Ibn Al-Haitham Journal for Pure and Applied Sciences
Journal homepage: jih.uobaghdad.edu.iq
PISSN: 1609-4042, EISSN: 2521-3407
IHJPAS. 2025, 38(3)

Improving Modularity Regularization Techniques for Estimating Structures in Complex Network

Samaa F. Ibraheem[1*] **and Basad Al-Sarray** [2]
[1]Department of Mathematics, College of Applied Science, University of Technology, Baghdad, Iraq.
[2]Department of Computer Science, College of Science, University of Baghdad, Baghdad, Iraq.
*Corresponding Author.

Received: 17 September 2023 Accepted: 19 February 2024 Published: 20 July 2025
doi.org/10.30526/38.3.3728

Abstract

Complex systems in the real world have networks differ significantly from random graphs and have non-trivial structures. In fact, they have a community structure that needs to be recognized and recovered. The stochastic block models (SBMs) are popular models for community detection in networks, where nodes are divided into groups based on their connectivity patterns. Maximum likelihood estimation is a common method for estimating the parameters of SBMs. In this paper, a model selection for stochastic block models is presented based on the optimization of the log-likelihood function to find the best number of communities K detected by the regularized convex modularity maximization method. This work deals with many assumptions on K because it is necessary to study the behavior of the network and the best optimum K is assumed to select the best partition over the AIC, BIC metric. The proposed model selection method is presented in an algorithm that is implemented for both real and synthetic networks. This method enables the detection of networks with small communities that are more likely to provide a better fit to the observed data.

Keywords: Community detection, likelihood function, model selection, Bayesian Information Criterion, Stochastic Block Models.

1. Introduction

The stochastic block model (SBM) in its standard form is a popular tool for detecting communities in networks. It is based on the assumption that the nodes in a network can be divided into groups or blocks, so that the probability of an edge between two nodes depends only on which block they belong to. This enables the estimation of block membership and the detection of communities within a network (1). The SBM has been used to detect communities in social networks, biological networks and other types of networks (2) and (3). More specifically, it can be used to estimate the number of communities in a network as well as their size and connectivity patterns. In addition, it can be used to identify outliers or anomalous nodes that do not fit into any of the detected clusters (4).

The SBM model works by randomly assigning nodes to different blocks and then connecting nodes within each block with a certain probability, while connecting nodes between blocks with a different probability. In this way, networks can be created with different communities that are connected to different degrees (5, 6). Model selection is an important task in SBM analysis as it helps to select the most accurate model from a set of candidates (7). This includes selecting the number of groups or communities as well as the parameters that determine the edge probabilities within and between groups. There are several approaches to model selection for SBM, including likelihood-based methods (8, 9), information-theoretic methods (10), spectral clustering (11), and cross-validation (12).

Likelihood-based methods are commonly used for model selection in SBM analysis. Here, the maximum likelihood estimation (MLE) method is used to estimate the parameters of the SBM to find the parameter values that maximize the likelihood function. MLE can also be used to compare different SBMs by comparing their likelihood values using different penalized likelihood information criteria, such as the Akaike's Information Criterion (AIC), the Consistent AIC, the Bayesian Information Criterion (BIC) and the Adjusted BIC (13). Thus, AIC is an estimate of a constant plus the relative distance between the fitted likelihood function of the model and the unknown true likelihood function of the data, while BIC is an estimate of a function of the posterior probability of a model being true under a given Bayesian setup, such that a low AIC and BIC mean that the model is considered closer to the truth (14,15). However, optimization of the likelihood function can be practically performed by optimizing the profile likelihood criteria over all possible partitions for each K. Subsequently, the best model can be selected using AIC or BIC.

Overall, likelihood-based methods provide a rigorous framework for model selection in SBM analysis and help researchers to select the best-fitting model based on statistical criteria. Moreover, model selection for SBM can be challenging due to the high-dimensional parameter space and non-convex optimization problem (16). Therefore, it is important to use multiple approaches and carefully evaluate their results before drawing conclusions about the underlying structure of the network.

The stochastic block models (SBMs) are popular probabilistic models for community detection in networks. They provide a powerful and modern way to define and understand network structure (17). There are many types of SBMs, including symmetric SBMs and overlapping SBMs (18), but the focus is on SBMs in standard form and SBMs with corrected degree. To represent the standard SBM, an undirected and unweighted random graph G with n nodes is considered, where $A = [a_{ij}] \in R^{n\times n}$ is the adjacency matrix of G that characterizes the relation between each pair of vertices i and j. A is symmetric. To model A by SBM, assume that a_{ij} s are independent Bernoulli variables, that is

$$a_{ij} \sim Ber(p_{ij}),\ p_{ij} \in (0,1), i < j, a_{ii} = 0 \quad (1)$$

and the vertices in G are partitioned into K groups (communities) such that the vertex i belongs to the community $C_k, k = 1, \ldots, K$, according to the map $C : \{1, \ldots, n\} \rightarrow \{1, \ldots, K\}$ such that $p_{ij} = B_{z_i z_j}$. Let $Z \in \{1, \ldots, K\}^n$ be the vector of labels that members the vertices i's to one of K communities, i.e., $Z = \{z_1, z_2, \ldots, z_n\}$, where $z_i = C(i), i = 1, \ldots, n$. The membership label Z that defines community assignment is drawn independently from a multinomial distribution on K communities with probability $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)$ $\sum_{k=1}^{K} \pi_k = 1$, that is $Z \overset{\text{i.i.d}}{\sim}$ multinomial $(1, \pi)$ (2)

In SBM, the probability that vertices i and j are connected ($Pr(a_{ij} = 1) = p_{ij}$) depends on their community membership; therefor, p_{ij} is identical for any i and j that belong to the same community. By this way, a symmetric $K \times K$, matrix B is formed such that $p_{ij} = B_{z_i z_j}$, where $B \in R^{K \times K}$, and $K \ll n$. This probability matrix is called a stochastic block matrix or connectivity matrix, so there are K $(K + 1)/2$ connectivity parameters to estimate in this model (19). In networks with community structures, the edge probabilities within a same block exceed the probabilities that vertices of different blocks are connected, i.e., the vertices within a same block are more likely to be connected, so in community detection, we seek for $B_{KK} > B_{qr}$, $\forall k, q, r = 1, 2, \cdots, K$, with $q \neq r$.

The degree-corrected stochastic block model is a probabilistic model for network data that extends the traditional stochastic block model (SBM) by incorporating node-specific degree information. In the degree corrected SBM, each node is assigned to one of K blocks, and the probability of an edge between two nodes depends on their block assignments as well as their individual degrees. So the probability matrix B, with size $K \times K$, is formed as $\mathrm{p_{ij}} = \theta_i \theta_j \, \mathrm{B}_{z_i z_j}$, where $\theta_i > 0$ is an additional parameter assigned to each node to control its expected degree (1). Moreover, the value a_{ij} is a Poisson-distributed random variable with mean $\lambda = \mathrm{p_{ij}}$, i.e.,

$$\mathrm{A_{ij}} = \mathrm{Poisson}(\theta_i \theta_j \, \mathrm{B}_{z_i z_j}) \tag{3}$$

In this paper, a model selection for stochastic block models is performed based on optimizing the log likelihood function to find the best number K of communities. The main aim is to select the model for communities that are detected using a convex optimization approach under degree corrected SBM, which is regularized convex modularity maximization method (RCMM) that is suggested in (20), and compare this result with others detected using various algorithms, which are convex modularity maximization CMM (4, 21), modularity maximization algorithm (22), Danon algorithm (23) and spectral clustering (11). The article is organized as follows: Section 2 introduces stochastic block models. Section 3 presents the objective, and some details on algorithmic settings are also given. Numerical results are presented in Section 4 that show the performance of the model selection algorithm on the RCMM method compared with other methods. Finally, concluding remarks are given in Section 5.

2.Optimizing Likelihood Function for Model Selection and Estimation

In community detection, the goal of optimization methods with stochastic block models is to estimate the parameters of the model, such as the number of communities and the probability of connections between nodes within and between communities, in order to get the optimal membership label vector Z. This is done by maximizing a likelihood function that measures how well the model fits the observed network data, therefore detecting the network community.

In the stochastic block model SBM, the A_{ij}'s are independent Bernoulli variables with probability p_{ij}, which is the expectation of A_{ij} That is

$E(A_{ij}) = pij$ which performs a $K \times K$ symmetric matrix (connectivity matrix) $\mathrm{B}_{z_i z_j}$

Given the node labels (the vector Z), the likelihood of observing A with probability B under SBM is:

$$L(A|B) = \prod_{i<j} {B_{z_i z_j}}^{A_{ij}} \, (1 - B_{z_i z_j})^{(1 - A_{ij})} \tag{4}$$

whereas the DCSBM, which limits the applicability of the SBM, has an expected value of the adjacency matrix element A_{ij} equal to:

$$E(A_{ij}) = \theta_i \theta_j B_{z_i z_j} \tag{5}$$

but a Poisson distribution with this mean is used to determine the actual number of edges between any two vertices. The likelihood function is defined as

$$L(A|\theta, B, Ƶ) = \prod_{i<j} \frac{(\theta_i\theta_j B_{z_iz_j})^{A_{ij}}}{A_{ij}!} \exp(-\theta_i\theta_j B_{z_iz_j}) \times \prod_i \frac{(\frac{1}{2}\theta_i{}^2 B_{z_iz_i})^{A_{ij}/2}}{(A_{ij}/2)!} \exp(-\frac{1}{2}\theta_i^2 B_{z_iz_i}) \quad (6)$$

To find the optimal partition, the likelihood function is maximized over *B*, giving the profile likelihood criteria to be optimized over all possible partitions. Actually, the Bernoulli likelihood is replaced by the Poisson likelihood, to simplify derivations. So for these two models, a profile likelihood was driven from the log-likelihood function of each of them to give the following criteria (1):

$$Q_{SBM} = \sum_{kl} m_{kl} \log \frac{m_{kl}}{n_k n_l} \quad (7)$$

$$Q_{DCSBM} = \sum_{kl} m_{kl} \log \frac{m_{kl}}{d_k d_l} \quad (8)$$

where $m_{kl} = \sum_{ij} A_{ij}\, I\{z_i = k, z_j = l$ is the number of edges from community k to community $l, k \neq l$. $d_k = \sum_{l=1}^{K} m_{kl}$ be the node degrees in community *k*. In fact m_{kk} gives twice the number of edgeswithin community k, and n_k is the number of nodes of community *k*, also $\frac{m_{kl}}{n_k n_l}$ and m_{kl} give the maximum likelihood values $\hat{B}_{kl}$ of the SBM and DCSBM model parameters respectively.

3.Enhancement of CMM for Community Detection

There are many methods for detecting communities, distinguishing between traditional methods and optimization- and model-based methods. The optimization algorithm tries to identify communities within the network (graph) based on the optimization of some quality functions. Depending on the representation of the network, the modularity function, denoted by *Q*, can be defined as follows: It is a function that measures the strength of the partition of a network into communities (subgraphs) considering the degree distribution, such that networks with a high modularity value have dense connections between nodes within a cluster but sparse connections between nodes of different clusters. The modularity function is given by:

$$Q = \frac{1}{2m} \sum_{ij} \left(a_{ij} - \frac{d_i d_j}{2m}\right) \delta(C_i, C_j) \quad (9)$$

where ,a_{ij}.=A the adjacency matrix and Kronecker delta ($\delta(C_i$,$C_j)$ is a function that yield 0 or 1 if vertices *i* of degree ,d_i.and j of degree ,d_j. are in the same community or not). So, the possible existing communities are gained by balancing between the actual density of edges and the expected density that would appear in the communities if the vertices were connected regardless of community structure. Many methods for community detection are used in this work for in comparison, these methods are summarized as follows:

Newman Modularity Maximization: this is an agglomerative hierarchical clustering technique in which smaller communities of vertices are gradually merged into larger ones, leading to an increase in modularity (22). The algorithm works as follows:

- Step 1: Assign each of the n vertices to a single cluster.
- Step 2: Reduce the number of clusters by adding an edge to the set of unconnected vertices such that the resulting partition has a modularity Equation (9) greater than the value of the previous configuration.
- Step 3: Choose the optimal fusion by calculating the variation ΔQ
- Step 4: Repeat steps 2 and 3 until you have added all edges and obtained one cluster. Save the modularity score for each partition.
- Step 5: Select the partition with the highest modularity score.

3.1. Danon Method

In this method, (23) proposed to normalize the modularity variation Q resulting from the union of two communities by the proportion of edges belonging to one of the two communities, since the Newman technique tends to form large communities at the expense of small ones. Thus, this approach leads to larger modularity optima, especially when the communities have very different sizes.

3.2.Convex Modularity Maximization

The CMM algorithm makes use of convex optimization, where it is based on a convex programming relaxation of the modularity optimization and works under block models (21). Thus, so this algorithm tends to maximize modularity by solving the following optimization problem:

$$\begin{aligned} &\max_X \langle X, A - \lambda d d^T \rangle \\ &subject\ to\ X \succcurlyeq 0,\ \ 0 \le X \le J,\ \ diag(X) = 1 \end{aligned} \qquad (10)$$

where X is the partitioning matrix for the network nodes. The CMM approach combines convex optimization techniques with modularity maximization and a weighted $\ell 1$-norm k-medoids as follows. After obtaining the optimal solution to Equation (10), $\hat{X}$, by the alternating direction method of multipliers (ADMM), the communities is extracted from weighted $\hat{X}$ by weighted $\ell 1$-norm k-medoids, i.e by solving the optimization:

$$\begin{aligned} &min_{C_k,\ c_r} \quad \textstyle\sum_{1 \le r \le K} \sum_{i \in C_k} d_i \|\hat{w}_i - c_r\|_1 \\ &s.t \quad c_r \in rows(\widehat{W}) \qquad \forall r \in \{1, \dots, K\} \end{aligned} \qquad (11)$$

where $\widehat{W} = D\hat{X}$, $D = diag(d)$ and c_r are the centers, so the clustering is performed on the rows of $\widehat{W}$ instead of $\hat{X}$, thus, this procedure is summarized as follows:

- Step 1: Solve the optimization problem (10) for X using the alternate direction method of multipliers (22).
- Step 2: Find the k-partition of n nodes by the k-medoids approach, weighted by node degrees.

3.3. Regularized Convex Modularity Maximization

The RCMM algorithm is considered a regularization of the CMM algorithm. Where the common neighbor's similarity measurement is used to weight the $\ell 1$-norm of k-mediod clustering (20) instead by node degrees. Where the weight was given twice, once to the rows of the partition matrix X and secondly to the terms of the $\ell 1$-norm of k-medoides. The number σ_{ij} of common neighbors between two vertices i and j is defined as

$$\sigma_{ij} = \textstyle\sum_{k=1}^{n} A_{ik} A_{kj} \qquad (12)$$

So the weighted $\ell 1$-norm k-medoids clustering will be done for the matrix $\hat{S} = \sigma\hat{X}$ and the partition yields by solving the optimization problem:

$$\begin{aligned} &min_{C_k,\ c_r} \quad \textstyle\sum_{1 \le r \le K} \sum_{i \in C_k} \sigma_{ij} \|\hat{s}_i - c_r\|_1 \\ &\quad s.t \quad c_r \in rows(\hat{S}), r \in \{1, \dots, K\} \end{aligned} \qquad (13)$$

3.4. Spectral Clustering

In general it is a method uses of information which is taken from the eigenvalues of special matrices like Affinity Matrix, Degree Matrix and Laplacian Matrix derived from the data set (graph). It takes the matrix of eigenvectors associated with the K largest eigenvalues of one of these special matrices and performs one of the distance-based clustering methods on it (25). So mathematically, it is summarized in the following steps:

- Step 1: Take the leading eigenvectors $[V_1, V_2, \dots, V_K] \in R^{n \times K}$ of the adjacency A or its graph normalized Laplacian $L(A) = D^{-1/2} A D^{-1/2}$, where $D = diag(d_i)$ is the diagonal matrix of degrees and the eigenvectors are associated with the K largest eigenvalues of the chosen matrix.

- Step 2: Normalize each rows of the matrix $V = [V_1, V_2, \dots, V_K]$ to find V^* with elements $V_{ij}^* = V_{ij}/(\sum_j V_{ij}^2)^{1/2}$.
- Step 3: Run any distance clustering method (e.g., *K*-means) on the rows of the resulted vectors V^* to partition them into desired number of clusters (*K*), where the output $C_1, \dots, C_K$ and the node *i* is assigned to cluster *r* if the *i*"th row of $V^*$is assigned to cluster C_r.

4. Model Selection

Model selection for SBMs involves choosing not only the marginal probabilities but also the number of communities K in order to determine the best fitting model from a set of candidate models (26). This is done in two stages: In the first, the parameters for the likelihood function are estimated for many chosen *K*, while in the second, the model that best fits the observed data is selected, i.e.,. the model that has the minimum AIC or BIC value (27-29).

In likelihood-based methods, the likelihood of the observed network is maximized among different SBM models with different numbers of communities and parameter values. The Akaike information criterion (AIC) and Bayesian information criterion (BIC) are commonly used to compare models with different numbers of parameters and select an appropriate model. Let *r* be the number of estimated parameters in the model

$$AIC = 2r - 2\,ln(L) \tag{14}$$

and

$$BIC = r\,ln(n) - 2\,ln(L) \tag{15}$$

The model that has the lowest AIC and BIC values from a group of possible models for the data is the one that should be selected. In other words, the AIC or BIC can be used to compare different models with different numbers of parameters and select the one that best fits the data, where the lower the AIC or BIC value, the better the model fits the data (30).

Therefore, two stages are needed to estimate the parameters , .and choose the optimum model. The first is to choose the best membership label Z_{best}, which represents the partition of the network, using any community detection algorithm for many values $K<<n$ such that, $Z_{\text{best}} = \arg\max_Z \ell(A|Z)$. and find the likelihood for every partition using Equation (7) or (8). This probability indicates how likely it is that the observed data *A* matches the model specified by the optimized parameters *B* for the best label *Z*. In the second phase, the corresponding AIC or BIC values are determined, Equations (6) and (7) respectively. As a result, the model that best fits the data and has a lower AIC or BIC value is selected. This procedure can be summarized using the flowchart in **Figure (1)** with its **Algorithm (1)**.

Algorithm (1) can be used with any community detection algorithm of a network with *n* nodes for any chosen number of communities *K* to identify the model that is more likely to fit the observed network or discover *K* for a model that is more likely to do so.

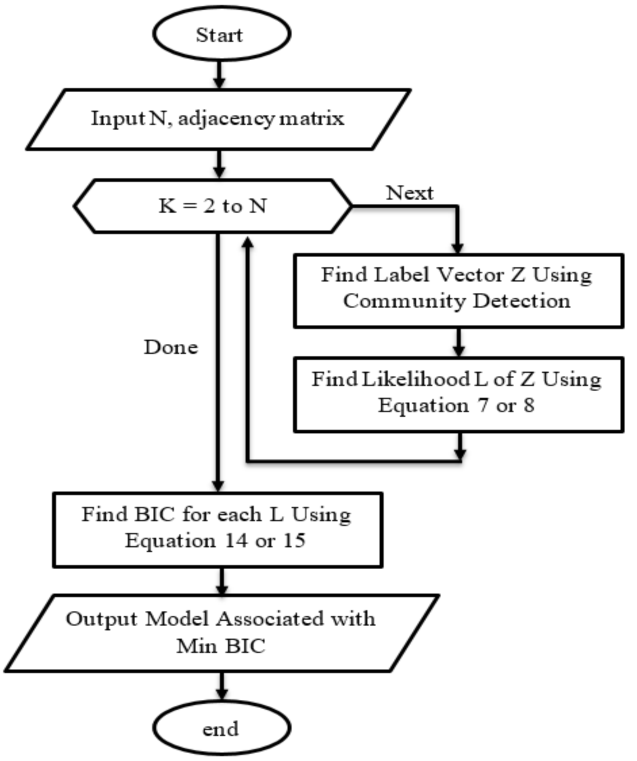


Figure1: Flowchart for Model Selection.

Algorithm 1. Model selection algorithm for SBMs

Step1: **Input:** Network adjacency matrix *A* of *n* by *n* elements.
Step 2: For K=2 to N , where $N<<n$
Find the label vector Z by using any community detection algorithm.
Find its likelihood using Equation (7) or (8).
Step3: Put the likelihood values in statistical model selection criteria AIC or BIC Equation (14) or (15).
Step4: Pick up the model with smallest AIC or BIC.
Step5:**Output:**The selected model with communities *K* and its associated likelihood for given *Z*.

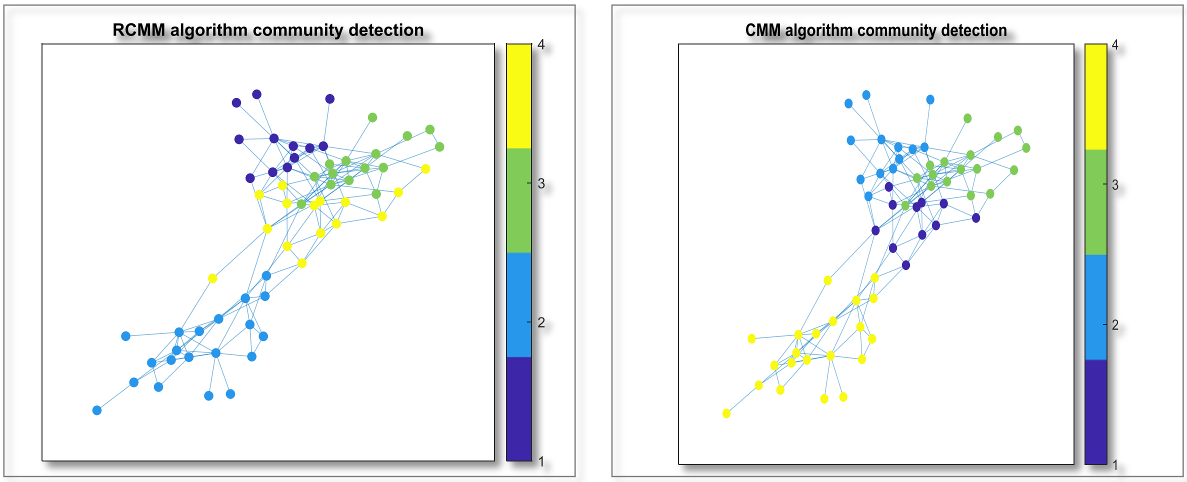
5.Empirical Example and Results

Algorithm (1) was implemented on various real and synthetic networks using a dataset from the Stanford Network Analysis Platform (SNAP) network analysis library. **Table (1**) shows the total number of nodes and edges in each network analyzed.

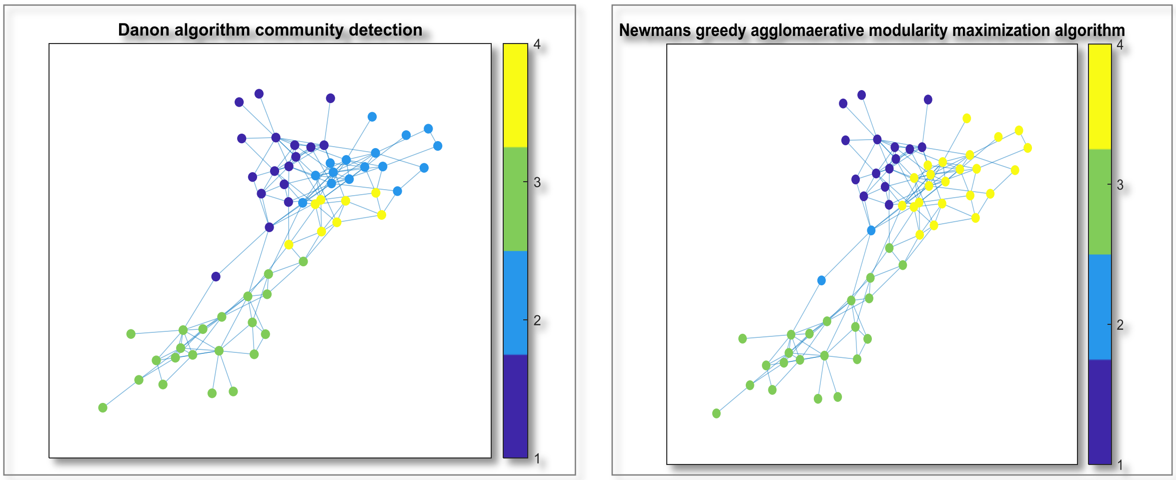
Table 1. Datasets for various networks from Stanford large network dataset collection

No.	Network	Vertices No.	Edges No.
1	Dolphin	62	159
2	Karate	34	78
3	American Football College (AFC)	115	613
6	Chesapeake Synthetic	39	170
7	Delaunay Synthetic	1024	3056

As an example, the experiment is performed with a dolphin network (62 nodes and 159 edges), as shown in **Table (1).** First, the detection is performed at K=4 with many algorithms including RCMM, CMM, Danon method, Newman modularity maximization, and spectral clustering with the corresponding likelihood values shown in **Figure (2).** Then, as mentioned earlier, the best label vector corresponding to the minimum likelihood value is chosen. All calculations are performed with a degree-corrected model, as it is more suitable for modeling networks.



Log-likelihood=-1.6117e+03 Log-likelihood=-1.6118e+03

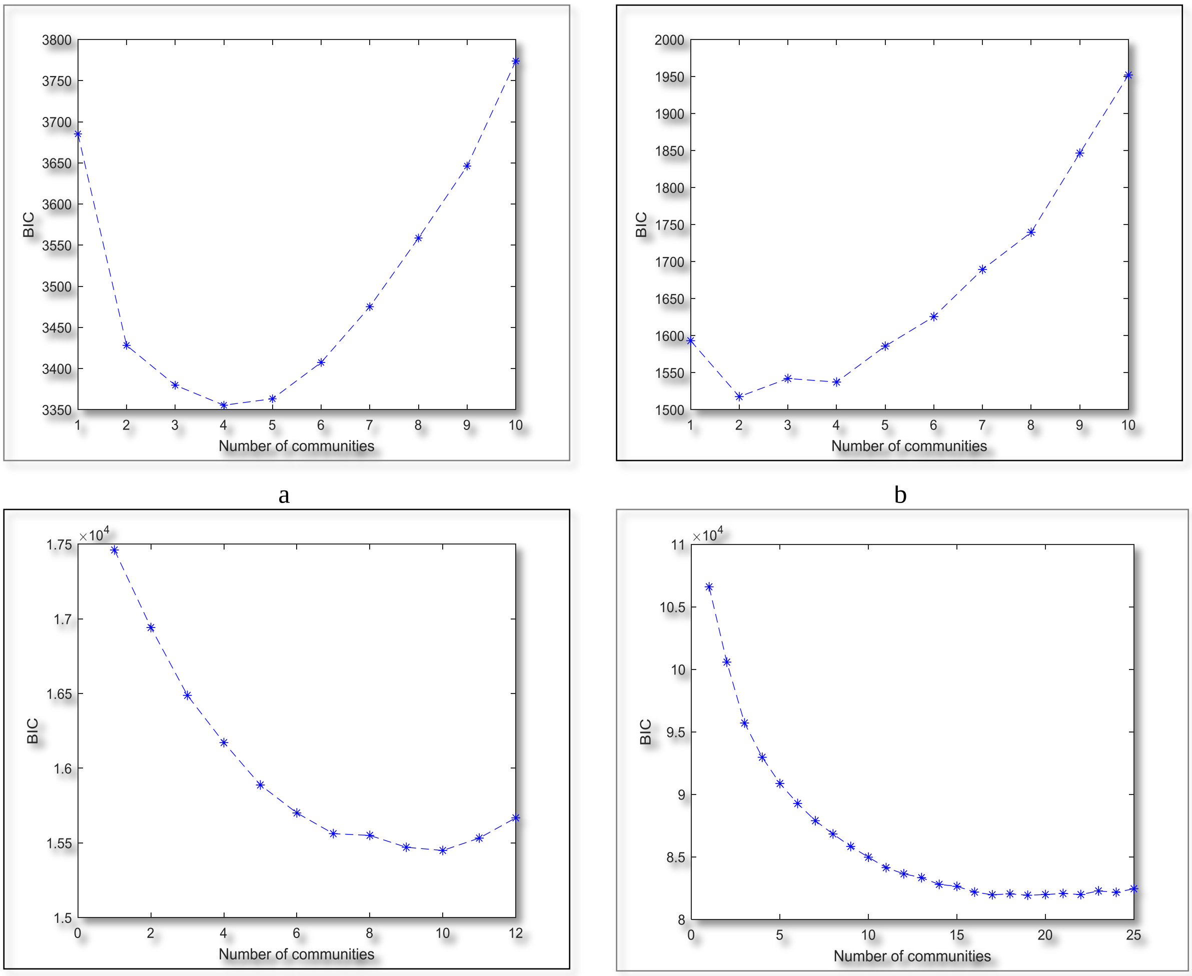


Log-likelihood=-1.6218e+03 Log-likelihood=-1.6386e+03

Figure 2. Dolphin network partition for K=4, using different community detection algorithms under degree corrected model with each log-likelihood values.

For model selection, the above experiment is applied for many *K* and on each network with the same community detection algorithm (see **Algorithm 1**), selecting the model that has the lower likelihood criteria for model selection, and here BIC is chosen. The applied community detection algorithm is RCMM to test whether it is sufficient to detect the correct model (**Figure 3**). These results are compared with the results of other community detection methods, namely: Danon algorithm, CMM algorithm, modularity maximization algorithm and spectral clustering for many networks (see **Table 2**).

In **Table (2),** the comparison is done for model selection between the suggested method RCMM and the other methods, including: Newman modularity maximization method, Danon method, CMM algorithm, and spectral clustering.



a b

c d

Figure 3. Model selection for RCMM community detection method: a) Dolphin network, b) Karate c)AFC d) Delaunay Synthetic

Table 2. Model selection for various networks with theirs minimum BIC and associated log-likelihood value using various community detection algorithms.

Network	RCMM			Danon method		
	K	**BIC**	**Log-likelihood**	**K**	**BIC**	**Log-likelihood**
Dolphin	4	3.3555×10^{3}	-1.6117×10^{3}	4	3.3757×10^{3}	-1.6218×10^{3}
Karate	2	1.5177×10^{3}	-7.3943×10^{2}	4	1.5464×10^{3}	-7.1679×10^{2}
American Football College (AFC)	10	1.5450×10^{4}	$-7.3286e\times10^{3}$	7	1.5662×10^{4}	-7.6268×10^{3}
Delaunay Synthetic	19	8.1922×10^{4}	-3.8976×10^{4}	9	8.6799×10^{4}	-4.2925×10^{4}
Chesapeake Synthetic	2	3.9148×10^{3}	-1.9373×10^{3}	4	4.0696×10^{3}	-1.9762×10^{3}

Network	CMM			Newman Modularity Maximization		
	K	**BIC**	**Log-likelihood**	**K**	**BIC**	**Log-likelihood**
Dolphin	4	3.3557×10^{3}	-1.6118×10^{3}	4	3.4093×10^{3}	-1.6386×10^{3}
Karate	2	1.5177×10^{3}	-7.3943×10^{2}	4	1.5735×10^{3}	-7.30324×10^{2}
American Football College (AFC)	10	1.5450×10^{4}	-7.3286×10^{3}	7	1.5662×10^{4}	-7.6268×10^{3}
Delaunay Synthetic	19	8.1885×10^{4}	-3.8942×10^{4}	7	8.9932×10^{4}	-4.4668×10^{4}
Chesapeake Synthetic	3	3.9262×10^{3}	-1.9265×10^{3}	3	3.9294×10^{3}	-1.9281×10^{3}

Network	Spectral clustering		
	K	**BIC**	**Log-likelihood**
Dolphin	4	3.3940×10^{3}	-1.6310×10^{3}
Karate	3	1.5119×10^{3}	-7.2069×10^{2}
American Football College (AFC)	8	1.5686×10^{4}	-7.5821×10^{3}
Delaunay Synthetic	19	8.268×10^{4}	-3.9358×10^{4}
Chesapeake Synthetic	4	3.9553×10^{3}	-1.9190×10^{3}

6. Conclusion

A model selection for stochastic block models based on the optimization of the log_likelihood function is given for the communities detected by the RCMM method under degree-corrected SBM to find the best number *K* of communities that have the minimum BIC. The results are compared with the results of other different methods, including CMM under degree-corrected SBM, Danon method, Newman modularity maximization, and spectral clustering, and applied to both real and synthetic networks. The results show that model selection for many networks partitioned with RCMM using degree- corrected SBM is identical to many results obtained with other community detection methods, and that in most cases it is better because it has a lower BIC value, so it fits the observed data better. Moreover, RCMM is able to detect small communities as it analyzes the networks with a larger value of *K*. However, the goal of model selection is to find a model that provides both a good fit to the existing observations and reasonable predictions for future observations.

Acknowledgment

Our researcher extends his Sincere thanks to the editor and members of the preparatory committee of the Ibn AL-Haitham Journal of Pure and Applied Sciences.

Conflict of Interest

There are no conflicts of interest.

Funding

There is no funding for the article.

References

1. Karrer B., Newman M.E.J. Stochastic blockmodels and community structure in networks. Phys Rev E Stat Nonlin Soft Matter Phys 2011;83(1):1–11.https://doi.org/10.1103/PhysRevE.83.016107
2. Abdul-Emeer S.S. An evolutionary bi-clustering algorithm for community detection in social networks. Iraqi J Sci 2016;57(3):2111–2120.
3. Mohammed S.K., Taha M.M., Taha E.M., Mohammad M.N.A. Cluster analysis of biochemical markers as predictor of COVID-19 severity. Baghdad Sci J 2022;19(6):1423–1429. https://doi.org/10.21123/BSJ.2022.7454
4. Li X., Chen Y., Xu J. Convex relaxation methods for community detection. Stat Sci 2021;36(1):2–15. https://doi.org/10.1214/19-STS715
5. Holland P.W., Laskey K.B., Leinhardt S. Stochastic blockmodels: first steps. Soc Networks 1983;5(2):109–137. https://doi.org/10.1016/0378-8733(83)90021-7.
6. Alsaleem N.Y.A., Alsaleem M.Y.A., Zakar N.A. Network performance analysis based on network simulator NS-2. Ibn Al-Haitham J Pure Appl Sci 2018;31:222–229.
7. Hassan F., Behadili S.F. Modeling social networks using data mining approaches—Review. Iraqi J Sci 2022;63(3):1313–1338. https://doi.org/10.24996/ijs.2022.63.3.35
8. Wang Y.X.R., Bickel P.J. Likelihood-based model selection for stochastic block models. Ann Stat 2017;45(2):500–528. https://doi.org/10.1214/16-AOS1457
9. Al-obedy J.A.N. Posterior estimates for the parameter of the Poisson distribution by using two different loss functions. Ibn Al-Haitham J Pure Appl Sci 2022;35(1):60–72. https://doi.org/10.30526/35.1.2800
10. Abbe E. Community detection and stochastic block models: recent developments. J Mach Learn Res 2017;18(177):1–86.
11. Von Luxburg U. A tutorial on spectral clustering. Stat Comput 2007;17(4):395–416. https://doi.org/10.1007/s11222-007-9033-z

12. Dabbs B., Junker B. Comparison of cross-validation methods for stochastic block models. arXiv Methodol 2016. http://arxiv.org/abs/1605.03000
13. Newman M.E.J. Networks: an introduction. Oxford Univ Press 2018. ISBN: 9780198805090.
14. Nishii R. Maximum likelihood principle and model selection when the true model is unspecified. J Multivar Anal 1988;27(2):392–403. https://doi.org/10.1016/0047-259X(88)90137-6
15. Ponveni P., Visumathi J. A review on community detection algorithms and evaluation measures in social networks. 2021 7th Int Conf Adv Comput Commun Syst 2021:1892–1899. https://doi.org/10.1109/ICACCS51430.2021.9441868
16. Yaqoob A.F., Al-Sarray B. Community detection under stochastic block model likelihood optimization via tabu search – Fuzzy C-mean method for social network data. Iraqi J Sci 2020;61(10):2695–2704. https://doi.org/10.24996/ijs.2020.61.10.27
17. Funke T., Becker T. Stochastic block models: a comparison of variants and inference methods. PLoS One 2019;14(4):e0215296. https://doi.org/10.1371/journal.pone.0215296
18. Abbe E., Sandon C. Community detection in general stochastic block models: fundamental limits and efficient algorithms for recovery. Proc Annu IEEE Symp Found Comput Sci 2015;2015:670–688. https://doi.org/10.1109/FOCS.2015.47
19. Daudin J., Picard F., Robin S. A mixture model for random graphs. Stat Comput 2008;18(2):173–183.
20. Ibraheem S.F., Al-Sarray B. A heuristic technique for community detection in complex networks. AIP Conf Proc 2024;3097:080012. https://doi.org/10.1063/5.0209446
21. Chen Y., Li X., Xu J. Convexified modularity maximization for degree-corrected stochastic block models. Ann Stat 2018;46(4):573–1602. https://doi.org/10.1214/17-AOS1595
22. Newman M.E.J. Fast algorithm for detecting community structure in networks. Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Top 2004;69(6):5. https://doi.org/10.1103/PhysRevE.69.066133
23. Danon L., Díaz-Guilera A., Arenas A. The effect of size heterogeneity on community identification in complex networks. J Stat Mech Theory Exp 2006;2006(11):P11010. https://doi.org/10.1088/1742-5468/2006/11/P11010
24. Cai T.T., Li X. Robust and computationally feasible community detection in the presence of arbitrary outlier nodes. Ann Stat 2015;43(3):1027–1059. https://doi.org/10.1214/14-AOS1290
25. Zhou Z., Amini A.A. Analysis of spectral clustering algorithms for community detection: the general bipartite setting. J Mach Learn Res 2019;20:1–45.
26. Yan B., Sarkar P., Cheng X. Provable estimation of the number of blocks in block models. Int Conf Artif Intell Stat AISTATS 2018;84:1185–1194.
27. Jin D., Zhao Z., Zhang Y., Zheng V.W., Yu S., Li Y. A survey of community detection approaches: from statistical modeling to deep learning. IEEE Trans Knowl Data Eng 2023;35(2):1149–1170. https://doi.org/10.1109/TKDE.2021.3104155
28. Christensen A.P., Garrido L.E., Guerra-Peña K., Golino H. Comparing community detection algorithms in psychological data: a Monte Carlo simulation. 2020:1–25.
29. Ren R., Shao J., Bishop A.N., Zheng W.X. Modeling and detecting communities in node attributed networks. IEEE Trans Knowl Data Eng 2022. http://arxiv.org/abs/2101.03280
30. Ghasemian A. Limits of model selection, link prediction, and community detection. PhD Thesis, University of Colorado at Boulder; 2019.