


Article

Estimating the Number of Communities in Weighted Networks

Huan Qing 

School of Mathematics, China University of Mining and Technology, Xuzhou 221116, China;
qinghuan@cumt.edu.cn or qinghuan07131995@163.com

Abstract: Community detection in weighted networks has been a popular topic in recent years. However, while there exist several flexible methods for estimating communities in weighted networks, these methods usually assume that the number of communities is known. It is usually unclear how to determine the exact number of communities one should use. Here, to estimate the number of communities for weighted networks generated from arbitrary distribution under the degree-corrected distribution-free model, we propose one approach that combines weighted modularity with spectral clustering. This approach allows a weighted network to have negative edge weights and it also works for signed networks. We compare the proposed method to several existing methods and show that our method is more accurate for estimating the number of communities both numerically and empirically.

Keywords: community detection; degree-corrected distribution-free model; weighted modularity; network analysis

1. Introduction

For decades, network science provided substantial quantitative tools for the study of complex systems [1–4]. Networks emerge in numerous fields including physics, sociology, biology, economics, and so forth [5–15]. The elementary parts of a network are nodes, links, and link weights. A network is unweighted when all link weights are 1 and weighted otherwise [16]. Networks usually have community structure such that nodes within the same community have more connections than across communities [17,18]. For example, in social networks, communities can be groups of students who belong to the same school, be of the same club, be of the same graduation year, or be interested in the same movie; in scientific collaboration networks, communities are scientists in the same field [19–21]; in protein-protein interaction networks, communities are proteins enjoying similar functions [22,23]. However, in practice, the latent community structure of a network is generally not directly observable and we need to develop techniques to infer community structure.

Community detection for unweighted networks has been widely studied for decades [17,18]. Numerous community detection methods have been developed to fit a statistical model that can generate a random network with a community structure. The stochastic blockmodels (SBM) [24] is a classical and popular generative model for unweighted networks. The popular degree-corrected stochastic blockmodels (DCSBM) extends SBM by considering node heterogeneity. Based on SBM and DCSBM, substantial community detection methods have been developed, such as [25–36]. However, most methods require the number of communities K to be known in advance, and this is often not the case for real-world unweighted networks. To address this problem, some methods have been developed to estimate K under SBM or DCSBM [37–47], where approaches developed in [46] stand out as they estimate K for unweighted networks regardless of statistical models.

A significant drawback of the above SBM-based and DCSBM-based methods is that they ignore the impact of edge weights which are common in network data and could help us to understand the community structure of a network better [16]. Recently, community



Citation: Qing, H. Estimating the Number of Communities in Weighted Networks. *Entropy* **2023**, *25*, 551.
<https://doi.org/10.3390/e25040551>

Academic Editor: Adam Lipowski

Received: 7 February 2023

Revised: 13 March 2023

Accepted: 22 March 2023

Published: 23 March 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

detection in weighted networks has become a hot topic and many statistical models have been developed to fit weighted networks, such as the weighted stochastic blockmodels (WSBM) proposed in [48–54], the distribution-free model (DFM) of [55], and the degree-corrected distribution-free model (DCDFM) introduced in [56]. Among these models, DFM and its extension DCDFM stand out as they allow edge weights to follow any distribution as long as the expected adjacency matrix follows a block structure related to community partition. However, similar to SBM-based and DCSBM-based methods, algorithms developed for the above models also assume that K is known in advance, which is usually impractical for real-world weighted networks. To close this gap, we provide a simple approach to estimate K for weighted networks generated from DCDFM.

The main contributions of this work include:

(1) We propose a method by taking advantage of both spectral clustering and weighted modularity to estimate the number of communities for weighted networks generated from arbitrary distribution under DCDFM. The method determines K by increasing the number of communities until weighted modularity does not increase. The method is devised for DCDFM, but it can be naturally applied to weighted networks generated from DFM and unweighted networks generated from SBM and DCSBM since these three models are sub-models of DCDFM.

(2) We conduct a large number of experiments on both computer-generated weighted networks and real-world networks including signed networks. The experimental results show that our method can estimate the number of communities for weighted networks generated by different distributions under DCDFM even when the true K is 1 and it is more accurate than its competitors.

2. Methodology

2.1. The Degree-Corrected Distribution-Free Model

In this article, we work with the degree-corrected distribution-free model proposed in [56]. We assume that there exist K perceivable non-overlapping clusters $\mathcal{C}^{(1)}, \mathcal{C}^{(2)}, \dots, \mathcal{C}^{(K)}$, and each node only belongs to exactly one cluster. Let the $n \times 1$ vector ℓ denote the node label such that ℓ_i takes value from $\{1, 2, \dots, K\}$ and ℓ_i is the community label for node i for $i \in [n]$. Let $Z \in \{0, 1\}^{n \times K}$ be the community membership matrix such that $Z_{ik} = 1$ if $\ell_i = k$ and $Z_{ik} = 0$ otherwise. Let θ be an $n \times 1$ vector such that the positive number θ_i is the node heterogeneity of node i . Let Θ be an $n \times n$ diagonal matrix whose i -th diagonal entry is θ_i . Let P be the $K \times K$ symmetric connectivity matrix such that P 's rank is K , P 's elements can be any real values in $[-1, 1]$, and $\max_{k,l \in [K]} |P_{kl}| = 1$, where we let P 's maximum absolute element be 1 for convenience since we consider the node heterogeneity parameter θ . For $i, j \in [n]$, the DCDFM model [56] generates the (i, j) -th element of the symmetric adjacency matrix A for an un-directed weighted network \mathcal{N} in the following way:

$$A_{ij} \text{ is a random variable generated from arbitrary distribution } \mathcal{F} \text{ with expectation } \Omega_{ij}, \quad (1)$$

where Ω is defined as $\Omega = \Theta Z P Z' \Theta$.

DCDFM includes several previous models. For example, when $\theta_i = \sqrt{\rho}$ for all $i \in [n]$, DCDFM reduces to the distribution-free model [55]; when \mathcal{F} is Bernoulli distribution and P 's elements are non-negative, DCDFM reduces to the classical degree-corrected stochastic blockmodels [57]; when \mathcal{F} is Bernoulli distribution, all elements of θ are the same, and P 's elements are non-negative, DCDFM reduces to the popular stochastic blockmodels [24], i.e., SBM, DCSBM, and DFM are sub-models of DCDFM. As analyzed in [56], \mathcal{F} can be any distribution as long as A 's expectation matrix is Ω under distribution \mathcal{F} . Meanwhile, the fact that whether P 's elements can be negative depends on distribution \mathcal{F} . For example, when \mathcal{F} is Bernoulli, Binomial, Poisson, Geometric or Exponential distributions, P 's elements should be non-negative or positive; when \mathcal{F} is Normal, Laplace or A is the adjacency matrix of a signed network, P 's elements can be negative. DCDFM can generate A for weighted networks benefiting from the arbitrariness of distribution \mathcal{F} .

When n, K, ℓ, P , and θ are set, we can generate the adjacency matrix A for any distribution \mathcal{F} under DCDFM as long as Equation (1) holds. Given A and the known number of clusters K , ref. [56] designs an efficient spectral algorithm called nDFA to estimate the node label vector ℓ and shows that nDFA enjoys consistent estimation under DCDFM for any distribution \mathcal{F} satisfying Equation (1). However, the method nDFA requires K to be known in advance, and this is not the case in practice. To process this problem, in this article, we aim at developing an efficient method to estimate the number of communities K when only the adjacency matrix A is known, where A is generated from DCDFM with K communities for arbitrary distribution \mathcal{F} satisfying Equation (1).

2.2. Estimation of the Number of Communities

Our method for estimating K is closely related to the modularity for signed networks introduced in [58] and this modularity extends the popular Newman-Girvan modularity matrix [59] from unweighted networks to signed networks. Instead of simply considering signed networks, we extend the modularity developed in [58] to weighted networks with A 's elements being any finite real values by considering indicator functions. We let the $n \times n$ symmetric adjacency matrix A be generated from DCDFM for arbitrary distribution \mathcal{F} satisfying Equation (1), so we have $A \in \mathbb{R}^{n \times n}$. Let $A^+, A^- \in \mathbb{R}_{\geq 0}^{n \times n}$ such that $A_{ij} = A_{ij}^+ - A_{ij}^-$, where $A_{ij}^+ = \max(0, A_{ij})$ and $A_{ij}^- = \max(0, -A_{ij})$ for any $i, j \in [n]$. Let d^+ be the positive degree vector with i -th entry $d_i^+ = \sum_{j=1}^n A_{ij}^+$ and d^- be the negative vector with i -th entry $d_i^- = \sum_{j=1}^n A_{ij}^-$ for $i \in [n]$. Let $m^+ = \sum_{i=1}^n d_i^+ / 2$ and $m^- = \sum_{i=1}^n d_i^- / 2$. Let $\hat{\ell}$ be a $n \times 1$ node label vector returned by running a community detection method \mathcal{M} on A with k communities such that $\hat{\ell}_i$ takes value from $\{1, 2, \dots, k\}$. Based on the community partition $\hat{\ell}$ obtained from the method \mathcal{M} , the positive modularity Q^+ and the negative modularity Q^- are defined as

$$Q^+ = \frac{1}{2m^+} \sum_{i=1}^n \sum_{j=1}^n (A_{ij}^+ - \frac{d_i^+ d_j^+}{2m^+}) \delta(\hat{\ell}_i, \hat{\ell}_j) 1_{m^+ > 0}, \quad Q^- = \frac{1}{2m^-} \sum_{i=1}^n \sum_{j=1}^n (A_{ij}^- - \frac{d_i^- d_j^-}{2m^-}) \delta(\hat{\ell}_i, \hat{\ell}_j) 1_{m^- > 0},$$

where $\delta(\hat{\ell}_i, \hat{\ell}_j)$ is the Kronecker delta function, $1_{m^+ > 0}$ and $1_{m^- > 0}$ are indicator functions such that

$$\delta(\hat{\ell}_i, \hat{\ell}_j) = \begin{cases} 1 & \text{when } \hat{\ell}_i = \hat{\ell}_j, \\ 0 & \text{otherwise,} \end{cases}, \quad 1_{m^+ > 0} = \begin{cases} 1 & \text{when } m^+ > 0, \\ 0 & \text{otherwise,} \end{cases}, \quad 1_{m^- > 0} = \begin{cases} 1 & \text{when } m^- > 0, \\ 0 & \text{otherwise,} \end{cases},$$

The weighted modularity considered in this article is defined as

$$Q_{\mathcal{M}}(k) = \frac{2m^+}{2m^+ + 2m^-} Q^+ - \frac{2m^-}{2m^+ + 2m^-} Q^-. \quad (2)$$

When all edge weights are non-negative such that $m^- = 0$, the weighted modularity reduces to the Newman-Girvan modularity. When A has both positive and negative entries, the weighted modularity reduces to the modularity introduced in [58]. The weighted modularity obtained via Equation (2) measures the quality of community partition for a weighted network whose adjacency matrix has any finite real elements, and it is more general than the modularity introduced in [58]. Similar to the Newman-Girvan modularity, a larger weighted modularity $Q_{\mathcal{M}}(k)$ indicates a better community partition.

In Equation (2), we write the weighted modularity as a function of the number of communities k and the community detection method \mathcal{M} to emphasize that the weighted modularity may be different for different k or different community detection methods. We estimate the number of communities K by increasing k until the weighted modularity function in Equation (2) does not increase. Suppose there is a cardinality choice of K such that K locates in $\{1, 2, \dots, K_0\}$. For a community detection algorithm \mathcal{M} , our strategy for estimating K is

$$\hat{K}_{\mathcal{M}} = \arg \max_{k \in [K_0]} Q_{\mathcal{M}}(k). \quad (3)$$

In this paper, to estimate the number of communities for weighted networks generated from DCDFM, we choose the method \mathcal{M} as the nDFA algorithm designed in [56] because nDFA enjoys consistent estimation of community memberships under DCDFM and it is computationally fast. For convenience, when \mathcal{M} is the nDFA algorithm, we call our method for estimating K via Equation (3) as nDFAwm, where “wm” means weighted modularity. The details of the nDFA algorithm [56] are written below.

Input: A, k . Output: $\hat{\ell}$.

- Let $\tilde{A} = \hat{U} \hat{\Lambda} \hat{U}'$ be the top- k eigendecomposition of A .
- Let the $n \times k$ matrix \hat{U}_* be the row normalization of \hat{U} such that $\hat{U}_*(i, :) = \frac{\hat{U}(i, :)}{\|\hat{U}(i, :)\|_F}$ for $i \in [n]$.
- Apply k-means algorithm on all rows of \hat{U}_* with k clusters to obtain $\hat{\ell}$.

3. Experimental Results

In this section, we present both simulation results and real-world experiments to compare our nDFAwm with three model-free methods in the literature for estimating the number of communities: the modularity eigengap (ME for short) method proposed in [60], the non-backtracking (NB) method designed in [46], and the Bethe Hessian matrix-based method BHac developed in [46].

3.1. Simulations

In this section, we investigate the performance of nDFAwm and competing algorithms to adjacency matrices generated from nine distributions under DCDFM. For each parameter setting, we report the accuracy rate over 100 repetitions for each method, where the accuracy rate is the fraction of times that the estimated number of clusters \hat{K} equals the true number of clusters K .

To generate simulated weighted networks from DCDFM, first, we need to define n, K, θ, Z , and P . For n , unless specified, we let $n = 50K$. For Z , we let each node belong to one of the K clusters with equal probability, i.e., there are around 50 nodes in each cluster. For θ , unless specified, we let $\theta_i = \text{rand}(1) \sqrt{\rho}$, where the positive number ρ controls network sparsity and $\text{rand}(1)$ is a random number drawn from the uniform distribution in the interval $(0, 1)$. We set n, K, P , and ρ independently for each simulation. After setting these model parameters, we generate A under DCDFM for several distributions \mathcal{F} satisfying Equation (1). For our nDFAwm, we set $K_c = 20$ since the largest K in our simulations is six. In this paper, we consider Bernoulli, binomial, Poisson, geometrical, exponential, normal, laplace, and uniform distributions, where details on probability mass function or probability density function of these distributions can be found in <http://www.stat.rice.edu/~dobelman/courses/texts/distributions.c&b.pdf> (accessed on 9 November 2022). Meanwhile, we also consider the signed network case in our simulation studies.

3.1.1. Bernoulli Distribution

When \mathcal{F} is Bernoulli distribution such that $A_{ij} \sim \text{Bernoulli}(\Omega_{ij})$, i.e., $A_{ij} \in \{0, 1\}$ for $i, j \in [n]$ and DCDFM reduces to DCSBM for this case. By the property of Bernoulli distribution, $\mathbb{E}[A_{ij}] = \Omega_{ij}$ satisfies Equation (1) and Ω_{ij} is a probability ranging in $[0, 1]$. So, ρ 's range is $(0, 1]$, and all elements of P should be non-negative. For Bernoulli distribution, we consider the following simulations.

Experiment 1 (a): changing ρ . Let $K = 3$ and P be

$$P = \begin{bmatrix} 1 & 0.2 & 0.3 \\ 0.2 & 0.8 & 0.2 \\ 0.3 & 0.2 & 0.9 \end{bmatrix}.$$

Let ρ range in $\{0.2, 0.3, \dots, 1\}$.

Experiment 1 (b): changing K . Let P 's diagonal entries be 1 and off-diagonal entries be 0.2. Let $\rho = 0.9$ and K range in $\{2, 3, \dots, 6\}$.

Experiment 1 (c): changing ρ when $K = 1$. Let $K = 1, P = 1$, and ρ range in $\{0.1, 0.2, \dots, 1\}$.

Experiment 1 (d): connectivity across communities. Let $K = 2, \rho = 1$, P 's diagonal entries be 1, P 's off-diagonal entries be β , and β range in $\{0.1, 0.2, \dots, 0.8\}$.

Figure 1 shows the accuracy rate of Experiment 1. Panel (a) of Figure 1 shows that as the network becomes denser, all methods provide more accurate estimations of the number of clusters. For Experiment 1 (a), all methods perform similarly. For Experiment 1 (b), from panel (b) of Figure 1, we see that our nDFAwm performs the best. From panel (c) of Figure 1, we see that our nDFAwm performs poorer than NB and BHac while ME fails to work. Meanwhile, except ME, all methods perform better as the network becomes denser for Experiment 1 (c). From panel (d) of Figure 1, we see that all methods perform poorer as the off-diagonal entries of P are closer to the diagonal entries and our nDFAwm performs slightly poorer than ME while it outperforms NB and BHac.

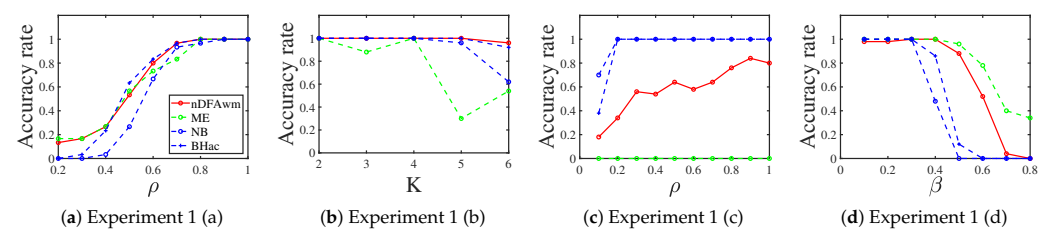


Figure 1. Bernoulli distribution.

3.1.2. Binomial Distribution

When \mathcal{F} is binomial distribution such that $A_{ij} \sim \text{Binomial}(m, \frac{\Omega_{ij}}{m})$ for any positive integer m , i.e., $A_{ij} \in \{0, 1, 2, \dots, m\}$ for $i, j \in [n]$. By the property of binomial distribution, $\mathbb{E}[A_{ij}] = \Omega_{ij}$ satisfies Equation (1) and $\frac{\Omega_{ij}}{m}$ is a probability ranging in $[0, 1]$. So, ρ 's range is $(0, m]$ and all elements of P should be non-negative.

Experiment 2 (a): changing ρ . Let $K = 3, m = 5$, and P be the same as that of Experiment 1 (a). Let ρ range in $\{0.5, 1, \dots, 5\}$.

Experiment 2 (b): changing K . Let P be the same as Experiment 1 (b), $\rho = 2$, $m = 5$, and K range in $\{2, 3, \dots, 6\}$.

Experiment 2 (c): changing ρ when $K = 1$. Let $K = 1, P = 1, m = 5$, and ρ range in $\{0.5, 1, \dots, 5\}$.

Experiment 2 (d): connectivity across communities. Let $K = 2, \rho = 1, m = 5$, and P be the same as Experiment 1 (d).

Figure 2 shows the accuracy rate of Experiment 2. For Experiments 2 (a), 2 (b), and 2 (c), the results are similar to that of Experiments 1 (a), 1 (b), and 1 (c), respectively, and we omit the analysis here. For Experiment 2 (d), panel (d) of Figure 2 says that our nDFAwm perform similarly to NB and BHac while ME performs best.

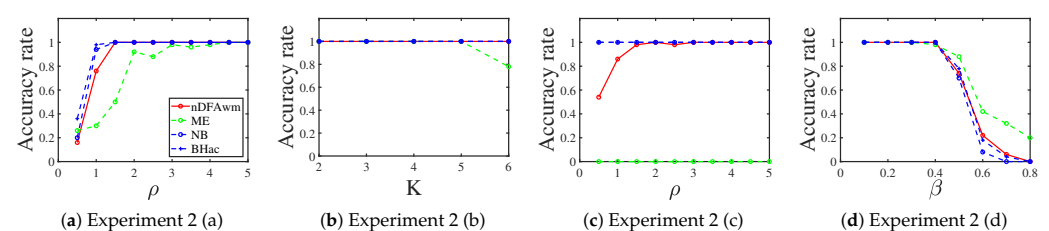


Figure 2. Binomial distribution.

3.1.3. Poisson Distribution

When \mathcal{F} is Poisson distribution such that $A_{ij} \sim \text{Poisson}(\Omega_{ij})$, i.e., A_{ij} is a non-negative integer for $i, j \in [n]$. By the property of Poisson distribution, $\mathbb{E}[A_{ij}] = \Omega_{ij}$ satisfies Equation (1) and Ω_{ij} is non-negative. So, ρ 's range is $(0, +\infty)$ and all elements of P should be non-negative.

Experiment 3 (a): changing ρ . Let $K = 3$ and P be the same as that of Experiment 1 (a). Let ρ range in $\{0.5, 1, \dots, 5\}$.

Experiment 3 (b): changing K . Let P be the same as Experiment 1 (b), $\rho = 2$, and K range in $\{2, 3, \dots, 6\}$.

Experiment 3 (c): changing ρ when $K = 1$. Let $K = 1, P = 1$, and ρ range in $\{0.5, 1, \dots, 5\}$.

Experiment 3 (d): connectivity across communities. Let $K = 2, \rho = 2$, and P be the same as Experiment 1 (d).

Figure 3 shows the accuracy rate of Experiment 3. The results are similar to that of Experiment 2, and we omit the analysis here.

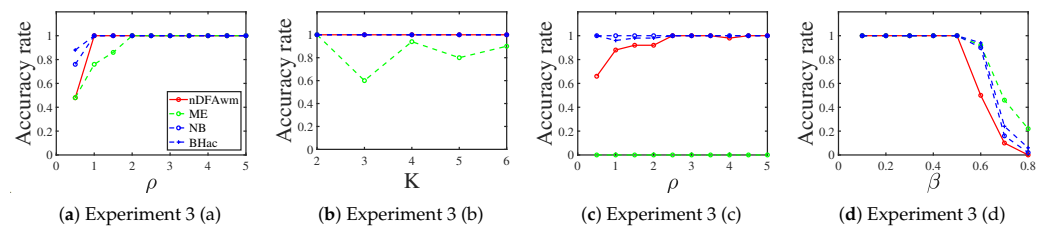


Figure 3. Poisson distribution.

3.1.4. Geometric Distribution

When \mathcal{F} is a geometric distribution such that $A_{ij} \sim \text{Geometric}(\frac{1}{\Omega_{ij}})$, i.e., A_{ij} is positive integer for $i, j \in [n]$. For geometric distribution, since $\mathbb{P}(A_{ij} = m) = \frac{1}{\Omega_{ij}}(1 - \frac{1}{\Omega_{ij}})^{m-1}$ for $m = 1, 2, \dots$, and $0 < \frac{1}{\Omega_{ij}} \leq 1$, all elements of P must be positive. By the property of geometric distribution, we have $\mathbb{E}[A_{ij}] = \Omega_{ij}$ satisfying Equation (1). For convenience, we let $\theta_i = \sqrt{\rho}$ for $i \in [n]$ to make DCDFM reduce to DFM for this case. Then, we have $\Omega = \rho Z P Z'$. Since $\Omega_{ij} \geq 1$ for $i, j \in [n]$, we have $\rho \min_{k,l \in [K]} P_{kl} \geq 1$.

Experiment 4 (a): changing ρ . Let $K = 3$ and P be the same as that of Experiment 1 (a). Let ρ range in $\{5, 6, \dots, 15\}$.

Experiment 4 (b): changing K . Let P be the same as Experiment 1 (b), $\rho = 10$, and K range in $\{2, 3, \dots, 6\}$.

Experiment 4 (c): changing ρ when $K = 1$. Let $K = 1, P = 1$, and ρ range in $\{2, 4, \dots, 20\}$.

Experiment 4 (d): connectivity across communities. Let $K = 2, \rho = 10$, and P be the same as Experiment 1 (d).

Figure 4 shows the accuracy rate of Experiment 4. Unlike Experiments 1–3, the numerical results of Experiment 4 say that our nDFAwm successfully estimates the number of communities for all cases while NB and BHac fail to work when the network is generated from geometric distribution under the DCDFM model. For the method ME, it fails to work when the true K is 1 and it performs similarly to our nDFAwm for other cases.

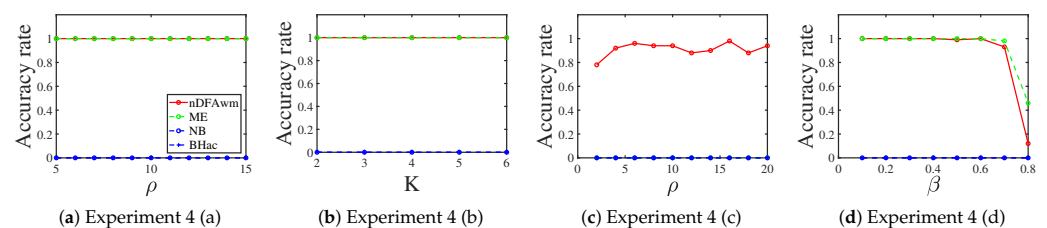


Figure 4. Geometric distribution.

3.1.5. Exponential Distribution

When \mathcal{F} is a exponential distribution such that $A_{ij} \sim \text{Exponential}(\frac{1}{\Omega_{ij}})$, i.e., $A_{ij} \in \mathbb{R}_+$ for $i, j \in [n]$. For exponential distribution, since $\frac{1}{\Omega_{ij}} > 0$, all elements of P must be positive and ρ range in $(0, +\infty)$. By the property of exponential distribution, $\mathbb{E}[A_{ij}] = \Omega_{ij}$ satisfies Equation (1).

Experiment 5 (a): changing ρ . Let $K = 3$ and P be the same as that of Experiment 1 (a). Let ρ range in $\{1, 2, \dots, 10\}$.

Experiment 5 (b): changing K . Let P be the same as Experiment 1 (b), $\rho = 5$, and K range in $\{2, 3, \dots, 6\}$.

Experiment 5 (c): changing ρ when $K = 1$. Let $K = 1, P = 1$, and ρ range in $\{1, 2, \dots, 10\}$.

Experiment 5 (d): connectivity across communities. Let $K = 2, \rho = 5$, and P be the same as Experiment 1 (d).

Figure 5 shows the accuracy rate of Experiment 5. In general, we see that our nDFAwm estimates K more accurately than its competitors except Experiment 5 (d) where ME performs slightly better than our nDFAwm. From panels (a) and (c) of Figure 5, it is interesting to find that NB and BHac perform poorer as ρ increases. Panels (b) and (d) of Figure 5 say that NB and BHac fail to work for Experiments 5 (b) and 5 (d).

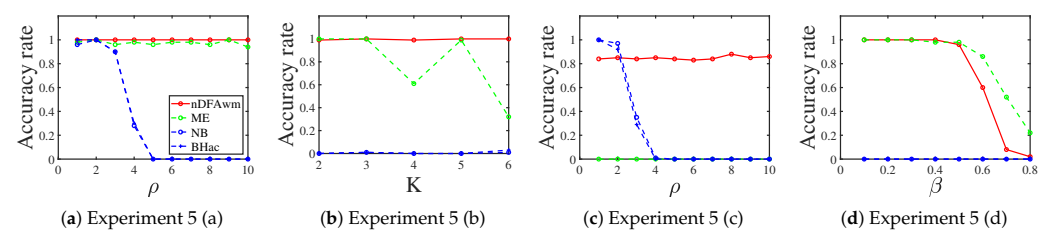


Figure 5. Exponential distribution.

3.1.6. Normal Distribution

When \mathcal{F} is normal distribution such that $A_{ij} \sim \text{Normal}(\Omega_{ij}, \sigma^2)$, i.e., $A_{ij} \in \mathbb{R}$ for $i, j \in [n]$, where $\Omega(i, j), \sigma^2$ are the expectation and variance terms of normal distribution, respectively. By the property of normal distribution, $\mathbb{E}[A_{ij}] = \Omega_{ij}$ satisfies Equation (1) and all entries of P are real values. So, ρ 's range is $(0, +\infty)$ and P 's elements can be negative.

Experiment 6 (a): changing ρ . Let $K = 3, \sigma^2 = 1$, and P be

$$P = \begin{bmatrix} 1 & -0.2 & -0.3 \\ -0.2 & 0.8 & 0.2 \\ -0.3 & 0.2 & 0.9 \end{bmatrix}.$$

Let ρ range in $\{1, 2, \dots, 10\}$.

Experiment 6 (b): changing K . Let P be the same as Experiment 1 (b), $\sigma^2 = 1$, $\rho = 3$, and K range in $\{2, 3, \dots, 6\}$.

Experiment 6 (c): changing ρ when $K = 1$. Let $K = 1, \sigma^2 = 1, P = 1$, and ρ range in $\{0.5, 1, \dots, 10\}$.

Experiment 6 (d): connectivity across communities. Let $K = 2, \sigma^2 = 1, \rho = 2$, P 's diagonal entries be 1, P 's off-diagonal entries be β , and β range in $\{-0.5, -0.4, \dots, 0.9\}$.

Figure 6 shows the accuracy rate of Experiment 6. In general, we see that our nDFAwm outperforms its competitors except for Experiment 6 (d) where it performs similarly to ME. From panels (a), (b), and (d) of Figure 6, we see that NB and BHac fail to work. Panel (c) of Figure 6 says that though NB and BHac perform poorer than our nDFAwm, they provide more accurate estimations as ρ increases for Experiment 6 (c).

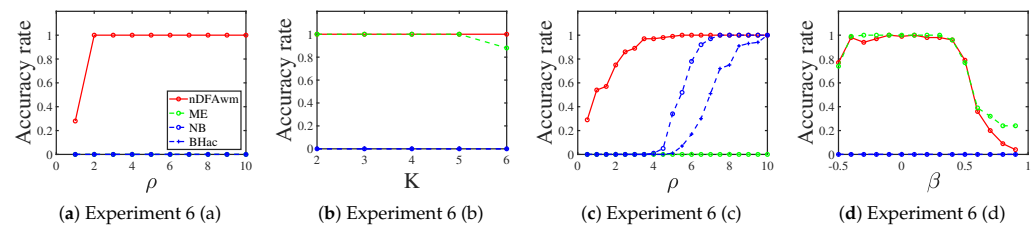


Figure 6. Normal distribution.

3.1.7. Laplace Distribution

When \mathcal{F} is laplace distribution such that $A_{ij} \sim \text{Laplace}(\Omega_{ij}, \frac{\sigma^2}{2})$, i.e., $A_{ij} \in \mathbb{R}$ for $i, j \in [n]$, where $\Omega(i, j), \sigma^2$ are the expectation and variance terms of laplace distribution, respectively. Similar to normal distribution, $\mathbb{E}[A_{ij}] = \Omega_{ij}$ satisfies Equation (1), all elements of P are real values, and ρ 's range is $(0, +\infty)$.

Experiment 7 (a): changing ρ . Let $K = 3, \sigma^2 = 1$, P be the same as Experiment 6 (a), and ρ range in $\{1, 2, \dots, 10\}$.

Experiment 7 (b): changing K . Let P be the same as Experiment 1 (b), $\sigma^2 = 1$, $\rho = 3$, and K range in $\{2, 3, \dots, 6\}$.

Experiment 7 (c): changing ρ when $K = 1$. Let $K = 1, \sigma^2 = 1, P = 1$, and ρ range in $\{0.5, 1, \dots, 10\}$.

Experiment 7 (d): connectivity across communities. Let $K = 2, \sigma^2 = 1, \rho = 2$, P 's diagonal entries be 1, P 's off-diagonal entries be β , and β range in $\{-0.5, -0.4, \dots, 0.9\}$.

Figure 7 displays the accuracy rate of Experiment 7. The numerical results are similar to that of Experiment 6 and we omit the analysis here.

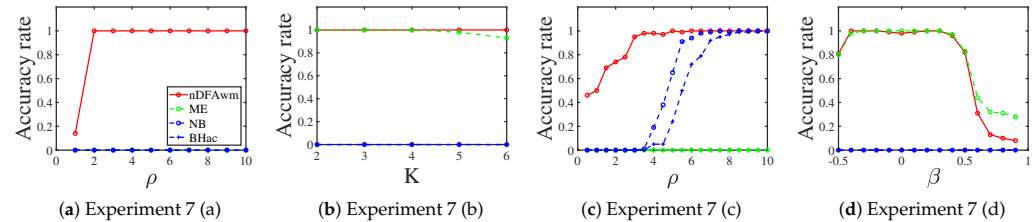


Figure 7. Laplace distribution.

3.1.8. Uniform Distribution

When \mathcal{F} is uniform distribution such that $A_{ij} \sim \text{Uniform}(0, \Omega_{ij})$. For this case, $\mathbb{E}[A_{ij}] = \Omega_{ij}$ satisfies Equation (1), all elements of P are non-negative, and ρ 's range is $(0, +\infty)$ because $A_{ij} \in (0, \max_{i,j \in [n]} \Omega_{ij})$ and it has no limitation on ρ as long as ρ is positive.

Experiment 8 (a): changing ρ . Let $K = 3$, P be the same as Experiment 1 (a), and ρ range in $\{2, 4, \dots, 20\}$.

Experiment 8 (b): changing K . Let P be the same as Experiment 1 (b), $\rho = 0.3$, and K range in $\{2, 3, \dots, 6\}$.

Experiment 8 (c): changing ρ when $K = 1$. Let $K = 1, P = 1$, and ρ range in $\{2, 4, \dots, 20\}$.

Experiment 8 (d): connectivity across communities. Let $K = 2, \rho = 1$, and P be the same as Experiment 1 (d).

Figure 8 displays the accuracy rate of Experiment 8. We see that our approach nDFAwm outperforms its competitors in all cases except for Experiment 8 (d) where it performs slightly poorer than ME. For ME method, it enjoys similar performances as our nDFAwm for Experiments 8 (a), 8 (b), and 8 (d) while it fails to estimate the number of clusters when the true K is 1. For NB and BHac, they perform poorer as ρ increases for Experiments 8 (a), 8 (c), and 8 (d). Meanwhile, NB and BHac fail to work for Experiment 8 (b).

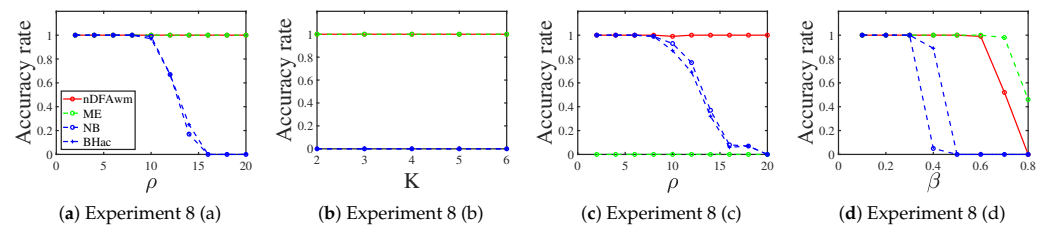


Figure 8. Uniform distribution.

3.1.9. Signed Networks

Let $\mathbb{P}(A_{ij} = 1) = \frac{1+\Omega_{ij}}{2}$ and $\mathbb{P}(A_{ij} = -1) = \frac{1-\Omega_{ij}}{2}$ such that A is the adjacency matrix of a signed network. For this case, $\mathbb{E}[A_{ij}] = \Omega_{ij}$ satisfies Equation (1), all elements of P are real values, and ρ 's range is $(0, 1]$. For signed networks, we let $n = 100K$, each node belong to one of the K communities with equal probability, and $\theta_i = \sqrt{\rho}$ for $i \in [n]$.

Experiment 9 (a): changing ρ . Let $K = 3$, P be the same as Experiment 6 (a), and ρ range in $\{0.1, 0.2, \dots, 1\}$.

Experiment 9 (b): changing K . Let P be the same as Experiment 1 (b), $\rho = 0.5$, and K range in $\{2, 3, \dots, 6\}$.

Experiment 9 (c): changing ρ when $K = 1$. Let $K = 1$, $P = 1$, and ρ range in $\{0.1, 0.2, \dots, 1\}$.

Experiment 9 (d): connectivity across communities. Let $K = 2$, $\rho = 0.5$, P 's diagonal entries be 1, P 's off-diagonal entries be β , and β range in $\{-0.5, -0.4, \dots, 0.9\}$.

Figure 9 displays the accuracy rate of Experiment 9. We see that our approach nDFAwm provides a more accurate estimation of the number of clusters than its competitors except Experiment 9 (d) where it performs similarly to ME. For ME, it fails to work in Experiments 9 (a) and 9 (c). For NB and BHac, they fail to estimate K except for Experiment 9 (c) where they have better estimations as ρ increases.

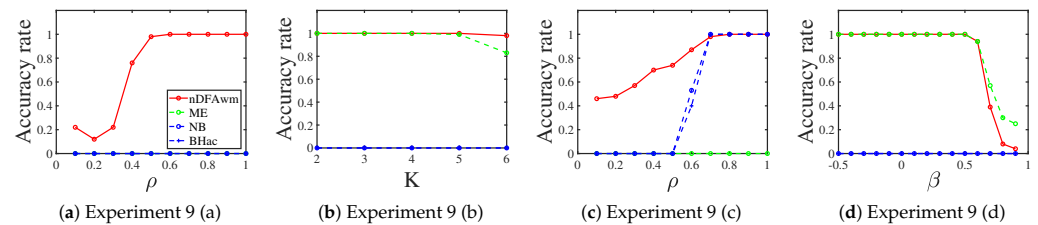


Figure 9. Signed networks.

3.2. Real-World Networks

For real-world networks, we consider eight data sets in Table 1. The ground truth numbers of communities of these eight networks are known and they provide a reasonable baseline to compare estimators. The Karate club (weighted) network is a weighted network with non-negative edge weights, the Gahuku-Gama subtribes is a signed network, the Slovene Parliamentary Party network is a weighted network with positive and negative edge weights, and the other five data sets are unweighted. For visualization, Figure 10 displays adjacency matrices of weighted networks considered in this paper. The Karate club (weighted) network can be downloaded from <http://vlado.fmf.uni-lj.si/pub/networks/data/ucinet/ucidata.htm#kazalo> (accessed on 12 November 2022) and it is the weighted version of the classical Karate club network. The Gahuku-Gama subtribes network can be downloaded from <http://konect.cc/networks/ucidata-gama/> (accessed on 12 November 2022) and its ground truth of node labels can be found in Figure 9 (b) of [61]. The Slovene Parliamentary Party network can be downloaded from <http://vlado.fmf.uni-lj.si/pub/networks/data/soc/Samo/Stranke94.htm> (accessed on 12 November 2022). The other five data sets with ground truth of node labels can be downloaded from <http://www-personal.umich.edu/~mejn/netdata/> (accessed on

12 November 2022). In particular, for the Dolphins network, as analyzed in [62], both $K = 2$ or $K = 4$ are reasonable.

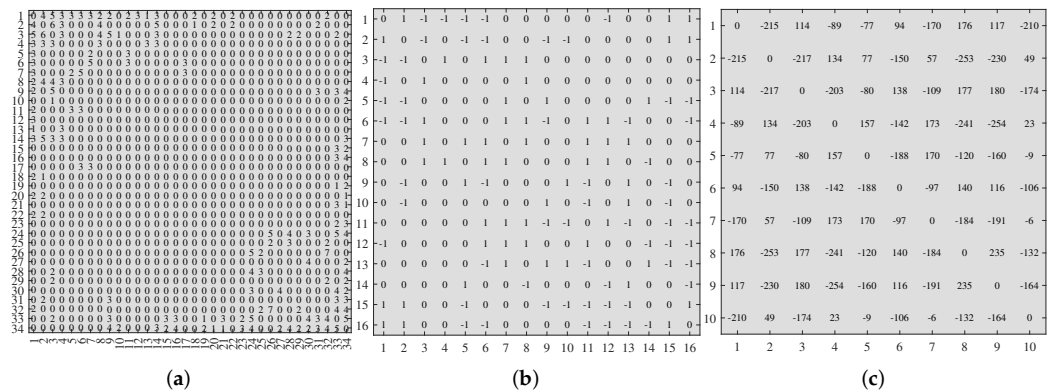


Figure 10. Adjacency matrices of Karate club (weighted), Gahuku-Gama subtribes, and Slovene Parliamentary Party network. (a) Karate club (weighted). (b) Gahuku-Gama subtribes. (c) Slovene Parliamentary Party network.

Table 1. Comparison of estimated K in real-world networks.

Dataset	Source	n	K	Weighted?	nDFAwm	ME	NB	BHm	BHa	BHmc	BHac
Karate club (weighted)	[63]	34	2	Yes	2	2	4	4	4	4	4
Gahuku-Gama subtribes	[64]	16	3	Yes	3	N/A	1	1	12	N/A	13
Slovene Parliamentary Party	[65]	10	2	Yes	2	2	N/A	N/A	N/A	N/A	N/A
Dolphins	[66]	62	2, 4	No	4	2	2	2	2	2	2
College football	[67]	110	11	No	11	10	10	10	10	10	10
Karate club	[63]	34	2	No	2	34	2	2	2	2	2
Political books	[68]	105	3	No	4	2	3	3	4	4	4
Political blogs	[69]	1222	2	No	2	2	7	7	7	8	8

For real-world networks, we compare our nDFAwm with the modularity eigengap (ME) [60], NB [46], BHm [46], BHa [46], BHmc [46], and BHac [46]. For our nDFAwm, we take $K_c = n$. Figure 11 displays the weighted modularity from Equation (2) by the nDFA algorithm for different choices of the number of clusters and we can find the nDFAwm's estimated K of the eight real-world networks from Figure 11 directly. Table 1 shows the estimated number of clusters for these networks. For all networks except for the Political books network, our nDFAwm successfully determines the correct number of communities. For the ME method, it estimates the correct K for Karate club (weighted), Slovene Parliamentary Party Network, Dolphins, and Political blogs while it fails for the other four networks. For NB and BHm methods, they only estimate K correctly for Dolphins, Karate club, and Political books. For BHa, BHmc, and BHac, they only estimate K successfully for Dolphins and Karate club. In particular, the non-backtracking method and Bethe Hessian matrix-based methods proposed in [46] fail to estimate the number of communities for the three real-world weighted networks in Table 1. As a result, our nDFAwm outperforms its competitors in these real-world networks.

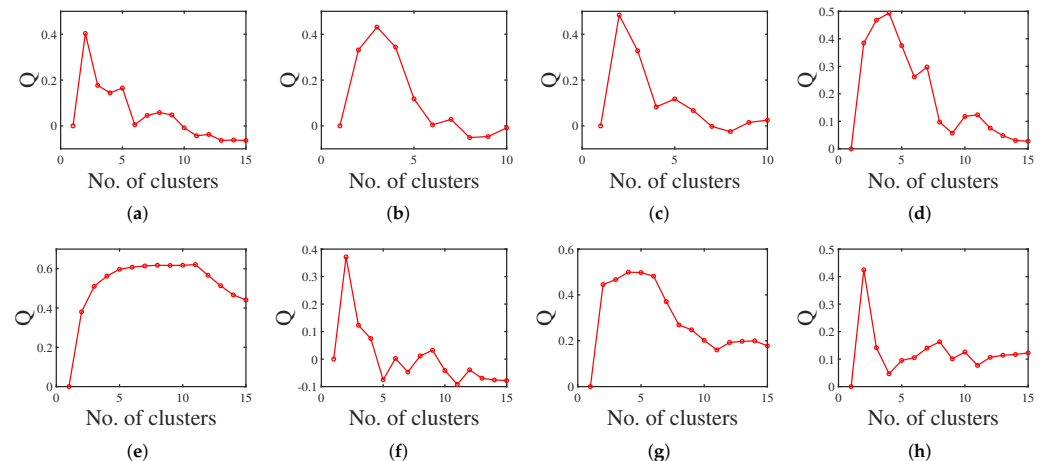


Figure 11. Weighted modularity Q obtained from Equation (2) against the number of clusters by the nDFA algorithm for real-world networks considered in this paper. (a) Karate club (weighted). (b) Gahuku-Gama subtribes. (c) Slovene Parliamentary Party network. (d) Dolphins. (e) College football. (f) Karate club. (g) Political books. (h) Political blogs.

4. Conclusions and Future Work

In this paper, we proposed a method for determining the number of communities for weighted networks in DCDFM. The method is designed based on a combination of weighted modularity and a spectral clustering algorithm. This estimation method enables us to estimate the number of communities even in the case where there is only one community in a weighted network generated by different distributions under DCDFM. Through substantial computer-generated weighted networks from DCDFM and several real-world networks, the numerical results show that the estimation accuracy of our approach is better than its competitors and our method also works for signed networks.

There are some open questions. First, building a theoretical guarantee on the consistency of our estimator for the true number of clusters under DCDFM is an attractive and challenging task. Second, determining the exact condition under which estimating the number of clusters is possible under DCDFM is a challenging problem. Third, in this paper, we are mainly interested in DCDFM for non-overlapping weighted networks, but the idea can be extended to overlapping weighted networks [70]. Fourth, in this paper, we estimate the number of communities for weighted networks generated from DCDFM by Equation (3) when we choose the method \mathcal{M} as the spectral method nDFA. If we let \mathcal{M} be algorithms developed in [48–54] to fit their weighted stochastic blockmodels for weighted networks, we wonder that we can also estimate K for these models through Equation (3). We leave them for the future.

Funding: This research was funded by the Scientific research start-up fund of CUMT No. 102520253, the High-level personal project of Jiangsu Province No. bJSSCBS20211218.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Barabási, A.L.; Albert, R. Emergence of scaling in random networks. *Science* **1999**, *286*, 509–512. [[CrossRef](#)] [[PubMed](#)]
2. Albert, R.; Barabási, A.L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **2002**, *74*, 47. [[CrossRef](#)]
3. Newman, M.E. The structure and function of complex networks. *SIAM Rev.* **2003**, *45*, 167–256. [[CrossRef](#)]

4. Boccaletti, S.; Latora, V.; Moreno, Y.; Chavez, M.; Hwang, D.U. Complex networks: Structure and dynamics. *Phys. Rep.* **2006**, *424*, 175–308. [\[CrossRef\]](#)
5. Lusseau, D.; Newman, M.E. Identifying the role that animals play in their social networks. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **2004**, *271*, S477–S481. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Guimera, R.; Nunes Amaral, L.A. Functional cartography of complex metabolic networks. *Nature* **2005**, *433*, 895–900. [\[CrossRef\]](#)
7. Barabasi, A.L.; Oltvai, Z.N. Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **2004**, *5*, 101–113. [\[CrossRef\]](#)
8. Palla, G.; Barabási, A.L.; Vicsek, T. Quantifying social group evolution. *Nature* **2007**, *446*, 664–667. [\[CrossRef\]](#)
9. Bullmore, E.; Sporns, O. Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* **2009**, *10*, 186–198. [\[CrossRef\]](#)
10. Foster, J. From simplistic to complex systems in economics. *Camb. J. Econ.* **2005**, *29*, 873–892. [\[CrossRef\]](#)
11. Schweitzer, F.; Fagiolo, G.; Sornette, D.; Vega-Redondo, F.; Vespignani, A.; White, D.R. Economic networks: The new challenges. *Science* **2009**, *325*, 422–425. [\[CrossRef\]](#)
12. Pastor-Satorras, R.; Castellano, C.; Van Mieghem, P.; Vespignani, A. Epidemic processes in complex networks. *Rev. Mod. Phys.* **2015**, *87*, 925. [\[CrossRef\]](#)
13. Chow, K.; Ay, A.; Elhesha, R.; Kahveci, T. ANCA: Alignment-based network construction algorithm. In Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, Washington, DC, USA, 29 August–1 September 2018; pp. 21–26.
14. Elhesha, R.; Sarkar, A.; Cinaglia, P.; Boucher, C.; Kahveci, T. Co-evolving patterns in temporal networks of varying evolution. In Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics, New York, NY, USA, 7–10 September 2019; pp. 494–503.
15. Cinaglia, P.; Cannataro, M. Network alignment and motif discovery in dynamic networks. *Netw. Model. Anal. Health Inform. Bioinform.* **2022**, *11*, 38. [\[CrossRef\]](#)
16. Newman, M.E. Analysis of weighted networks. *Phys. Rev. E* **2004**, *70*, 056131. [\[CrossRef\]](#)
17. Fortunato, S. Community detection in graphs. *Phys. Rep.* **2010**, *486*, 75–174. [\[CrossRef\]](#)
18. Fortunato, S.; Hric, D. Community detection in networks: A user guide. *Phys. Rep.* **2016**, *659*, 1–44. [\[CrossRef\]](#)
19. Newman, M.E. The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 404–409. [\[CrossRef\]](#)
20. Ji, P.; Jin, J. Coauthorship and citation networks for statisticians. *Ann. Appl. Stat.* **2016**, *10*, 1779–1812. [\[CrossRef\]](#)
21. Ji, P.; Jin, J.; Ke, Z.T.; Li, W. Co-citation and Co-authorship Networks of Statisticians. *J. Bus. Econ. Stat.* **2022**, *40*, 469–485. [\[CrossRef\]](#)
22. Schwikowski, B.; Uetz, P.; Fields, S. A network of protein–protein interactions in yeast. *Nat. Biotechnol.* **2000**, *18*, 1257–1261. [\[CrossRef\]](#)
23. Ideker, T.; Sharan, R. Protein networks in disease. *Genome Res.* **2008**, *18*, 644–652. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Holland, P.W.; Laskey, K.B.; Leinhardt, S. Stochastic blockmodels: First steps. *Soc. Netw.* **1983**, *5*, 109–137. [\[CrossRef\]](#)
25. Rohe, K.; Chatterjee, S.; Yu, B. Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Stat.* **2011**, *39*, 1878–1915. [\[CrossRef\]](#)
26. Amini, A.A.; Chen, A.; Bickel, P.J.; Levina, E. Pseudo-likelihood methods for community detection in large sparse networks. *Ann. Stat.* **2013**, *41*, 2097–2122. [\[CrossRef\]](#)
27. Lei, J.; Rinaldo, A. Consistency of spectral clustering in stochastic block models. *Ann. Stat.* **2015**, *43*, 215–237. [\[CrossRef\]](#)
28. Jin, J. Fast community detection by SCORE. *Ann. Stat.* **2015**, *43*, 57–89. [\[CrossRef\]](#)
29. Joseph, A.; Yu, B. Impact of regularization on spectral clustering. *Ann. Stat.* **2016**, *44*, 1765–1791. [\[CrossRef\]](#)
30. Mao, X.; Sarkar, P.; Chakrabarti, D. On Mixed Memberships and Symmetric Nonnegative Matrix Factorizations. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 2324–2333.
31. Chen, Y.; Li, X.; Xu, J. Convexified modularity maximization for degree-corrected stochastic block models. *Ann. Stat.* **2018**, *46*, 1573–1602. [\[CrossRef\]](#)
32. Zhang, Y.; Levina, E.; Zhu, J. Detecting overlapping communities in networks using spectral methods. *SIAM J. Math. Data Sci.* **2020**, *2*, 265–283. [\[CrossRef\]](#)
33. Mao, X.; Sarkar, P.; Chakrabarti, D. Overlapping Clustering Models, and One (class) SVM to Bind Them All. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 3–8 December 2018; Volume 31, pp. 2126–2136.
34. Mao, X.; Sarkar, P.; Chakrabarti, D. Estimating Mixed Memberships With Sharp Eigenvector Deviations. *J. Am. Stat. Assoc.* **2020**, *116*, 1928–1940. [\[CrossRef\]](#)
35. Li, X.; Chen, Y.; Xu, J. Convex relaxation methods for community detection. *Stat. Sci.* **2021**, *36*, 2–15. [\[CrossRef\]](#)
36. Jing, B.; Li, T.; Ying, N.; Yu, X. Community detection in sparse networks using the symmetrized laplacian inverse matrix (slim). *Stat. Sin.* **2022**, *32*, 1. [\[CrossRef\]](#)
37. Newman, M.E.; Reinert, G. Estimating the number of communities in a network. *Phys. Rev. Lett.* **2016**, *117*, 078301. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Bickel, P.J.; Sarkar, P. Hypothesis testing for automated community detection in networks. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2016**, *78*, 253–273. [\[CrossRef\]](#)
39. Lei, J. A goodness-of-fit test for stochastic block models. *Ann. Stat.* **2016**, *44*, 401–424. [\[CrossRef\]](#)

40. Riolo, M.A.; Cantwell, G.T.; Reinert, G.; Newman, M.E. Efficient method for estimating the number of communities in a network. *Phys. Rev. E* **2017**, *96*, 032310. [[CrossRef](#)]
41. Saldaña, D.F.; Yu, Y.; Feng, Y. How many communities are there. *J. Comput. Graph. Stat.* **2017**, *26*, 171–181. [[CrossRef](#)]
42. Wang, Y.R.; Bickel, P.J. Likelihood-based model selection for stochastic block models. *Ann. Stat.* **2017**, *45*, 500–528. [[CrossRef](#)]
43. Yan, B.; Sarkar, P.; Cheng, X. Provable estimation of the number of blocks in block models. In Proceedings of the International Conference on Artificial Intelligence and Statistics, PMLR, Virtual Event, 28–30 March 2022; pp. 1185–1194.
44. Chen, K.; Lei, J. Network cross-validation for determining the number of communities in network data. *J. Am. Stat. Assoc.* **2018**, *113*, 241–251. [[CrossRef](#)]
45. Ma, S.; Su, L.; Zhang, Y. Determining the number of communities in degree-corrected stochastic block models. *J. Mach. Learn. Res.* **2021**, *22*.
46. Le, C.M.; Levina, E. Estimating the number of communities by spectral methods. *Electron. J. Stat.* **2022**, *16*, 3315–3342. [[CrossRef](#)]
47. Jin, J.; Ke, Z.T.; Luo, S.; Wang, M. Optimal estimation of the number of network communities. *J. Am. Stat. Assoc.* **2022**. [[CrossRef](#)]
48. Aicher, C.; Jacobs, A.Z.; Clauset, A. Learning latent block structure in weighted networks. *J. Complex Netw.* **2015**, *3*, 221–248. [[CrossRef](#)]
49. Jog, V.; Loh, P.L. Recovering communities in weighted stochastic block models. In Proceedings of the 2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton), Monticello, IL, USA, 29 September–2 October 2015; pp. 1308–1315.
50. Ahn, K.; Lee, K.; Suh, C. Hypergraph Spectral Clustering in the Weighted Stochastic Block Model. *IEEE J. Sel. Top. Signal Process.* **2018**, *12*, 959–974. [[CrossRef](#)]
51. Palowitch, J.; Bhamidi, S.; Nobel, A.B. Significance-based community detection in weighted networks. *J. Mach. Learn. Res.* **2018**, *18*, 1–48.
52. Peixoto, T.P. Nonparametric weighted stochastic block models. *Phys. Rev. E* **2018**, *97*, 12306–12306. [[CrossRef](#)] [[PubMed](#)]
53. Xu, M.; Jog, V.; Loh, P.L. Optimal rates for community estimation in the weighted stochastic block model. *Ann. Stat.* **2020**, *48*, 183–204. [[CrossRef](#)]
54. Ng, T.L.J.; Murphy, T.B. Weighted stochastic block model. *Stat. Methods Appl.* **2021**, *30*, 1365–1398. [[CrossRef](#)] [[PubMed](#)]
55. Qing, H. Distribution-Free Model for Community Detection. *Prog. Theor. Exp. Phys.* **2023**, *2023*, 033A01. [[CrossRef](#)]
56. Qing, H. Degree-corrected distribution-free model for community detection in weighted networks. *Sci. Rep.* **2022**, *12*, 15153. [[CrossRef](#)]
57. Karrer, B.; Newman, M.E.J. Stochastic blockmodels and community structure in networks. *Phys. Rev. E* **2011**, *83*, 16107. [[CrossRef](#)] [[PubMed](#)]
58. Gómez, S.; Jensen, P.; Arenas, A. Analysis of community structure in networks of correlated data. *Phys. Rev. E* **2009**, *80*, 016114. [[CrossRef](#)] [[PubMed](#)]
59. Newman, M.E.J. Modularity and community structure in networks. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 8577–8582. [[CrossRef](#)] [[PubMed](#)]
60. Budel, G.; Van Mieghem, P. Detecting the number of clusters in a network. *J. Complex Netw.* **2020**, *8*, cnaa047. [[CrossRef](#)]
61. Yang, B.; Cheung, W.; Liu, J. Community mining from signed social networks. *IEEE Trans. Knowl. Data Eng.* **2007**, *19*, 1333–1348. [[CrossRef](#)]
62. Liu, W.; Jiang, X.; Pellegrini, M.; Wang, X. Discovering communities in complex networks by edge label propagation. *Sci. Rep.* **2016**, *6*, 22470. [[CrossRef](#)]
63. Zachary, W.W. An information flow model for conflict and fission in small groups. *J. Anthropol. Res.* **1977**, *33*, 452–473. [[CrossRef](#)]
64. Read, K.E. Cultures of the central highlands, New Guinea. *Southwest. J. Anthropol.* **1954**, *10*, 1–43. [[CrossRef](#)]
65. Ferligoj, A.; Kramberger, A. An analysis of the slovene parliamentary parties network. *Dev. Stat. Methodol.* **1996**, *12*, 209–216.
66. Lusseau, D.; Schneider, K.; Boisseau, O.J.; Haase, P.; Slooten, E.; Dawson, S.M. The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. *Behav. Ecol. Sociobiol.* **2003**, *54*, 396–405. [[CrossRef](#)]
67. Girvan, M.; Newman, M.E. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 7821–7826. [[CrossRef](#)] [[PubMed](#)]
68. Newman, M.E. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E* **2006**, *74*, 036104. [[CrossRef](#)] [[PubMed](#)]
69. Adamic, L.A.; Glance, N. The political blogosphere and the 2004 US election: Divided they blog. In Proceedings of the 3rd International Workshop on Link Discovery, Chicago, IL, USA, 21–25 August 2005; pp. 36–43.
70. Qing, H. Mixed membership distribution-free model. *arXiv* **2021**, arXiv:2112.04389.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.