



PAPER

Optimization of identifiability for efficient community detection

OPEN ACCESS

RECEIVED

13 January 2020

REVISED

23 April 2020

ACCEPTED FOR PUBLICATION

29 April 2020

PUBLISHED

19 June 2020

Original content from
this work may be used
under the terms of the
[Creative Commons
Attribution 4.0 licence](#).

Any further distribution
of this work must
maintain attribution to
the author(s) and the
title of the work, journal
citation and DOI.

Hui-Jia Li¹ , Lin Wang² , Yan Zhang³ and Matjaž Perc^{4,5,6,7} ¹ School of Science, Beijing University of Posts and Telecommunications, Beijing 100876, People's Republic of China² Department of Genetics, University of Cambridge, Cambridge, CB2 3EH, United Kingdom³ Alibaba Local Services Lab, Alibaba Group, Shanghai 200333, People's Republic of China⁴ Faculty of Natural Sciences and Mathematics, University of Maribor, Koroška cesta 160, 2000 Maribor, Slovenia⁵ Department of Medical Research, China Medical University Hospital, China Medical University, Taichung, Taiwan⁶ Complexity Science Hub Vienna, Josefstädterstraße 39, 1080 Vienna, Austria⁷ Author to whom any correspondence should be addressed.E-mail: lw660@cam.ac.uk and matjaz.perc@gmail.com**Keywords:** complex system, complex network, community detection, matrix factorizationSupplementary material for this article is available [online](#)**Abstract**

Many physical and social systems are best described by networks. And the structural properties of these networks often critically determine the properties and function of the resulting mathematical models. An important method to infer the correlations between topology and function is the detection of community structure, which plays a key role in the analysis, design, and optimization of many complex systems. The nonnegative matrix factorization has been used prolifically to that effect in recent years, although it cannot guarantee balanced partitions, and it also does not allow a proactive computation of the number of communities in a network. This indicates that the nonnegative matrix factorization does not satisfy all the nonnegative low-rank approximation conditions. Here we show how to resolve this important open problem by optimizing the identifiability of community structure. We propose a new form of nonnegative matrix decomposition and a probabilistic surrogate learning function that can be solved according to the majorization–minimization principle. Extensive *in silico* tests on artificial and real-world data demonstrate the efficient performance in community detection, regardless of the size and complexity of the network.

1. Introduction

Describing and modeling real-world complex systems such as protein–protein interactions [1–3], the brain connectome [4–7], friendship relationships [8–12], or public transport [13–16, 18]) by means of complex networks provide many insights towards better understanding their structure and function. As a critical properties of real networks, community structure asserts that some groups of nodes—the communities—are much more densely connected with each other than with other nodes outside of the group [19–22]. The existence and properties of the community structure can have far-reaching consequences for many processes that unfold on networks, including the prolonged duration of epidemics [23], difficulties in eliminating vaccine hesitancy [24, 25], and biased recommendations [26] and accelerated rumor spreading [27–30]. Due to the theoretical and practical importance [31–34], the detection of network communities has been a popular and vibrant research area, with numerous applications across the physical and social sciences [35–38].

Existing methods for detecting network communities mainly rely on the optimization of quality functions that are defined according to the similarity between different nodes. Nonnegative matrix factorization is a mainstream technology for detecting network communities [39]. The basic protocol of the nonnegative matrix factorization methods uses nonnegative low-rank approximations to factorize a node-to-node similarity matrix into node-to-cluster low-rank matrices. Ding *et al* [40] verified the approximate equivalence between the nonnegative matrix factorization and the seminal *k*-means clustering.

However, the nonnegative matrix factorization methods do not satisfy all of the nonnegative low-rank approximation conditions, so that they may lead to unbalanced network partitions in which some communities are assigned with much more nodes than others [39]. Besides, nonnegative matrix factorization methods are limited to account for curved manifolds and to proactively compute the optimal number of communities [41].

To resolve these limitations, we introduce the *identifiability* of network communities as the objective of optimization, which is defined as the discrepancy between the network similarity matrix and the doubly-stochastically normalized community incidence matrix. This objective function is solvable for all possible community partitions, and the optimal number of communities can also be determined proactively. To deal with sparse similarity inputs, we use the Kullback–Leibler divergence instead of the conventional distance-based measures. We also propose a new form of nonnegative matrix factorization to explore the solution space, which yields a probabilistic objective function that can be solved by an efficient algorithm using the majorization–minimization principle. We perform extensive experiments on multiple types of datasets, showing that our approach substantially improves the accuracy and computational efficiency of community detection, regardless of the size and complexity of the network.

2. Identifiability

Consider a network with N nodes that form r communities. The operation of assigning nodes into each community is described by an $N \times r$ binary community indicator matrix \bar{W} , in which each entry $\bar{W}_{ik} = 1$ (or 0) indicates that the node i is (not) assigned to the community k . The dot product of the community indicator matrix is defined as the community incidence matrix $\bar{B} = \bar{W}\bar{W}^T$, in which each entry $\bar{B}_{ij} = 1$ (or 0) indicates that the nodes i and j are (not) assigned to the same community. To reduce the unbalanced partitions, it is more convenient to use the normalized community indicator matrix W with each entry denoting $W_{ik} = \bar{W}_{ik} / \sqrt{\sum_{v=1}^r \bar{W}_{vk}}$, since the normalized community incidence matrix $B = WW^T$ becomes a doubly stochastic matrix (i.e. $\sum_{i=1}^N B_{ij} = 1$, $\sum_{j=1}^N B_{ij} = 1$) [40, 43].

Let S be the similarity matrix, in which each entry S_{ij} accounts for the probability of similarity between nodes i, j . For example, the similarity of two nodes in the protein–protein interaction networks can be quantified by the Hamming distance between their amino acid sequences [1]. The probability of similarity is expected to be higher for the nodes within the same community than those part in different communities, so that the normalized community incidence matrix B converges to the similarity matrix S when nodes are accurately assigned to each community.

We can also interpret matrix B in a probabilistic form. Denote $W_{ik} = P(i|k)$ as the conditional probability that any node i belongs to a given community k , and $P(k|i)$ the conditional probability that any community k has a given node i . Considering an uninformative prior for selecting nodes, i.e. $P(i) = 1/N$, we have

$$P(i|k) = \frac{P(k|i)P(i)}{\sum_{j=1}^N P(k|j)P(j)} = \frac{P(k|i)}{\sum_{i=1}^N P(k|i)}, \quad (1)$$

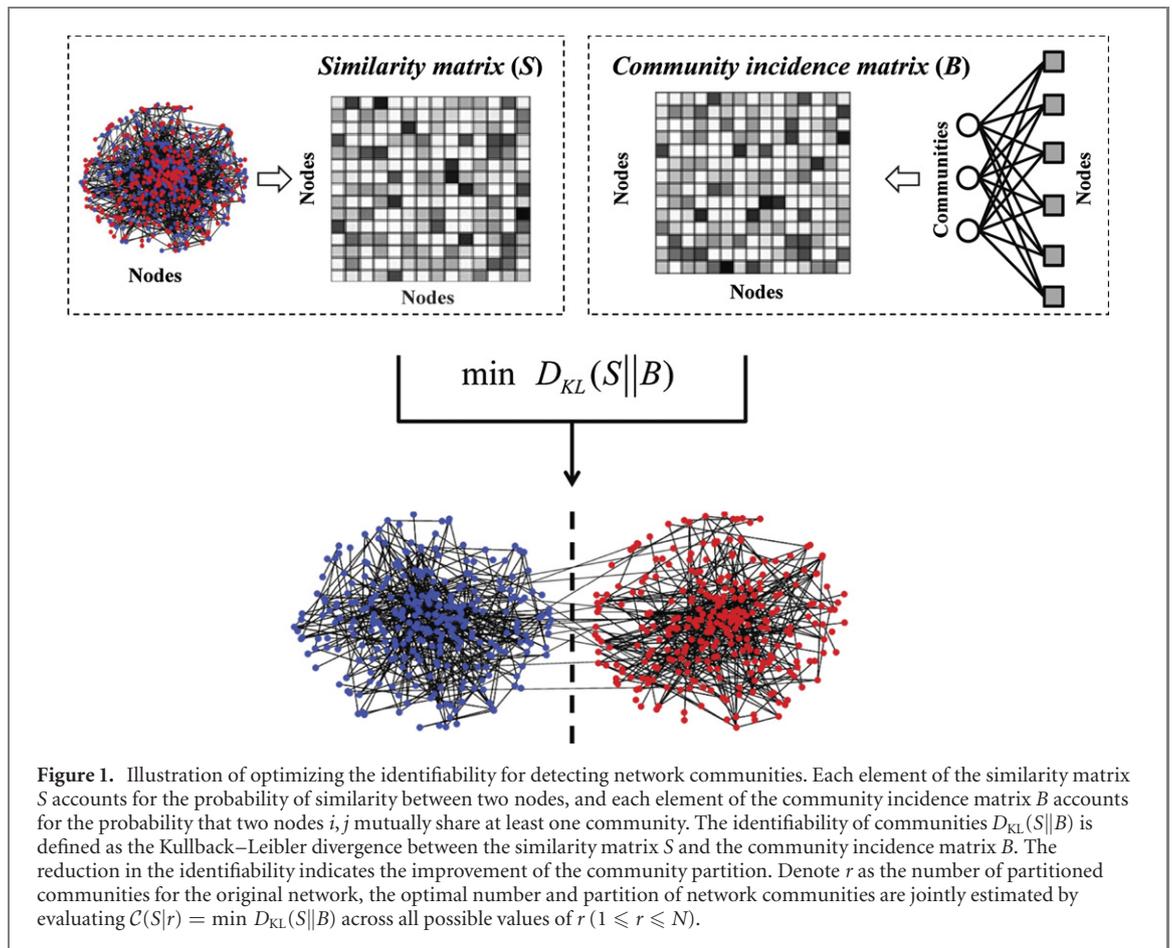
according to the Bayes' rule. The probability that two nodes i, j mutually share at least one community is k is:

$$B_{ij} = \sum_{k=1}^r \frac{P(i|k)P(j|k)}{\sum_{\ell=1}^r P(\ell|k)} = \sum_{k=1}^r \frac{P(k|i)P(k|j)}{\sum_{j=1}^N P(k|j)} = \sum_{k=1}^r P(j|k)P(k|i) = P(j|i). \quad (2)$$

The identifiability of communities $D(S||B)$ is defined as the discrepancy between the similarity matrix S and the normalized community incidence matrix B , so that the reduction in the identifiability indicates the improvement of the community partition (figure 1). Let $\mathcal{C}(S|r) = \min_{(B|r)} D(S||B)$ be the optimized (or minimized) identifiability, given that the original network is partitioned into r communities. Evaluating $\mathcal{C}(S|r)$ over all possible number of communities r ($1 \leq r \leq N$), we can jointly estimate the optimal number and partition of network communities. The optimization of $\mathcal{C}(S|r)$ often need to search a sparse solution space [43–45], so we propose a new framework for minimizing the identifiability.

3. Optimization of identifiability via matrix decomposition

In this section, we first formulate the identifiability using the Kullback–Leibler divergence, and then introduce a multiplicative minimization algorithm that efficiently optimizes the identifiability.



3.1. Kullback–Leibler divergence

The use of simple discrepancy measures such as the Euclidean distance or Hamming distance does not work well for many real-world problems that only contain weakly informative data. Inspired by the seminal work of probabilistic latent semantic indexing [46], we resort to use the Kullback–Leibler divergence to measure the discrepancy (or approximation error) between the dense normalized community incidence matrix B and the sparse node similarity matrix S . As such, the optimization problem is summarized as follows:

$$\begin{aligned} \min_{W \geq 0} D_{KL}(S||B) &= \sum_{i=1}^N \sum_{j=1}^N \left(S_{ij} \log \frac{S_{ij}}{B_{ij}} - S_{ij} + B_{ij} \right) \\ \text{subject to } B_{ij} &= \sum_{k=1}^r \frac{W_{ik} W_{jk}}{\sum_v W_{vk}}, \\ \sum_{k=1}^r W_{jk} &= 1, \quad i = 1, \dots, N. \end{aligned} \quad (3)$$

where the normalized community incidence B_{ij} is estimated with the relative probability of randomly selecting two nodes i, j that belong to the same community k , and nonnegative $W_{ik} = P(k|i)$ denotes the probability that a given node i belongs to the community k . We use $-S_{ij} + B_{ij}$ to measure the discrepancy between these two different matrices, since we can naturally consider that the community incidence matrix B should be close to the similarity matrix S for a good community partition. By eliminating all the constant terms (e.g. $\sum_{i=1}^N \sum_{j=1}^N B_{ij} = N$) from the objective function of equation (3), the optimization is transformed to the maximization of $\sum_{i=1}^N \sum_{j=1}^N S_{ij} \log B_{ij}$.

It offers several advantages to improve the efficiency of optimization: (1) only non-zero elements in the similarity matrix S need to be evaluated by the objective and gradient functions; and (2) the logarithmic functions in the low-rank community incidence matrix B only need additive calculations. Therefore, we use a Dirichlet prior to lower the computational complexity (section 3.2) and use a convex–concave procedure [47] for the optimization (section 3.3).

3.2. Regularization

In equation (3), each row of matrix W is normalized. Suppose that the elements of each row in W follow a Dirichlet distribution [46], the complexity of matrix W can be constrained with log-Dirichlet prior, i.e., $\sum_{i=1}^N \sum_{k=1}^r \log W_{ik}$. Here, the complexity of a matrix also means the sparsity and regularity, and less the log-Dirichlet prior, less the complexity of a matrix. Therefore, we use the following regularized cost function:

$$\mathcal{J}(W) = -\sum_{i=1}^N \sum_{j=1}^N S_{ij} \log B_{ij} - (\alpha - 1) \sum_{i=1}^N \sum_{k=1}^r \log W_{ik}. \quad (4)$$

If all similarities S_{ij} are integer-valued, the objective function of equation (3) can be the log-likelihood of generative models as follows: (1) drawing the rows of W according to a uniform Dirichlet distribution with parameter α ; and (2) for $t = 1, \dots, T$, add one to each entry $W_{ij} \sim \text{multinomial}(\frac{1}{N}B, 1)$. If $\alpha = 1$, the Dirichlet prior disappears; whereas if $\alpha > 1$, by smoothing each entry of matrix W , the prior provides a wider relaxation that is usually desired in preliminary stages of W learning. In this paper, we employ the Dirichlet prior only to simplify the optimization steps, without change the control of KL divergence.

3.3. Optimization

We implement the optimization using nonnegative matrix factorization with multiplicative updating rule [40]. We first split the gradient of the objective function \mathcal{J} with respect to the nonnegative matrix W into two nonnegative components:

$$\nabla_{ik} \stackrel{\text{def}}{=} \frac{\partial \mathcal{J}}{\partial W_{ik}} = \nabla_{ik}^+ - \nabla_{ik}^-. \quad (5)$$

We then iterate the optimization using the following multiplicative updating rule:

$$W_{ik} \leftarrow W_{ik} \frac{\nabla_{ik}^-}{\nabla_{ik}^+}, \quad (6)$$

which retains the nonnegativity of matrix W and does not need to adjust the step size per iteration. Since the optimization monotonically reduces the objective function \mathcal{J} , the multiplicative updating rule can warrant the convergence of W . To deal with the row-based probability constraint on W , we introduce a Lagrangian multiplier $\{\lambda_i, i = 1, \dots, N\}$ which is subject to the following constraint:

$$\mathcal{L}(W, \lambda) = \mathcal{J}(W) + \sum_i \lambda_i \left(\sum_{k=1}^r W_{ik} - 1 \right). \quad (7)$$

This improves the multiplicative update rule as follows:

$$W'_{ik} = W_{ik} \frac{\nabla_{ik}^- - \lambda_i}{\nabla_{ik}^+}, \quad (8)$$

in which the gradient of the objective function \mathcal{J} is split as:

$$\frac{\partial \mathcal{J}}{\partial W} = \underbrace{[(W^T Z W)_{kk} s_k^{-2} + W_{ik}^{-1}]}_{\nabla_k^+} - \underbrace{[2(ZW)_{ik} s_k^{-1} + \alpha W_{ik}^{-1}]}_{\nabla_k^-}, \quad (9)$$

where $s_k = \sum_{v=1}^N W_{vk}$, and the ratio $Z_{ij} = S_{ij}/B_{ij}$ only calculates non-zero entries of S , without requirement to specify the entire matrix B . Given $\sum_k W_{ik}' = 1$, we have

$$\lambda_i = \frac{b_i - 1}{a_i}, \quad (10)$$

where

$$a_i = \sum_{l=1}^r \frac{W_{il}}{\nabla_{il}^+} \quad \text{and} \quad b_i = \sum_{l=1}^r W_{il} \frac{\nabla_{il}^-}{\nabla_{il}^+}. \quad (11)$$

Combining equation (11) and equation (8), we have

$$W_{ik} \leftarrow W_{ik} \frac{\nabla_{ik}^- a_i + 1 - b_i}{\nabla_{ik}^+ a_i}. \quad (12)$$

To retain the positivity of W , we add the term b_i to both the denominator and numerator of equation (12), which gives the following updating rule:

$$W_{ik} \leftarrow W_{ik} \frac{\nabla_{ik}^- a_i + 1}{\nabla_{ik}^+ a_i + b_i}. \tag{13}$$

This algorithm warrants the convergence of the optimization because the cost function \mathcal{J} can be monotonically decreased in each iteration. We validate this point using the majorization–minimization principle in the following. W and \tilde{W} distinguish the current estimate and variable, respectively.

(1) Majorization

$$\begin{aligned} \text{Let } \phi_{ijk} &= \frac{W_{ik} W_{jk}}{\sum_{v=1}^N W_{vk}} \left(\sum_{l=1}^r \frac{W_{il} W_{jl}}{\sum_{v=1}^N W_{vl}} \right)^{-1}. \\ \mathcal{L}(\tilde{W}, \lambda) &\leq - \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^r S_{ij} \phi_{ijk} \left[\log \tilde{W}_{ik} + \log \tilde{W}_{jk} - \log \sum_v \tilde{W}_{vk} \right] \\ &\quad - (\alpha - 1) \sum_{i=1}^N \sum_{k=1}^r \log \tilde{W}_{ik} + \sum_{i=1}^N \sum_{k=1}^r \lambda_i \tilde{W}_{ik} + C_1 \\ &\leq - \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^r S_{ij} \phi_{ijk} \left[\log \tilde{W}_{ik} + \log \tilde{W}_{jk} - \frac{\sum_{v=1}^N \tilde{W}_{vk}}{\sum_{v=1}^N W_{vk}} \right] \\ &\quad - (\alpha - 1) \sum_{i=1}^N \sum_{k=1}^r \log \tilde{W}_{ik} + \sum_{i=1}^N \sum_{k=1}^r \lambda_i \tilde{W}_{ik} + C_2 \\ &\leq - \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^r S_{ij} \phi_{ijk} \left[\log \tilde{W}_{ik} + \log \tilde{W}_{jk} - \frac{\sum_{v=1}^N \tilde{W}_{vk}}{\sum_{v=1}^N W_{vk}} \right] \\ &\quad - (\alpha - 1) \sum_{i=1}^N \sum_{k=1}^r \log \tilde{W}_{ik} + \sum_{i=1}^N \sum_{k=1}^r \lambda_i \tilde{W}_{ik} \\ &\quad + \sum_{i=1}^N \sum_{k=1}^r \left(\frac{1}{a_i} + \frac{\alpha}{W_{ik}} \right) W_{ik} \left(\frac{\tilde{W}_{ik}}{W_{ik}} - \log \frac{\tilde{W}_{ik}}{W_{ik}} - 1 \right) + C_2 \\ &\stackrel{\text{def}}{=} G(\tilde{W}, W), \end{aligned} \tag{14}$$

where

$$C_1 = \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^r S_{ij} \phi_{ijk} \log \phi_{ijk}, \tag{15}$$

$$C_2 = C_1 + \sum_{i=1}^N \sum_{j=1}^N \sum_{k=1}^r S_{ij} \phi_{ijk} \left(\log \sum_{v=1}^N W_{vk} - 1 \right) \tag{16}$$

are constants that are independent of the variable \tilde{W} . The first two inequalities depend on the convexity and concavity of the logarithmic functions. By further adding the same constant $\frac{1}{a_i} + \frac{\alpha}{W_{ik}}$ to both the denominator and numerator, the third inequality warrants the positivity of each entry in the renewed matrix, because $x \geq 1 + \log x$ when $x > 0$. The upper boundaries mentioned above are tight at $\tilde{W} = W$, i.e. $G(W, W) = \mathcal{L}(W, \lambda)$.

(2) Minimization

$$\begin{aligned} \frac{\partial G}{\partial \tilde{W}_{ik}} &= \nabla_{ik}^+ - \frac{\alpha}{W_{ik}} - \frac{W_{ik}}{\tilde{W}_{ik}} \left(\nabla_{ik}^- - \frac{\alpha}{W_{ik}} \right) + \lambda_i + \left(\frac{1}{a_i} + \frac{\alpha}{W_{ik}} \right) W_{ik} \left(\frac{1}{W_{ik}} - \frac{1}{\tilde{W}_{ik}} \right) \\ &= - \frac{W_{ik}}{\tilde{W}_{ik}} \left(\nabla_{ik}^- + \frac{1}{a_i} \right) + \left(\nabla_{ik}^+ + \frac{b_i}{a_i} \right). \end{aligned} \tag{17}$$

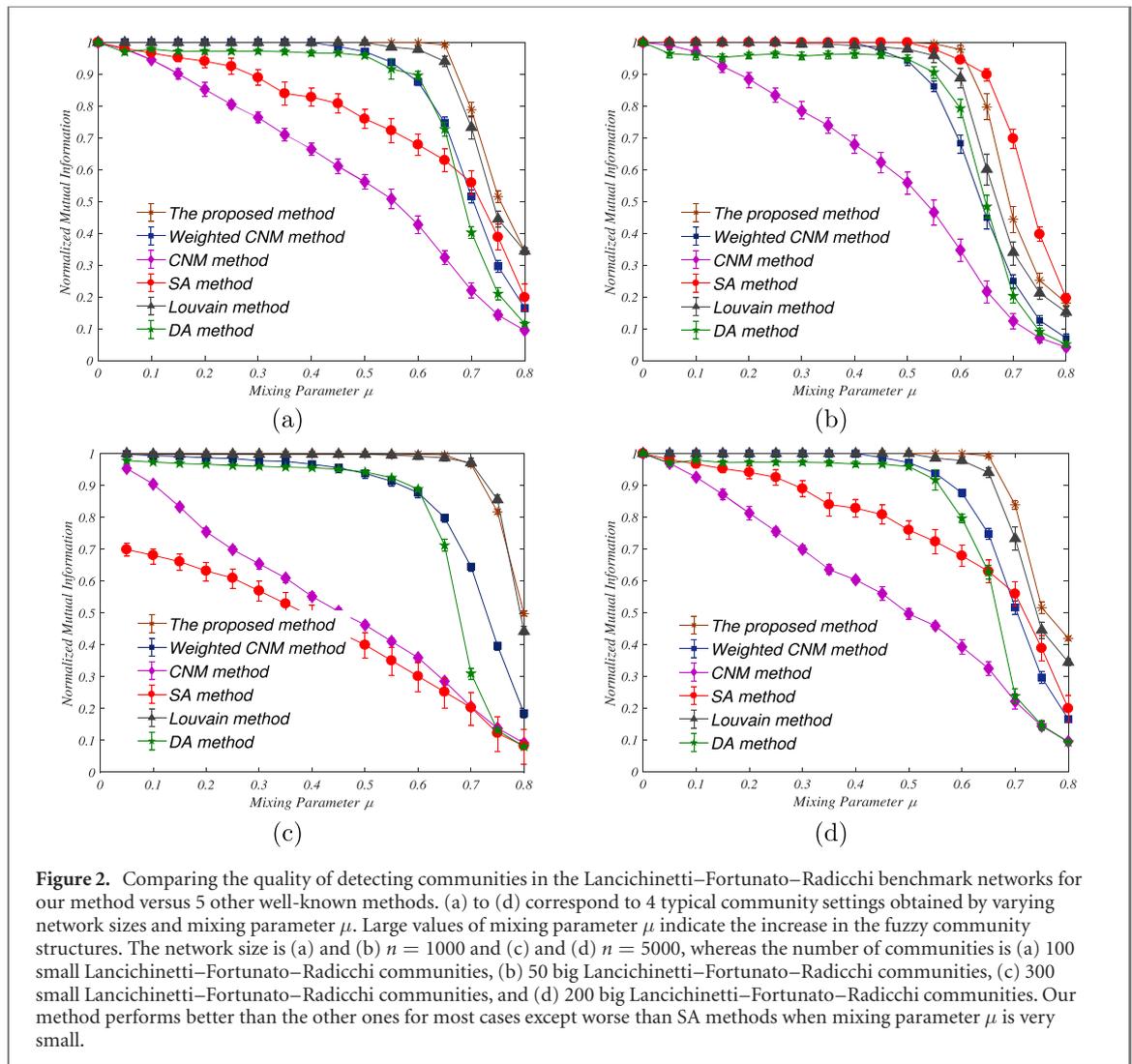


Figure 2. Comparing the quality of detecting communities in the Lancichinetti–Fortunato–Radicchi benchmark networks for our method versus 5 other well-known methods. (a) to (d) correspond to 4 typical community settings obtained by varying network sizes and mixing parameter μ . Large values of mixing parameter μ indicate the increase in the fuzzy community structures. The network size is (a) and (b) $n = 1000$ and (c) and (d) $n = 5000$, whereas the number of communities is (a) 100 small Lancichinetti–Fortunato–Radicchi communities, (b) 50 big Lancichinetti–Fortunato–Radicchi communities, (c) 300 small Lancichinetti–Fortunato–Radicchi communities, and (d) 200 big Lancichinetti–Fortunato–Radicchi communities. Our method performs better than the other ones for most cases except worse than SA methods when mixing parameter μ is very small.

Since the gradient is reduced to zero when the cost function converges, we have

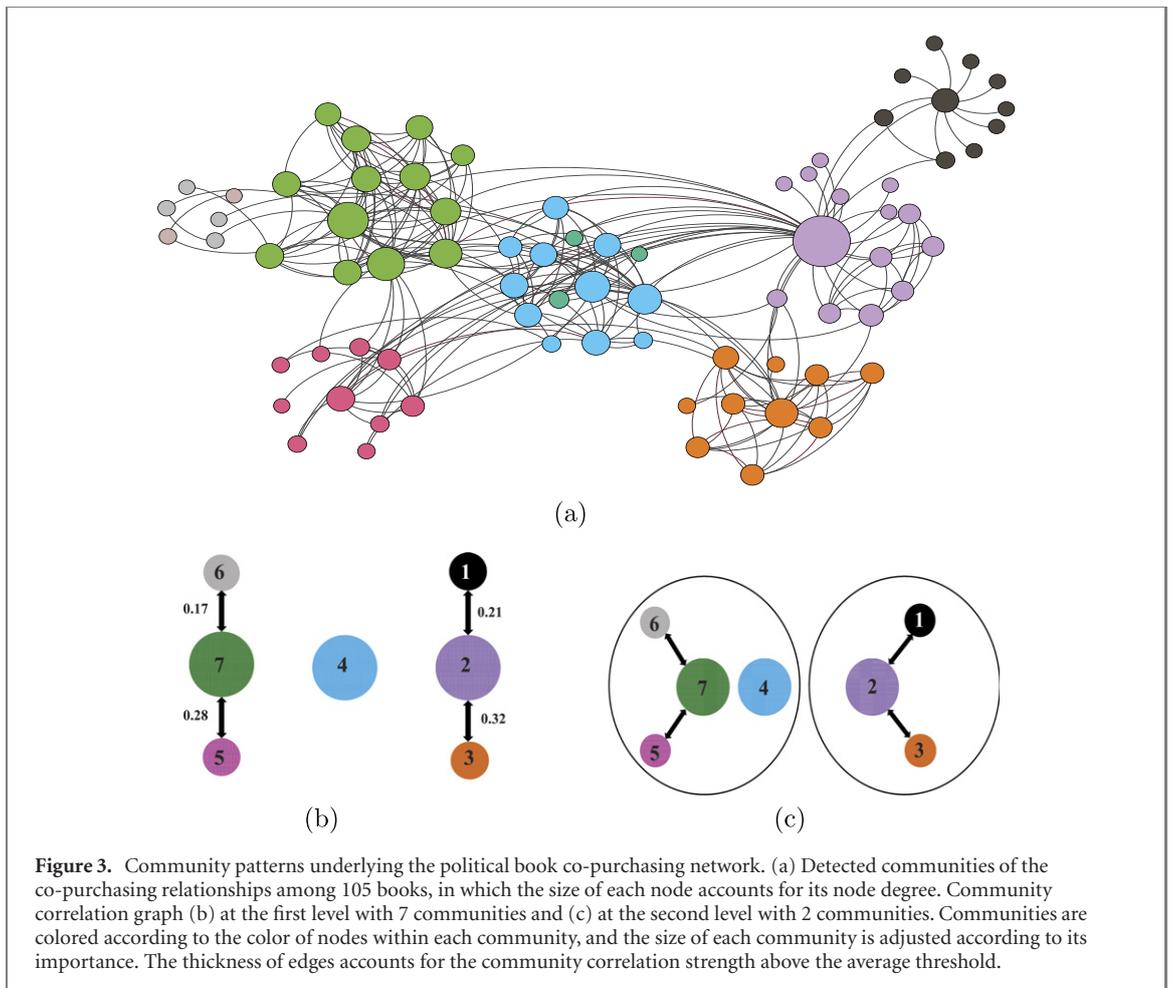
$$W_{ik}^{\text{new}} = W_{ik} \frac{a_i \nabla_{ik}^- + 1}{a_i \nabla_{ik}^+ + b_i}, \quad (18)$$

where W_{ik}^{new} denotes the updated result. The above discussions verify that $\mathcal{L}(W^{\text{new}}, \lambda) \leq G(W^{\text{new}}, W) \leq \mathcal{L}(W, \lambda)$. This indicates that our framework reduces each row of W as a probabilistic simplex, in which b_i accounts for the sum of the rows of unconstrained multiplicative results, and a_i accounts for the balance between the probabilistic simplex attraction and the gradient learning force.

3.4. Initialization and optimal number of communities

Our framework can start from any initial partition of communities. Given a randomly initialized community indicator matrix W , we provide a tiny positive perturbation (e.g. 0.05) to each entry of W before the optimization (i.e. with $\alpha = 1$). The regularized condition (i.e. with various $\alpha \neq 1$) can also be used to offer non-regularized initialization (i.e. with $\alpha = 1$). Therefore, the parameter α is only used at the initialization, with its optimal value determined by the minimum of $D(S||B)$ (i.e. the optimal partition of communities). In addition, we set the similarity matrix S as the adjacent matrix A , which is sufficient to extract the reasonable community partitions.

The number of partitioned communities r can be treated as an independent variable, whose optimal value can be calculated proactively. Specifically, we evaluate the residual $D(S||B)$ with W discretized to the community indicator matrix against each possible value of r , and then choose the optimal number of communities r that lead to the lowest residual $D(S||B)$. We provide the pseudo-code of this algorithm in the Supplementary Material.



4. Results

In this section, extensive experiments on multiple types of dataset are applied to demonstrate the high-level performance and efficiency of the proposed method.

We first consider the seminal Lancichinetti–Fortunato–Radicchi benchmark networks, which comprises a series of synthesized network models that maintain several major structural features observed in real-world networks [42, 48]. We use the following set of parameters to generate the synthesized network models: the average node degree $k = 10$, maximum node degree $k_{\max} = 50$, minimum community size $\min_c = 100$, and maximum community size $\max_c = 300$. We examine several typical community configurations by varying the network size and mixing parameter, as specified in figure 2

We apply our method and 5 different well-known algorithms to the benchmark networks and evaluate the resultant quality of detected communities using normalized mutual information [48]. The normalized mutual information computes the accuracy in detecting the true communities: the higher the normalized mutual information, the better the performance of community partition. Figure 2 demonstrates the improved performance of our method.

4.1. Case study with political book co-purchasing network

We further apply our method to the famous co-purchasing network of Amazon political book [49] to show the detection of community correlation patterns. In the political book co-purchasing network, the nodes represent the books sold by the Amazon online website, and edges represent the pairs of books frequently co-purchased. More detailed data description can be found at [49].

As in figure 3(a), our method partitions the network into a 2 level hierarchy, with seven nonparametric and two parametric communities. Figure 3(b) shows the community correlation graph at the first level, which reveals two hubs (communities 2 and 7), an outlier (community 4), and two wings (communities

Table 1. Structural properties and community partitions of six real-world large-scale networks obtained from the Stanford Network Analysis Platform [50]. The quality of community partitions is compared using the normalized mutual information (NMI), the modularity (Q) and the variation of information (VOI) with perturbation probability 0.3. Here, N and M represent the number of nodes and edges, R_g and R_m represent the number of communities obtained from the ground-truth and from our method, respectively.

Network	N	M	R_g	R_m	NMI	Q	VOI
<i>Amazon</i>	334 863	925 872	75 149	73 961	0.423	0.5353	0.106
<i>Dblp</i>	317 080	1049 866	13 477	14 531	0.431	0.5213	0.127
<i>Youtube</i>	1134 890	2987 624	8385	10 156	0.356	0.4353	0.133
<i>Livejournal</i>	3997 962	34 681 189	287 512	176 351	0.323	0.5013	0.184
<i>Orkut</i>	3072 441	117 185 083	6288 363	5963 451	0.394	0.4633	0.205
<i>Friendster</i>	65 608 366	1806 067 135	957 154	937 640	0.441	0.3253	0.321

1–3 vs communities 5–7). The two wings of books are linked by the two hub communities. The books in each wing share the same labels, so that they have a strong correlation even if they may never be purchased together. Figure 3(c) shows the two macroscopic modules at the second level, which is obtained according to the types of books. This indicates that the customers of Amazon are more likely to purchase the books attached with the same labels (about 24%), and they rarely or never purchase books attached with different labels (less than 1%).

Based on the above analysis, we can build a correlation strength matrix C , in which each entry C_{ks} calculates the coarse-grained relationship (i.e. the ratio of interconnect edges) between first-level communities k and s . The correlation strengths of each community can be ranked to guide the strategic planning for booksellers. Because of the high correlation among communities 1–3, the storage of books in these communities can be increased together if the purchasing of books in community 2 is increasing. In addition, since the community 2 is segregated from communities 5 and 7, the storage of books in the communities 5 and 7 can be reduced, because they are never co-purchased with the books in community 2.

4.2. Scalability of the algorithm

To further validate the efficiency of our method, we apply it to six large-scale real-world data-sets obtained from the Stanford Network Analysis Platform [50]: (1) the *Amazon* product network in which two products (nodes) are connected together if they are co-purchased frequently; (2) the coauthor cooperation network *Dblp*, in which two scientists are connected by an edge if they have coauthored a paper; and 4 social networks from the following sources (3) *Youtube*, (4) *Livejournal*, (5) *Orkut*, (6) *Friendster*. The ground truth partition is known for each of these networks. Table 1 presents the key structural properties and the community partition using our method for these networks. We use the modularity [1], the normalized mutual information [42] and the variation of information (VOI) to evaluate the performance. Since our results are very close to the ground-truth results (real communities) for almost all tested networks, our method is satisfactory to deal with large-scale networks with millions of nodes and edges.

5. Conclusions

In summary, we have developed a new community detection method, using nonnegative low-rank approximations. We have two major contributions: (1) an efficient definition of identifiability is introduced, which can be used to estimate the optimal number and partition of communities simultaneously; and (2) a new relaxed formulation with low-rank doubly stochastic matrix decomposition and corresponding multiplicative majorization–minimization algorithm are proposed, which allows high efficiency optimization. We use extensive in multiple types of experiments to demonstrate the near optimal performance of our method in dealing with a wide range of networks. We also show the scalability of our method using real world large-scale datasets with billions of nodes.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (Grant Nos. 71871233, 71871109, 71873039, 71573051, 11975025), the Beijing Natural Science Foundation (Grant No. 9182015), and the Slovenian Research Agency (Grant Nos. J4-9302, J1-9112, P1-0403). LW acknowledges the research computing facilities and advisory services offered by Information Technology Services of the University of Hong Kong.

ORCID iDs

Hui-Jia Li  <https://orcid.org/0000-0003-1000-717X>

Lin Wang  <https://orcid.org/0000-0002-5371-2138>

Matjaž Perc  <https://orcid.org/0000-0002-3087-541X>

References

- [1] Girvan M and Newman M E J 2002 Community structure in social and biological networks *Proc. Natl Acad. Sci. USA* **99** 7821
- [2] Pereda M and Estrada E 2019 Visualization and machine learning analysis of complex networks in hyperspherical space *Pattern Recogn.* **86** 320
- [3] De Domenico M, Granell C, Porter M A and Arenas A 2016 The physics of spreading processes in multilayer networks *Nat. Phys.* **12** 901
- [4] Bullmore E and Sporns O 2009 Complex brain networks: graph theoretical analysis of structural and functional systems *Nat. Rev. Neurosci.* **10** 186
- [5] Lynn C W and Basset D S 2019 The physics of brain network structure, function and control *Nat. Rev. Phys.* **1** 318
- [6] Estrada E 2019 Communicability geometry of multiplexes *New J. Phys.* **21** 015004
- [7] De Domenico M and Arenas A 2017 Modeling structure and resilience of the dark network *Phys. Rev. E* **95** 022313
- [8] Wang Z, Wang L, Szolnoki A and Perc M 2015 Evolutionary games on multilayer networks: a colloquium *Eur. Phys. J. B* **88** 1
- [9] Lazer D M J, Baum M A, Benkler Y, Berinsky A J, Greenhill K M et al 2018 The science of fake news *Science* **359** 1094
- [10] Orsini C, Dankulov Mitrović M, Colomer-de-Simon P, Jamakovic A, Mahadevan P et al 2015 Quantifying randomness in real networks *Nat. Commun.* **6** 8627
- [11] Tadić B, Dankulov Mitrović M and Melnik R 2017 Mechanisms of self-organized criticality in social processes of knowledge creation *Phys. Rev. E* **96** 032307
- [12] Tsvetkova M, Garcia-Gavilanes R and Yasseri T 2016 Dynamics of disagreement: large-scale temporal network analysis reveals negative interactions in online collaboration *Sci. Rep.* **6** 36333
- [13] Wang L, Li X, Zhang Y Q, Zhang Y and Zhang K 2011 Evolution of scaling emergence in large-scale spatial epidemic spreading *Plos One* **6** e21197
- [14] Wang L and Li X 2014 Spatial epidemiology of networked metapopulation: an overview *Chin. Sci. Bull.* **59** 3511
- [15] Wang J B, Wang L and Li X 2015 Identifying spatial invasion of pandemics on metapopulation networks via anatomizing arrival history *IEEE Trans. Cybern.* **46** 2782
- [16] He D, Lui R, Wang L, Tse C K, Yang L and Stone L 2015 Global spatio-temporal patterns of influenza in the post-pandemic era *Sci. Rep.* **5** 11013
- [17] Wang L and Wu J 2018 Characterizing the dynamics underlying global spread of epidemics *Nat. Commun.* **9** 218
- [18] Barthelemy M 2019 The statistical physics of cities *Nat. Rev. Phys.* **1** 406
- [19] Newman M E J 2003 The structure and function of complex networks *SIAM Rev.* **45** 167
- [20] Arenas A, Daz-Guilera A, Kurths J, Moreno Y and Zhou C 2008 Synchronization in complex networks *Phys. Rep.* **469** 93
- [21] Boccaletti S, Bianconi G, Criado R, Del Genio C I, Gomez-Gardenes J, Romance M, Sendina-Nadal I, Wang Z and Zanin M 2014 The structure and dynamics of multilayer networks *Phys. Rep.* **544** 1
- [22] De Arruda G, Rodrigues F A and Moreno Y 2018 Fundamentals of spreading processes in single and multilayer complex networks *Phys. Rep.* **756** 1
- [23] Wang W, Liu Q H, Liang J, Hu Y and Zhou T 2019 Coevolution spreading in complex networks *Phys. Rep.* **820** 1
- [24] Wang Z, Andrews M A, Wu Z X, Wang L and Bauch C T 2015 Coupled disease-behavior dynamics on complex networks: a review *Phys. Life Rev.* **15** 1
- [25] Wang Z, Bauch C T, Bhattacharyya S, d'Onofrio A, Manfredi P, Perc M, Perra N, Salathe M and Zhao D 2016 Statistical physics of vaccination *Phys. Rep.* **664** 1
- [26] Ward J K, Peretti-Watel P, Bocquier A, Seror V and Verger P 2019 Vaccine hesitancy and coercion: all eyes on France *Nat. Immunol.* **20** 1257
- [27] Zhang Y, Wang L, Zhang Y Q and Li X 2012 Towards a temporal network analysis of interactive WiFi users *Europhys. Lett.* **98** 68002
- [28] Wang Z, Liu Y, Wang L and Zhang Y 2014 Freezing period strongly impacts the emergence of a global consensus in the voter model *Sci. Rep.* **4** 3597
- [29] Zhao D, Wang L, Li S, Wang Z, Wang L and Gao B 2014 Immunization of epidemics in multiplex networks *PLoS One* **9** e112018
- [30] Wang Z, Zhao D W, Wang L, Sun G Q and Jin Z 2015 Immunity of multiplex networks via acquaintance vaccination *Europhys. Lett.* **112** 48002
- [31] Fortunato S and Hric D 2016 Community detection in networks: a user guide *Phys. Rep.* **659** 1
- [32] Xia C, Li X, Wang Z and Perc M 2018 Doubly effects of information sharing on interdependent network reciprocity *New J. Phys.* **20** 075005
- [33] Battiston F, Perc M and Latora V 2017 Determinants of public cooperation in multiplex networks *New J. Phys.* **19** 073017
- [34] Perc M, Jordan J J, Rand D G, Wang Z, Boccaletti S and Szolnoki A 2017 Statistical physics of human cooperation *Phys. Rep.* **687** 1
- [35] Liu Q H, Ajelli M, Aleta A, Merler S, Moreno Y and Vespignani A 2018 Measurability of the epidemic reproduction number in data-driven contact networks *Proc. Natl Acad. Sci. USA* **50** 12680
- [36] De Arruda G, Cozzo E, Peixoto T P, Rodrigues F A and Moreno Y 2017 Disease localization in multilayer networks *Phys. Rev. X* **7** 011014
- [37] Battiston F, Perc M and Latora V 2017 Determinants of public cooperation in multiplex networks *New J. Phys.* **19** 073017
- [38] Tejedor A, Longjas A, Fofoula-Georgiou E, Georgiou T T and Moreno Y 2018 Diffusion dynamics and optimal coupling in multiplex networks with directed layers *Phys. Rev. X* **8** 031071
- [39] Lee D D and Seung H S 1999 Learning the parts of objects by non-negative matrix factorization *Nature* **401** 788
- [40] Ding C, He X and Simon H 2005 On the equivalence of nonnegative matrix factorization and spectral clustering *Proc. of 2005 SIAM Int. Conf. on Data Mining* (Philadelphia, PA: Society for Industrial and Applied Mathematics)
- [41] Ding C H Q, Li T and Jordan M I 2010 Convex and semi-nonnegative matrix factorizations *IEEE Trans. Pattern Anal.* **32** 45

- [42] Lancichinetti A and Fortunato S 2009 Community detection algorithms: a comparative analysis *Phys. Rev. E* **80** 056117
- [43] Shi J and Malik J 2000 Normalized cuts and image segmentation *IEEE Trans. Pattern Anal.* **22** 888
- [44] Aloise D, Deshpande A, Hansen P and Popat P 2009 NP-hardness of Euclidean sum-of-squares clustering *Mach. Learn.* **75** 245
- [45] Mahajan M, Nimbhorkar P and Varadarajan K 2009 *The Planar k-Means Problem is NP-Hard* (Lecture Notes in Computer Science vol 5431) (Berlin: Springer) p 274
- [46] Hofmann T 1999 Probabilistic latent semantic analysis *Proc. of the 15th Conf. on Uncertainty in Artificial Intelligence* (San Mateo, CA: Morgan Kaufmann Publishers)
- [47] Hunter D and Lange K 2004 A tutorial on MM algorithms *Am. Stat.* **58** 30
- [48] Lancichinetti A, Fortunato S and Radicchi F 2008 Benchmark graphs for testing community detection algorithms *Phys. Rev. E* **78** 046110
- [49] Fortunato S 2010 Community detection in graphs *Phys. Rep.* **486** 75
- [50] Leskovec J and Krevl A 2014 SNAP Datasets (Stanford: Stanford University) <http://snap.stanford.edu/data/index.html>