Overlapping community detection in complex networks using symmetric binary matrix factorization

Zhong-Yuan Zhang*

School of Statistics, Central University of Finance and Economics, Haidian District, Beijing 100081, China

Yong Wang

National Center for Mathematics and Interdisciplinary Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China

Yong-Yeol Ahn[†]

School of Informatics and Computing, Indiana University, Bloomington, Indiana 47408, USA (Received 21 March 2013; published 12 June 2013)

Discovering overlapping community structures is a crucial step to understanding the structure and dynamics of many networks. In this paper we develop a symmetric binary matrix factorization model to identify overlapping communities. Our model allows us not only to assign community memberships explicitly to nodes, but also to distinguish outliers from overlapping nodes. In addition, we propose a modified partition density to evaluate the quality of community structures. We use this to determine the most appropriate number of communities. We evaluate our methods using both synthetic benchmarks and real-world networks, demonstrating the effectiveness of our approach.

DOI: 10.1103/PhysRevE.87.062803

PACS number(s): 89.75.Hc

I. INTRODUCTION

Many complex networks possess community structure [1]. Intuitively speaking, a community is a set of nodes that are densely interconnected but loosely connected with the rest of the network. Nodes often belong to more than one community, causing communities to overlap with each other [2,3]. For example, a person in a social network belongs to many groups, such as a family, a company, and a baseball club; many proteins have more than one function and belong to multiple functional communities [2]. The overlap cannot be detected by the traditional hard-partitioning methods that assign each node to a single community [4].

Because of its abundance, overlapping community structure has become a popular research topic and many detection methods have been introduced. They can be classified into two categories based on their outputs [5]: fuzzy overlapping community detection and nonfuzzy overlapping community detection. Fuzzy overlapping community detection methods estimate the strength of memberships, while not being able to provide clear node membership to each community [6–8]. By contrast, nonfuzzy overlapping community detection methods give crisp partitions, allowing each node to have multiple community labels [3,9]. These methods do not provide any information about the strength of the nodes' membership in each community. In short, each approach has complementary benefits and drawbacks, raising the following question: Can we combine the advantages of both kinds of approaches?

In this paper we propose a symmetric binary matrix factorization (SBMF) model to combine the best of both. It is motivated by both symmetric non-negative matrix fac-

(BMF) [12,13]. The SBMF enables us not only to identify the community structures explicitly, but also to analyze the strength of membership based on the corresponding results from SNMF, providing a comprehensive picture exploiting both SNMF and SBMF. Furthermore, the model can distinguish the outliers, which do not belong to any communities, from the overlapping ones.

torization (SNMF) [10,11] and binary matrix factorization

An essential step in community detection is defining the quality of the detected community structure. One of the most widely used quantities is the modularity function Q [14,15]. However, the modularity function has several drawbacks. The most well-known problem is the resolution limit such that Q cannot capture small communities however strongly clustered they are [16]. Several other issues of Qhave been identified including bottleneck dependence [17], misidentification of communities [18], and a degeneracy problem [19]. Recently, *partition density*, a new measure designed for evaluating highly overlapping communities, was proposed [3]. In this paper we use a modified partition density to assess node-based communities instead of link communities.

In summary, the contributions of this paper are (i) proposing a parameter-free, simple-to-implement overlapping community detection method, SBMF; (ii) providing a method to infer an appropriate number of communities using a modified partition density; and (iii) demonstrating the effectiveness of the proposed method by systematically conducting experiments on both the synthetic and the real-world networks.

The rest of the paper is organized as follows. Section II introduces the symmetric binary matrix factorization. Section III presents the partition density. Section IV explains an illustrative example. Section V shows the experimental results. Finally, Sec. VI summarizes.

^{*}zhyuanzh@gmail.com

[†]yyahn@indiana.edu

II. SYMMETRIC BINARY MATRIX FACTORIZATION

A. Motivation

Let us begin from a matrix factorization approach for community detection [6]:

$$\min_{i=1}^{c} \|X - UU^{T}\|_{F}^{2} \quad \text{subject to } U \ge 0,$$

$$\sum_{i=1}^{c} U_{ij} = 1, \quad i = 1, 2, \dots, n,$$
(1)

where X is the adjacency matrix of the network of size $n \times n$ and U is the community membership matrix of size $n \times c$. Note that in this paper we set the diagonal elements of X equal to 1 in accordance with the assumption that the connected pairs of nodes are more similar to each other.

The standard non-negative matrix factorization (NMF) tries to factorize a non-negative matrix X of size $n \times m$ into two nonnegative matrices, U_1 of size $n \times c$ and U_2 of size $m \times c$, such that $X \approx U_1 U_2^T$ [10,11]. For the symmetric objective matrix X, NMF can be reduced to SNMF [see Eq. (1)].

The NMF is becoming one of the most popular and widely accepted models in unsupervised learning [10,20,21]. Several NMF-based models have been applied to community detection. Due to the flexibility of the model, NMF is particularly suitable for the detection of overlapping communities [7,22-25]. The result U can be interpreted as a cluster membership degree matrix, i.e., node i belongs to a community t with the strength U_{it} (note that $\sum_{j} U_{ij} = 1$), providing fuzzy overlapping communities. However, how does one determine whether a node really belongs to a community or not? Does the strength 0.9 mean that the node belongs to the community? What about the strength 0.5? It is often more useful to identify community structure explicitly and consider nodes as full members of their communities [3]. Another issue is that one cannot distinguish the outliers from the overlapping nodes based on the result U. For example, if a node's membership strength vector is [1/c, 1/c, ..., 1/c], one cannot tell whether the node belongs to all of the communities with same strength or does not belong to any of them.

To address these problems, we introduce the SBMF model. The model can explicitly assign community memberships to nodes and can distinguish outliers from overlapping nodes.

B. Model formulation

The SBMF can be defined as follows. Given a symmetric binary matrix A of size $n \times n$, we want to find a binary matrix U of size $n \times c$ such that $A \approx UU^T$. The objective matrix A is the adjacency matrix of the network and U is the community membership indicator matrix: $U_{it} = 1$ if node i is in the community t and 0 if not. If a node i belongs to multiple communities, then the sum of the corresponding row i of U will be larger than one $(\sum_j U_{ij} > 1)$. On the contrary, if the node i is an outlier, the corresponding row i will be zero $(\sum_j U_{ij} = 0)$.

We assume that there is a relatively small number of outliers, which do not belong to any communities in the network, and require that U should have as few zero rows $(\sum_{j} U_{ij} = 0)$ as possible. We achieve this by adding a penalty term into the optimization model. We use 1-norm¹ instead of Frobenius-norm because it gives better numerical results.

In summary, SBMF can be formulated as the following constrained nonlinear programming:

$$\min_{U} \quad \|A - UU^{T}\|_{1} + \sum_{i} \left[1 - \Theta\left(\sum_{j} U_{ij}\right)\right], \quad (2)$$

subject to $U_{ij}^2 - U_{ij} = 0$, i = 1, 2, ..., n; j = 1, 2, ..., c,

¹1-norm of a matrix X is the largest column sum of abs(X), where $abs(X)_{ij} = abs(X_{ij})$ and $abs(\cdot)$ is the absolute value.



FIG. 1. (Color online) Illustrative example, with overlapping nodes 5 and 6, to show how the results of the proposed SBMF model can reveal the community structure in complex networks.

where Θ is the Heaviside step function: For some matrix X,

$$\Theta(X)_{ij} := \begin{cases} 1 & \text{if } X_{ij} > 0 \\ 0 & \text{if } X_{ij} \leqslant 0 \end{cases}$$

As one can see, all the elements of U are variables that need to be determined, which means that it is a large-scale optimization problem. Furthermore, the function Θ is not continuous and the problem is nonsmooth. Hence the standard optimization algorithms are not suitable. To overcome these difficulties, we first initialize U by solving the NMF model (1) and then discretize it by solving the following simpler unconstrained nonlinear programming, instead of (2), to get an optimal approximation solution

$$\min_{u} \quad \|A - \Theta(U - u)\Theta(U - u)^{T}\|_{1} + \sum_{i} \left(1 - \sum_{j} \Theta(U - u)_{ij}\right),$$
(3)

where *u* is a scalar.

To solve (3), we fix U and discretize the domain $\{u : 0 \le u \le \max(U)\}$ to select \hat{u} , which minimizes the optimization problem (3). Finally, we obtain the binary matrix U as follows:

$$U := \Theta(U - \hat{u})$$

To initialize U, we employ the algorithm of multiplicative update rules developed for SNMF, which is summarized in Algorithm 1. We set the iteration number C equal to 100 in this paper. The time complexity is $C \times O(n^2c)$ for updating U [26].

Algorithm 1: Symmetric non-negative matrix factorization (least-squares error)

Input: *A*, *C*

Output: U

1: Initialize elements of U with non-negative random numbers drawn from [0,1]

2: for t = 1 : C do 3: $U_{ij} := U_{ij} \frac{(AU)_{ij}}{(UU^T U)_{ij}}$ 4: $U_{ij} := \frac{U_{ij}}{\sum_j U_{ij}}$ 5: end for



FIG. 3. (Color online) Mean generalized NMI with the standard deviation of SBMF on the overlapping LFR networks. The average degree is denoted by $\langle k \rangle$. We run the algorithm ten times with random initial conditions.

III. MODEL SELECTION

The problem of community detection is an unsupervised learning task and the number of communities is unknown in real applications. Several methods have been developed to infer the number of communities. The most popular one uses the modularity function Q [14,15], namely, choosing the number of communities c at which the modularity function achieves the maximum. However, as explained above, the modularity function may lead to both the underestimation and overestimation of community numbers [27]. To overcome these limitations, we modify the partition density to better estimate the appropriate number of communities. Originally, the partition density was defined in terms of edges [3].

Formally, the standard partition density of the community α can be defined as [3]

$$D_{\alpha} = \frac{m_{\alpha} - (n_{\alpha} - 1)}{n_{\alpha}(n_{\alpha} - 1)/2 - (n_{\alpha} - 1)}$$

where n_{α} and m_{α} are the number of nodes and the number of edges in the community α , respectively. Then the overall partition density of the network can be defined as the weighted sum of D_{α} , $\alpha = 1, 2, ..., c$. The weight for each community was given by the number of edges in the community in the original formulation. Here we use the number of nodes as the weights because this formulation gives us better results. Thus



FIG. 2. (Color online) Mean normalized mutual information with the standard deviation of Bayesian NMF and SBMF on (a) GN networks and (b) nonoverlapping LFR networks. We run the algorithm ten times with random initial conditions.



FIG. 4. Number of communities estimated by SBMF on overlapping LFR networks with the standard deviation errors. Here $\langle k \rangle$ represents the average degree of nodes in the network and "Actual Number" represents the actual number of communities in the corresponding LFR network. We run the algorithm ten times with random initial conditions.

the partition density of the whole network is

$$D = \frac{2}{N} \sum_{\alpha=1}^{c} n_{\alpha} \frac{m_{\alpha} - (n_{\alpha} - 1)}{(n_{\alpha} - 2)(n_{\alpha} - 1)},$$

where N is the sum of the sizes of different communities and the number of outliers. Here N may be larger than the number of nodes n in the network since the overlapping nodes are counted more than once.

Nodes can be assigned to multiple communities. However, in practice one can observe that the nodes with several community labels are not common due to limited energy, time, and resources. Hence we add a penalty term into the denominator to control the nodes' activity degree and the definition can thus be updated as follows:

$$D = \frac{2}{N} \sum_{\alpha=1}^{c} \frac{n_{\alpha}}{q_{\alpha}} \frac{m_{\alpha} - (n_{\alpha} - 1)}{(n_{\alpha} - 2)(n_{\alpha} - 1)},$$

where $q_{\alpha} = \max_{j \in \alpha} l_j$, $j \in \alpha$ means that the node *j* belongs to the community α , and l_j is the number of community labels that the node *j* has. Since the partition density *D* only considers the local information in each community, it does not suffer from the resolution limit problem [3].

IV. ILLUSTRATIVE EXAMPLE

In this section we use a toy example to illustrate how the proposed method works. Figure 1 shows a toy network with two communities, where nodes 5 and 6 are bridges between them. First, we build the adjacency matrix A and solve the SBMF model (2) with different numbers of communities c. Then we calculate the partition density for each c to select the best one. Finally, the corresponding binary matrix U is obtained. As one can observe, the two overlapping communities can be explicitly recovered from the output U of SBMF model: 1 means that the node belongs to the corresponding community and 0 means it does not.

V. EXPERIMENTAL RESULTS

In this section we test the effectiveness of our method on both the synthetic and the real-world networks. Codes are available through Ref. [28].

A. Description of data sets

1. The Girvan-Newman benchmark network. The Girvan-Newman (GN) benchmark network [1] has four equally sized nonoverlapping communities with 32 nodes each. On average each node has Z_{in} edges connecting with the others in its own community and Z_{out} edges connecting with the other three communities. As expected, the communities become less clear with increasing Z_{out} . Here $Z_{in} + Z_{out}$ is set equal to 16.

2. The Lancichinetti-Fortunato-Radicchi benchmark network. The Lancichinetti-Fortunato-Radicchi (LFR) benchmark network [29] model was proposed to address most characteristics of real networks, e.g., size of the network and heterogeneous degree distribution, that the GN networks do not capture. In LFR benchmarks both the degree and the community size distributions obey power laws with exponents γ and β . Each node has a fraction $1 - \mu$ of its neighbors in its own community and a fraction μ in the other communities. Furthermore, nodes can be assigned to multiple communities.

In this paper we set the parameters of the LFR benchmark as follows: The number of nodes is 1000, the maximum degree is 50, the exponent of the degree distribution γ is 2, and that of the community size distribution β is 1. We set the parameters for the nonoverlapping LFR benchmark as follows: The average degree of the nodes is 20 and the range of the mixing parameter



FIG. 5. Number of communities estimated by Bayesian NMF and SBMF on (a) GN and (b) nonoverlapping LFR networks with the standard deviation errors. The black horizontal line is the actual number of communities. We run the algorithm ten times with random initial conditions.



FIG. 6. Averaged partition density of SBMF versus community number on real-world networks: (a) football network, (b) political book network, (c) dolphin network, and (d) jazz band network. We run the algorithm ten times with random initial conditions.

 μ is from 0.1 to 0.6. For the overlapping LFR benchmark, we set the parameters as follows: The average degree of the nodes is from 15 to 25, the mixing parameter μ is 0.1, the minimum community size is 20, the maximum community size is 100, and the fraction of the overlapping nodes is from 0.1 to 0.5.

3. Football. This data set is the network of 115 American football teams [1]. There are 613 edges. The teams are represented as nodes and the nodes are connected if there is a game between them. The teams are divided into 12 conferences, where the teams usually play more games with others in the same conference, inducing community structure.

4. Political books.: This data set is the Amazon copurchasing network with 105 books about U.S. politics [30]. There are 441 edges. Nodes are books and edges represent copurchasing of books by the same buyers.

5. *Dolphins*. This data set is the social communication network of 62 bottlenose dolphins that lived in Doubtful Sound, New Zealand [31]. There are 159 edges.

6. Jazz bands. This data set is the collaboration network of jazz bands [32]. There are 198 nodes representing the bands and 2742 edges connecting the bands if there is at least one musician in common.

B. Assessment standards

To evaluate the detection performance on nonoverlapping synthetic networks (GN and nonoverlapping LFR), we use the normalized mutual information [33]. The value can be

TABLE I. Number of communities estimated by Bayesian NMF and SBMF on real-world networks. The last column lists the running time of SBMF.

Data set	Bayesian NMF	SBMF	Running time (s)
football	9.2 ± 0.63	$ \begin{array}{r} 12 \pm 0 \\ 6 \pm 0 \\ 5 \pm 0 \\ 2 \pm 0 \end{array} $	0.02
dolphin	7.6 ± 0.70		0.01
political books	6.9 ± 0.57		0.03
jazz	11.1 ± 0.74		0.07

formulated as follows:

$$I_{\text{norm}}(M_1, M_2) = \frac{\sum_{i=1}^k \sum_{j=1}^k n_{ij} \ln \frac{n_{ij}n}{n_i^{(1)}n_j^{(2)}}}{\sqrt{\left(\sum_{i=1}^k n_i^{(1)} \ln \frac{n_i^{(1)}}{n}\right)\left(\sum_{j=1}^k n_j^{(2)} \ln \frac{n_j^{(2)}}{n}\right)}}$$

where M_1 and M_2 are the ground-truth community label and the computed community label, respectively, k is the community number, n is the number of nodes, n_{ij} is the number of nodes in the ground-truth community i that are assigned to the computed community j, $n_i^{(1)}$ is the number of nodes in the ground-truth community i, $n_j^{(2)}$ is the number of nodes in the computed community j and ln here is the natural logarithm. The larger the normalized mutual information (NMI) value, the better the community partition. For overlapping LFR benchmarks, we use the generalized normalized mutual information [9].

C. Experimental results on synthetic networks

In this section we compare the Bayesian NMF [7] with our SBMF model and give the numerical results of NMI

TABLE II. List of the abnormal teams that played more frequently against the ones in the other conferences. The terms N_S and N_O denote the times that the team played against other teams in the same conference and in the other conferences, respectively.

Team ID	N_S	N_O	Team ID	N_S	No
37	0	8	60	2	6
43	0	7	64	2	7
81	1	10	70	3	8
83	1	10	98	3	5
91	0	9	111	0	11
12	4	6	29	0	9
25	3	7	59	2	8
51	3	6			



FIG. 7. (Color online) Comparison of (a) real grouping and (b) communities found by the SBMF model in the football network. There are no outliers and overlapping teams. All of the misclustered teams are abnormal ones, which are listed in Table II.

on the GN networks and the nonoverlapping LFR networks and generalized NMI on the overlapping LFR networks. We also compare the abilities of Bayesian NMF and SBMF to infer the appropriate number of communities. The results are averaged over ten trials and are shown in Figs. 2-5. From these figures one can observe the following. (i) The results of both the Bayesian NMF and the SBMF model decrease when Z_{out} or μ are increasing and the standard deviations are low. (ii) SBMF consistently outperforms Bayesian NMF, especially on the nonoverlapping LFR networks. For example, when $\mu = 0.6$, the NMI of the proposed model (96.32%) is 35% higher than that of the Bayesian model (60.32%). (iii) On the overlapping LFR networks, SBMF outperforms the methods in [34], where a systematic comparison of different methods on the overlapping LFR benchmarks with the same parameter settings is conducted. In addition, the generalized NMI does not decrease significantly with the increasing fraction of the overlapping nodes, indicating that our model is applicable for highly overlapping networks. (iv) The inferred community numbers based on the modified partition density are closer to the real ones and the standard deviations are lower. For example, the community number of the GN networks has been perfectly recovered. Note that the partition density can be applied to both the cases of nonoverlapping and overlapping communities, which makes it suitable for a wide range of applications.



FIG. 8. (Color online) Exponential entropy of the nodes in different conferences. The higher the value, the fuzzier the membership degrees. For instance, e^{H_i} will be *c* if the node belongs equally to *c* communities.

D. Experimental results on real networks

In this section we conduct experiments on the real-world networks. Figure 6 gives the results of partition densities under different community numbers c and Table I gives the inferred community numbers of different networks.

We use the football network as a case study. The teams in the network are assigned into 12 conferences, most of which play more games against others in the same conference. However, there are some abnormal teams that play against the ones in other conferences more frequently. Table II lists the basic information about these teams, among which teams 37, 43, 81, 83, and 91 are in the conference IA Independents [the black ones in Fig. 7(a)] and teams 12, 25, 51, 60, 64, 70, and 98 are in the conference Sunbelt [the dark green ones in Fig. 7(a), next to the black ones]. To determine the best community number c, we calculate the partition density given different c. The peak value is achieved at c = 12. The corresponding partitioning result of SBMF is shown in Fig. 7(b), from which one can observe that (i) our proposed BMF model only misclusters the abnormal teams, (ii) the abnormal teams are reallocated to the other conferences based on the real topology structures, and (iii) no outliers or overlapping teams are detected.

Based on the partitioning result, one can analyze the membership degrees of nodes using the matrix U obtained from Algorithm 1 before discretizing it. For example, one can use the exponential entropy e^{H_i} , i = 1, 2, ..., n [35], to analyze the positions of different nodes in the corresponding communities, where the entropy H_i is

$$H_i = -\sum_j U_{ij} \ln U_{ij}.$$

The exponential entropies of the nodes in ten normal conferences (i.e., not including the conferences IA Independents and Sunbelt) are shown in Fig. 8. The lower the value, the more likely the node is to stay in its own community.

VI. CONCLUSION

In this paper we presented a symmetric binary matrix factorization model to detect overlapping community structures. The model can explicitly identify the community memberships of the nodes, which are allowed to belong to multiple communities or to be outliers. We also gave a revised partition density to automatically infer the community number in the network. The experiments conducted on both the synthetic and the real-world networks showed the effectiveness of the proposed method. In summary, the SBMF model is parameter-free, easy to implement, and the partition density is good at determining the number of communities. Based on our works, there are several interesting problems for future work, including direct optimization of partition density, generalization of the proposed model to weighted networks and directed networks, and applying the model to recommendation systems. PHYSICAL REVIEW E 87, 062803 (2013)

ACKNOWLEDGMENTS

Z.-Y.Z was supported by the National Natural Science Foundation of China under Grant No. 61203295 and Program for Innovation Research in Central University of Finance and Economics. Y.W. was supported by the National Natural Science Foundation of China under Grants No. 11131009 and No. 61171007.

- [1] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. 99, 7821 (2002).
- [2] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, Nature (London) 435, 814 (2005).
- [3] Y. Ahn, J. Bagrow, and S. Lehmann, Nature (London) 466, 761 (2010).
- [4] X.-S. Zhang, Z. Li, R.-S. Wang, and Y. Wang, J. Comb. Opt. 23, 425 (2012).
- [5] S. Gregory, J. Stat. Mech. (2011) P02017.
- [6] T. Nepusz, A. Petróczi, L. Négyessy, and F. Bazsó, Phys. Rev. E 77, 016107 (2008).
- [7] I. Psorakis, S. Roberts, M. Ebden, and B. Sheldon, Phys. Rev. E 83, 066114 (2011).
- [8] H. Li, J. Zhang, Z. Liu, L. Chen, and X. Zhang, Eur. Phys. J. B 85, 1 (2012).
- [9] A. Lancichinetti, S. Fortunato, and J. Kertész, New J. Phys. 11, 033015 (2009).
- [10] D. Lee and H. Seung, Nature (London) 401, 788 (1999).
- [11] D. Seung and L. Lee, Adv. Neural Inf. Process. Syst. 13, 556 (2001).
- [12] Z. Zhang, C. Ding, T. Li, and X. Zhang, in *Proceedings of the Seventh IEEE International Conference on Data Mining*, edited by N. Ramakrishnan, O. R. Zaïane, Y. Shi, C. W. Clifton, and X. Wu (IEEE Computer Society, Los Alamitos, 2007), pp. 391–400.
- [13] Z. Zhang, T. Li, C. Ding, X. Ren, and X. Zhang, Data Min. Knowl. Disc. 20, 28 (2010).
- [14] M. E. J. Newman and M. Girvan, Phys. Rev. E **69**, 026113 (2004).
- [15] M. Newman, Proc. Natl. Acad. Sci. U.S.A. 103, 8577 (2006).

- [16] S. Fortunato and M. Barthelemy, Proc. Natl. Acad. Sci. 104, 36 (2007).
- [17] J. P. Bagrow, Phys. Rev. E 85, 066118 (2012).
- [18] X. Zhang, R. Wang, Y. Wang, J. Wang, Y. Qiu, L. Wang, and L. Chen, Europhys. Lett. 87, 38002 (2009).
- [19] B. H. Good, Y. A. de Montjoye, and A. Clauset, Phys. Rev. E 81, 046106 (2010).
- [20] J. Brunet, P. Tamayo, T. Golub, and J. Mesirov, Proc. Natl. Acad. Sci. U.S.A. 101, 4164 (2004).
- [21] Z. Zhang, T. Li, and C. Ding, Knowl. Inf. Syst. 34, 243 (2013).
- [22] Z. Zhang (unpublished).
- [23] F. Wang, T. Li, X. Wang, S. Zhu, and C. Ding, Data Min. Knowl. Disc. 22, 493 (2011).
- [24] X. Ma, L. Gao, X. Yong, and L. Fu, Physica A 389, 187 (2010).
- [25] M. Zarei, D. Izadi, and K. A. Samani, J. Stat. Mech. (2009) P11013.
- [26] C.-J. Lin, Neural Comput. 19, 2756 (2007).
- [27] A. Kehagias, arXiv:1209.2678.
- [28] https://github.com/ZhongYuanZhang/SBMF.git.
- [29] A. Lancichinetti and S. Fortunato, Phys. Rev. E 80, 016118 (2009).
- [30] http://www.orgnet.com/cases.html
- [31] D. Lusseau, K. Schneider, O. Boisseau, P. Haase, E. Slooten, and S. Dawson, Behav. Ecol. Sociobiol. 54, 396 (2003).
- [32] P. Gleiser and L. Danon, Adv. Complex Syst. 6, 565 (2003).
- [33] A. Strehl and J. Ghosh, J. Mach. Learn. Res. 3, 583 (2002).
- [34] A. Lancichinetti and S. Fortunato, Phys. Rev. E 80, 056117 (2009).
- [35] L. Campbell, Probab. Theory Relat. Fields 5, 217 (1966).