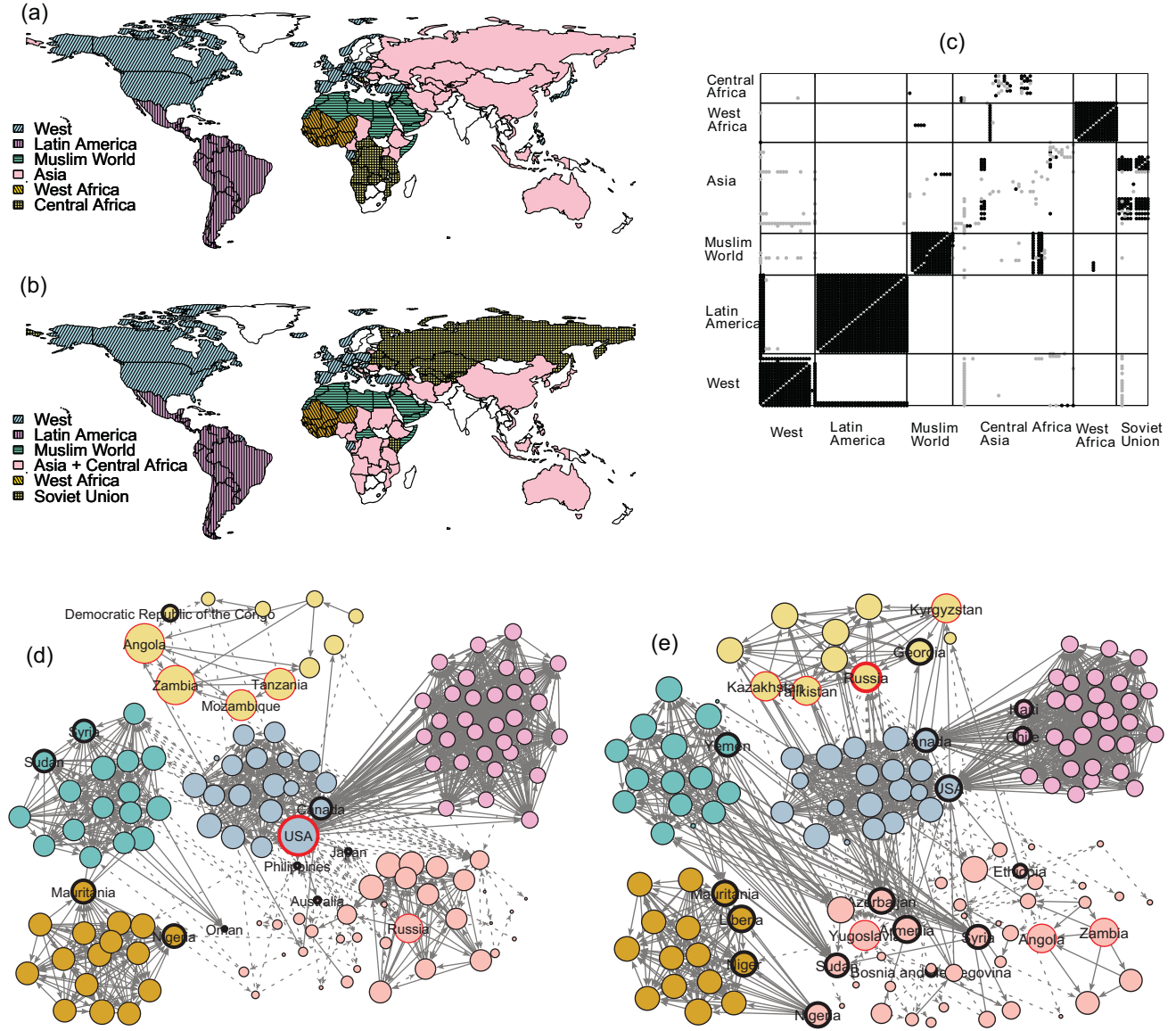


FIG. 8. (Color online) Exploratory analysis of the international conflict and alliance network [36]. Maps of the groups found using the SSBM model from the perspective of outgoing edges (a) and incoming edges (b). (c) The rearranged adjacency matrix, in which the black dots denote positive links and the gray dots represent negative edges. Six groups are separated by black solid lines. The partition of this network is obtained by the SSBM model from the perspective of outgoing edges (d) and incoming edges (e), where the solid edges denote positive links and the dashed edges represent negative links. The sizes of the vertices are proportional to their centrality degrees θ and ϕ , respectively. The red bold border vertices have large centrality degrees, while the black bold border vertices have large values of bridgeness and group entropy.

the vertices that are very difficult to divide into one group build a fuzzy watershed of the overlapping structures. In Fig. 8(d), the three vertices “Japan,” “Philippines,” and “Australia,” with high values of both bridgeness and group entropy, play a transitional role between the West and Asia groups. In reality, the above-mentioned Asian countries frequently collaborated with their counterparts in the West group in many areas, from economics to the military.

V. CONCLUSIONS

We propose an extension of the stochastic block model to study the mesoscopic structural patterns in signed networks. Without prior knowledge of what specific structure exists, our

model can not only accurately detect broad types of intrinsic structures, but it can also directly learn their types from the network data. Experiments on a number of synthetic and real-world networks demonstrate that our model outperforms the state-of-the-art approaches at extracting various structural features in a given network. Due to the flexibility inherited from the stochastic model, our method is an effective way to reveal the global organization of the networks in terms of the structural regularities, which further helps us to understand the relationship between the structure and function of networks. As future work, we will generalize our model by removing the requirement that the block matrices must be square matrices, and we will investigate the possible applications of the more flexible models.

ACKNOWLEDGMENTS

The author would like to thank Vincent A. Traag for providing the international conflict and alliance network used

in this paper. The author is also grateful to the anonymous reviewers for their valuable suggestions, which were very helpful in improving the manuscript.

-
- [1] M. E. J. Newman, *SIAM Rev.* **45**, 167 (2003).
 - [2] M. A. Porter, J.-P. Onnela, and P. J. Mucha, *Not. Am. Math. Soc.* **56**, 1082 (2009).
 - [3] S. Fortunato, *Phys. Rep.* **486**, 75 (2010).
 - [4] G. Palla, A.-L. Barabási, and T. Vicsek, *Nature (London)* **446**, 664 (2007).
 - [5] A. L. Traud, E. D. Kelsic, P. J. Mucha, and M. A. Porter, *SIAM Rev.* **53**, 526 (2011).
 - [6] G. W. Flake, S. R. Lawrence, C. L. Giles, and F. M. Coetzee, *IEEE Comp.* **35**, 66 (2002).
 - [7] R. Guimerà and L. A. N. Amaral, *Nature (London)* **433**, 895 (2005).
 - [8] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, *Nature (London)* **435**, 814 (2005).
 - [9] M. Huss and P. Holme, *IET Syst. Biol.* **1**, 280 (2007).
 - [10] M. E. J. Newman, *Phys. Rev. E* **74**, 036104 (2006).
 - [11] M. E. J. Newman and E. A. Leicht, *Proc. Natl. Acad. Sci. (USA)* **104**, 9564 (2007).
 - [12] M. Girvan and M. E. J. Newman, *Proc. Natl. Acad. Sci. (USA)* **99**, 7821 (2002).
 - [13] P. Holme, F. Liljeros, C. R. Edling, and B. J. Kim, *Phys. Rev. E* **68**, 056107 (2003).
 - [14] M. E. J. Newman and M. Girvan, *Phys. Rev. E* **69**, 026113 (2004).
 - [15] L. Danon, J. Duch, A. Diaz-Guilera, and A. Arenas, *J. Stat. Mech.* (2005) P09008.
 - [16] D. Gfeller, J. C. Chappelier, and P. De Los Rios, *Phys. Rev. E* **72**, 056135 (2005).
 - [17] T. Nepusz, A. Petróczy, L. Negyessy, and F. Bazsó, *Phys. Rev. E* **77**, 016107 (2008).
 - [18] I. Psorakis, S. Roberts, M. Ebdon, and B. Sheldon, *Phys. Rev. E* **83**, 066114 (2011).
 - [19] C. Granell, S. Gómez, and A. Arenas, *Chaos* **21**, 016102 (2011).
 - [20] J. Q. Jiang and L. J. McQuay, *Physica A* **391**, 854 (2012).
 - [21] P. W. Holland, K. B. Laskey, and S. Leinhardt, *Social Netw.* **5**, 109 (1983).
 - [22] B. Karrer and M. E. J. Newman, *Phys. Rev. E* **83**, 016107 (2011).
 - [23] T. Yang, Y. Chi, S. Zhu, Y. Gong, and R. Jin, *Mach. Learn.* **82**, 157 (2011).
 - [24] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing, *J. Mach. Learn. Res.* **9**, 1981 (2008).
 - [25] H. W. Shen, X. Q. Cheng, and J. F. Guo, *Phys. Rev. E* **84**, 056111 (2011).
 - [26] T. P. Peixoto, *Phys. Rev. X* **4**, 011047 (2014).
 - [27] A. Decelle, F. Krzakala, and L. Zdeborova, *Phys. Rev. Lett.* **107**, 065701 (2011).
 - [28] A. Clauset, C. Moore, and M. E. J. Newman, *Nature (London)* **453**, 98 (2008).
 - [29] R. Guimerà and M. Sales-Pardo, *Proc. Natl. Acad. Sci. (USA)* **106**, 22073 (2009).
 - [30] F. Lorrain and H. C. White, *J. Math. Sociol.* **1**, 49 (1971).
 - [31] J. Reichardt and D. R. White, *Eur. Phys. J. B* **60**, 217 (2007).
 - [32] M. J. Mason, G. Fan, K. Plath, Q. Zhou, and S. Horvath, *BMC Genomics* **10**, 327 (2009).
 - [33] G. Facchetti, G. Iacono, and C. Altafini, *Phys. Rev. E* **86**, 036116 (2012).
 - [34] B. Yang, W. K. Cheung, and J. M. Liu, *IEEE Trans. Knowl. Data Eng.* **19**, 1333 (2007).
 - [35] S. Gómez, P. Jensen, and A. Arenas, *Phys. Rev. E* **80**, 016114 (2009).
 - [36] V. A. Traag and J. Bruggeman, *Phys. Rev. E* **80**, 036115 (2009).
 - [37] M. Szell, R. Lambiotte, and S. Thurner, *Proc. Natl. Acad. Sci. (USA)* **107**, 13636 (2010).
 - [38] F. Heider, *J. Psychol.* **21**, 107 (1946).
 - [39] J. A. Davis, *Hum. Relat.* **20**, 181 (1967).
 - [40] F. Harary and D. Cartwright, *Elem. Math.* **23**, 85 (1968).
 - [41] J. Reichardt and S. Bornholdt, *Phys. Rev. E* **74**, 016110 (2006).
 - [42] T. D. Kaplan and S. Forrest, *arXiv:0801.3290*.
 - [43] A. Strehl and J. Ghosh, *J. Mach. Learn. Res.* **3**, 583 (2002).
 - [44] S. Kropivnik and A. Mrvar, in *Developments Statistics and Methodology*, edited by A. Ferligoj and A. Kramberger (Metodološki zvezki 12, FDV, Ljubljana, 1996), pp. 209–216.
 - [45] K. E. Read, *Southwest. J. Anthropol.* **10**, 1 (1954).
 - [46] S. P. Huntington, *The Clash of Civilizations and the Remaking of World Order* (Simon & Schuster, New York, 1996).