Review of Community Detection Approaches in Social Networks using Bayesian Method and Graph Theory

Harish kumar Shakya[1]
Department of Computer Engineering
Indian Institute of Technology (BHU), Varanasi
hkshakya.rs.cse@itbhu.ac.in[1]

Bhaskar Biswas[2]
Department of Computer Engineering
Indian Institute of Technology (BHU), Varanasi
Bhaskar.cse@itbhu.ac.in[2]

Abstract -- **Online social networks create significant challenges to computer scientists, physicists, and sociologists alike, for their massive size, fast evolution, and uncharted potential for social computing. One particular problem that has interested us is community identification. In this review, we focus on "community detection in social networks" through different approaches and techniques mainly Bayesian theorem and graph theory. At last, the authors point out some further research directions in SNA.**

Keywords- social networks, community detection, graph theory, Bayesian theorem.

1. INTRODUCTION

Recently, research on social network has become a hotspot in interdisciplinary subjects, Social networks actually can be performed as a abstract of a complex network, where a node represents an individual or a component in such systems, and an edge represents natural or artificial relationships.

Many systems are actually organized as a complex network [1, 2]. Technological networks, such as WWW World Wide Web, electric networks, global transportation networks and delay-tolerant networks [3, 5]; biological networks, such as epidemic networks [5,6], metabolism networks[7,8] and ecological networks[9,10] and social networks, such as scientists collaboration networks, online community networks and large organization networks[11,12]. Thus the theory on complex network is becoming more and more applicable and critical. The research methods applied in social network research mainly rely on graph theory, statistical mechanics and social network analysis. Yet a real social network always contains large quantity of nodes with complex structures, both of which boost the complexity of space and time during the network learning process, this tough problem has become the key challenge for researchers. To solve the problem of complexity, A. Lancichinetti et al [13], found that some nodes connect with each other so tightly that they can form a community, between each community the connection seems to be comparatively sparse. So dividing a network into some community in which the connections are tight and out of which are sparse, can help us decrease the complexity of whole network and improve our understanding of the structure, dynamic progress and function of the networks, since community could represent the potential relationships in the networks [14]. To detect community in information network, such as a world wide web which contains about 10^9 nodes, may inform us to divide the network into some small community, thus with higher efficiency during data mining and knowledge discovering [17]. So community detection has a basic research in social network.

In recent 10 years, researchers have made much exciting and important progress in community detection methods. Community detection methods mainly focus on two kind of networks, the one is common networks with positive connections only (weights of edge are positives), and the other is symbol networks, with both positive connections and negative connections within edges. In this review, we mainly discuss the first one. Partition algorithms in community detection mainly rely on two major aspects. One is graph theory. Its main idea is to divide a given network into some sub graphs within the same nodes, and the connections among each sub graphs are sparse. Graph partitioning methods belong to the area of computer science. Several famous methods are based on graph partitioning theory, such as Kernighan-Lin algorithms [25], spectrum equally division method based on the Laplace Eigen values [27, 28], clique percolation methods [16, 29] and W-H fast partitioning methods [30]. K-L algorithm is an equally partition methods based on greedy algorithms. The major drawback of K-L method is the size of two communities must be given, or it will result in error. So the application of K-L is limited when deal with real network problems. Spectrum partition methods need the accurate number of sub graphs before partition, since it only process the even number of sub graphs, and its stopping criterion is fuzzy.

The second kinds of methods are hierarchical clustering methods, which are mainly applied to analyze the similarity or intensity of connections between each node. Among them agglomerative and divisive are two major methodologies. For each pair i, j of nodes in networks, one calculate a weight $W_{i,j}$ which measures how closely the nodes connected. Starting from the set of all nodes with no edges, connections are iteratively added between pairs of nodes in order of decreasing the weight. Then nodes are grouped into larger and larger communities, and the trees built up to the root, which represents the whole network. Methods of this kind are called agglomerative, such as the methods based on measuring similarity [32]. The typical instance of agglomerative methods is Newman fast algorithms.

For divisive algorithms, the order of construction of the tree is reversed, one starts with the whole graph and iteratively cuts the edges, thus dividing the network progressively into smaller and smaller disconnected sub networks identified as the communities. The crucial point in a divisive algorithm is the selection of the edges to be cut, which have to be those connecting communities and not those within them, such as between ness methods [3,26,31].GN methods belong to divisive methodologies [26],Still GN methods have some troubles. It can hardly recognize the function of a community according to the topological structures. Additionally, keep the number of communities unknown, the stop criterion seems to be fuzzy. Some novel methods has been developed. Among them some are used to improve the time complexity [34, 36], some are focused on overlapping of communities [13, 37, 38], and some are majorly developed to deeply understand hierarchical structure of networks [39, 41].

This review is organized as following structure: After this brief introduction now we describe the organization of this review paper. In section II[nd] Bayesian method and all the graph theory methods have been described further of in this sub section II[nd] , we try to summaries the work of different researcher on community detection by using Bayesian method, in section III[rd] we conclude the paper also given the future direction.

II. Methods & Approaches

In social networks, extract meaningful communities and analyze their evolution with the help of some methods and techniques.

A. Generative Model

Whole approach includes the algorithm of Ioannis et al [40], consider the generative graphical model of Fig.1.The observed variable v_{ij} denotes the nonnegative count of interactions between two individuals i, j in a weighted undirected network with adjacency matrix $v \in R_+^{N*N}$ In the community detection context, we assume that there are a number K of 'hidden' classes of nodes in the network that affect v_{ij} . Thus define the allocations of nodes to communities as latent (unobserved) variables that allow us to explain the increased interaction density in certain regions of the network: the more two individuals interact the more likely they are to belong to the same communities, and vice versa.

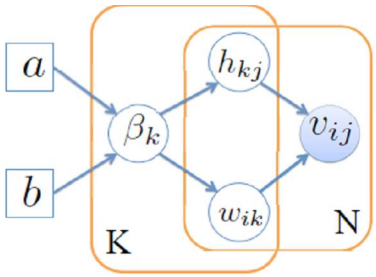

Fig. 1(Colour online) Graphical model [40]

Assume that the pair-wise interactions described in V are influenced by an unobserved expectation network $\bar{v}$ where each $\overline{v_{ij}}$ denotes the expected number of interactions (or expected link weight) that take place between i and j. The expectation network is composed of two nonnegative matrices $W \in R_+^{N*K}$ and $H \in R_+^{K*N}$ so that $\bar{v} = WH$ .Hence the model each interaction v_{ij} as drawn from a Poisson distribution with rate $\overline{v_{ij}} = \sum_{K=1}^{K} w_{ik} h_{kj}$. The inner rank K denotes the unknown number of communities and each element $k \in \{1,..., K\}$ in row i of W and column j of H represent the contribution of a single latent community to $\overline{v_{ij}}$. In other words, the expected number of times $\overline{v_{ij}}$ that two individuals i , j interact is a result of their mutual participation in the same communities.

In the typical community-detection setting, the value of K, which we call complexity or model order, is initially unknown. In previous work[47,48], the issue of inferring the appropriate number of communities has been addressed by performing multiple runs for various K and selecting one that yields the highest Newman modularity Q[45]. The appropriate model order arises naturally from a single run, by placing shrinkage or automatic relevance determination priors [49] with scale hyper parameters $\beta = \{\beta_k\}$ on the latent variables w_{ik} , h_{kj} , as presented [46]. By starting with a large K (say N, which is the maximum possible number of communities), the effect of these priors is to moderate complexity by 'shrinking' close to zero irrelevant columns of W and rows of H that do not contribute to explaining the observed interactions V. This is achieved by placing a distribution over the latent variables w_{ik} , h_{kj} , whose expectation approaches zero unless non-zero values are required by the data. This approach avoids the computational load of multiple runs and is free of the resolution bias problems [50] of modularity. Based on the graphical model of Fig. 1, where the distribution of β_k is parameterized by fixed hyper-hyper parameters a and b & express the joint distribution over all variables as [40]:

$$p(V, W, H, \beta) = P\left(V \middle/ W, H\right) P\left(W \middle/ \beta\right) P\left(H \middle/ \beta\right) P(\beta) \quad (1)$$

Hence the posterior over model parameters given the observations is:

$$P(V, W, H, \beta) = \frac{P\left(V \middle/ W, H\right) P\left(W \middle/ \beta\right) P\left(H \middle/ \beta\right) P(\beta)}{P(V)} \quad (2)$$

B. Posterior-based cost function

The aim of this function to maximize the model posterior given the observations, or equivalently, to minimize the negative log posterior, which may be regarded as an energy (or error) function U. Nothing that $p(v)$ is a constant w.r.t. the inference over the model's free parameters, we hence define [40]:

$$U = -\log p\left(V \middle/ W, H\right) - \log p\left(W \middle/ \beta\right) - \log p\left(H \middle/ \beta\right) - \log(\beta) \quad (3)$$

Where the first term is the log-likelihood of our data, derived from the probability $P\left(V \middle/ W, H\right) = p\left(V \middle/ \bar{V}\right)$ of observing every

interaction v_{ij} given a Poisson rate $\overline{v_{ij}}$. Therefore express the negative log-likelihood of a single observation v_{ij} as:

$$-\log\ p(v/\bar{v}) = -v\log\bar{v} + \bar{v} + \log v! \qquad (4)$$

Using the Sterling approximation to second order, namely:

$$\log\ v! \approx v\log\ v - v + \tfrac{1}{2}\log\left(2\Pi v\right) \qquad (5)$$

Eq. (4) can be written as:

$$-\log p(v/\bar{v}) \approx v\log(v/\bar{v}) + \bar{v} - v + \tfrac{1}{2}\log(2\Pi v) \quad (6)$$

Thus the full negative log-likelihood for all the observed data is:

$$\log\ p(v/\bar{v}) =$$

$$-\sum_{i=1}^{N}\sum_{j=1}^{N}\log p(\frac{v_{ij}}{v_{ij}}) \approx \sum_{i-1}^{N}\sum_{j=1}^{N}(v_{ij}\log\frac{v_{ij}}{v_{ij}} + \overline{v_{ij}} - v_{ij} + \frac{1}{2}\log(2\Pi v_{ij})) + k, \qquad (7)$$

where k is a constant. Following [46] and similar models for probabilistic PCA[51] and ICA[52,53], then place independent half-normal priors over the columns of W and rows of H with precision(inverse variance) parameters $\beta \in R^{K} = [\beta_1, \ldots, \beta_K]$.The negative log priors over W and H are then given by[40]:

$$-\log\ p\left(W/\beta\right) = -\sum_{i=1}^{N}\sum_{k=1}^{K}\log\ \mathrm{HN}\ (0, \beta_k^{-1})$$

$$\sum_{i=1}^{N}\sum_{k=1}^{K}(\tfrac{1}{2}\beta_k w_{ik}^2) - \frac{N}{2}\log\ \beta_k + k, \qquad (8)$$

$$-\log\ p\left(H/\beta\right) = -\sum_{k=1}^{K}\sum_{j=1}^{N}\log\ HN\ (0, \beta_k^{-1})$$

$$\sum_{k=1}^{K}\sum_{j=1}^{N}(\tfrac{1}{2}\beta_k h_{kj}^2) - \frac{N}{2}\log\ \beta_k + k \qquad (9)$$

Each β_k controls the importance of community k in explaining the observed interactions; large values of β_k denote that column k of W and row k of H have elements lying close to zero and therefore represent irrelevant communities. By assuming β_k are independent we place a standard Gamma distribution over them with fixed hyper-hyper parameters a, b [54]. The negative log hyper-priors are thus:

$$-\log\ p(\beta) = -\sum_{k=1}^{K}\log\ G(\beta_k / a, b)$$

$$\sum_{k=1}^{K}(\beta_k b - (a-1)\log\ \beta_k) + k. \qquad (10)$$

The objective function U of Eq. (3) can be expressed as the sum of Eq. (7) through (10):

$$u = \sum_i\sum_j\left[v_{ij}\log(\frac{v_{ij}}{v_{ij}}) + \overline{v_{ij}}\right]$$

$$+\frac{1}{2}\sum_k\left[\left(\sum_i\beta_k w_{ik}^2\right) + \left(\sum_j\beta_k h_{kj}^2\right) - 2N\log\beta_k\right] \qquad (11)$$

$$+\sum_k(\beta_k b_k) - (a_k - 1)\log\beta_k) + k.$$

C. Parameter inference

To optimize Eq. (11) for W,V and β, follow [46,55,56] by adopting the fast fixed-point algorithm presented in [46] that involves consecutive updates of W,H, and β until a convergence measure has been satisfied (a maximum number of iterations, or a tolerance on the cost function). The solution consists of $W_* \in R_+^{N\times K_*}$ and $H_* \in R_+^{K_*\times N}$ for which $\overline{V} = W_* H_*$ represents the expectation network given our observation data V and prior assumptions. The inner rank K_* denotes the inferred number of hidden modules in the network. Based on the above, model assumes that the joint membership of two nodes in the same community raises the probability of a link existing between them. Therefore, method performs best when modules are dense, with the best-case scenario being that each community is a fully connected sub-graph [40]. Many other initial statistical models based on a Bayesian non parametric model are available for community detection in social networks. This simple model capture the idea that it is more probable for two nodes to be connected if they are in the same community, as some of the Bayesian techniques such as MCMC(Markov chain Monte Carlo) algorithms, decision theory, CRP(Chinese restaurant process), DP(Dirichlet process).

In graph theory techniques, community detection methods can be classified into two categories; the first consists of algorithms that partition a graph into communities based on topological features (such as betweeness [24] centrality etc) and then use modularity to evaluate the result. The second consists of algorithms that try to directly optimize modularity using some approximation scheme. Another categorization is based on how these algorithms construct the community hierarchy dendrogram discussed in the previous section; agglomerate methods start by considering each individual node to be a community (the leaves at the bottom layer of the dendrogram) and proceed constructively by merging partitions based on some similarity metric. On the other hand, divisive methods start by the whole graph (parent node of the dendrogram) and break it down into different communities until they reach a point where each node where we have N communities with a single node at each. It is important to note that purely agglomerate or divisive methods do not allow overlapping communities [42].

D. Agglomerate

Given a collection of nodes V along with their edges D, we define some kind of similarity metric initially between nodes and then between groups, in order to successively merge them together and form the community hierarchy dendrogram. Initially consider that each individual node is a separate group. During the first iteration we group together each pair i; j of nodes with the highest similarity x_{ij} .Method proceed by merging pairs of similar groups until we reach the top level of the dendrogram where the whole network is a single group. This technique is also called hierarchical clustering [18, 21]. Unsurprisingly, the performance of these methods heavily relies on the similarity metric x_{ij} itself and some of them are presented in the paragraphs below.

1).Distance-based structural equivalence: [18, 42] is based on the concept that similar nodes q have same neighbors even if they are not adjacent themselves, thus

$$x_{ij} = \sqrt{\sum_{k \neq i,j} \left(A_{ik} - A_{jk}\right)^2}$$

2).Pearson correlation: [18, 42] is based on the similar concept of structural equivalence, but instead of a distance metric it measures the correlation between rows (or columns) of the adjacency matrix A. So $x_{ij} = \frac{\sum_k (A_{ik} - \mu_j)(A_{jk} - \mu_j)}{N \sigma_i \sigma_j}$

Where $\mu_i = \frac{1}{N} \sum_j A_{ij}$ *and* $\sigma_i = \sum_j (A_{ij} - \mu_i)^2$.

3).Donetti-Munoz method: [21, 42] utilizes a matrix L called the Laplacian, defined by inverting the sign of each element of A and then setting $A_{ij} = k_i \forall_i \in \{1,\ldots,N\}$. The idea is to use the values of D eigenvector components of L and project each individual to a D-dimensional space. Then, by defining a distance based similarity measure such as angular of Euclidean distance, So apply hierarchical clustering to produce a dendrogram of possible community partitions. The main drawback of this method is that the number D of eigenvectors use is not known a prior and the performance of this method relies on choosing a proper value of it.

4).Capocci method: [23, 42] utilizes the normal matrix N of a is defined by dividing each A_{ij} by the sum of the elements of A. Then, similarly using the eigenvectors of N it projects (similarly to Donetti-Munoz) the nodes to a high though easy to implement and relatively fast, it yields poor results in most real and artificial networks.

E. Divisive

Method starts by allowing for the whole network as a single community and precede deconsructively by breaking it down into smaller ones. Some very popular and efficient community detection methodologies fall into this category and they are presented below:

1).Newman-Girvan method: [43] was first introduced along with the concept of modularity in the influential paper. It utilizes the measure of edge betweeness, usually defined by the number of shortest paths between any pair of nodes that run through that given edge. Other (and less efficient) formulations are discussed in. The instinct behind this measure is that communities are linked together by a small number of edges that have significantly higher traffic than the others. The algorithm consists of ranking each edge in the network based on that edge centrality measure and by removing the most popular one. After the removal all edge betweenesses are re-calculated and another edge is removed. The algorithm iterates as shown above until some part of the network is isolated thus we have community separation. Apply the same procedure for each sub graph thus building the community hierarchy dendrogram, from which we select the layer with represents the partition with the highest value of modularity. The algorithm, although very popular, is relatively slow and performs poorly against densely connected graphs.

2).Spectral partitioning: is another divisive method from Mark Newman [8, 42] that builds the community hierarchy dendrogram by performing bisections of each community. It utilizes the modularity matrix B defined by $B_{ij} = A_{ij} - \frac{k_i k_j}{2M}$, which has the property that the elements of each of its rows and columns sum to zero. So given the full graph, each bisection is performed by computing the leading eigenvector of B and by dividing the vertices into two groups according to the signs of the elements of this vector. For each subgroup g we perform the same division scheme but with an updated modularity matrix $B^g_{ij} = B_{ij} - \delta_{ij} \sum_{k \in g} B_{ik}$ .The method is relatively fast and has the excellent property of identifying indivisible groups; a community cannot have further divisions with positive modularity if all the Eigen values of the modularity matrix are non-positive. It also provides good results in most real and artificial problems.

3).Extremal optimization: [42, 44] is a method for direct approximation of the modularity function Q. Given a network and a community partition, it breaks down its modularity Q into the individual contributions from each node. So given a node-I belonging to community-r, its contribution to the overall modularity is calculated from:

$$\lambda_i = \frac{k_{r(i)}}{k_i} - a_{r(i)}$$

where $k_{r(i)}$ is the number of neighbors of i that belong to the same community as i (intra-community degree) and $a_{r(i)}$ is the r-element of the a vector used . The quantity λ_i is normalized to the interval [-1, 1] to allow comparisons between individual nodes and the overall modularity can be easily recovered by

$$Q = \frac{1}{2M} \sum_{i=1}^{N} k_i \lambda_i$$

The algorithm starts by "creating a random bisection of the original graph. Then it calculates the lambda for each node and moves the least contributing nodes to another group. Due to the random initialization scheme and to avoid local maxima we usually select randomly one of the n-least contributing nodes to change its membership. After a number of moves, if the modularity does not improve so proceed recursively by applying the algorithm to each partition. From the resulting dendrogram we select the partition with the highest modularity". Although the method is heavily initialization dependent it provides state-of-the-art results in most problems.

F. *The Potts method* [35]

This is a community detection algorithm inspired by Statistical Mechanics. The model assumes that a network is a system of spins that can have q different states. Thus, each

node-i can take have a spin value $\sigma_i \in \{1,\ldots,q\}$ (that basically accounts for its community membership, C = q) and the interaction energy between spins is given by $-J_{ij}$ (where J = A) if the spins are in the same state and zero if they are not. Finding the appropriate partition for the network equals to finding the ground state minimum) of the Hamiltonian [42]:

$$H = -J_{ij} \sum_{i,j \in N} \delta_{\sigma_i \sigma_j} + \gamma \sum_{s=1}^{q} \frac{n_s (n_s - 1)}{2}$$

where n_s is the number of spins (nodes) in state (community) s, J_{ij} the interaction strength (given by the adjacency or weight matrix A), γ a positive parameter and the Kronecker $\delta_{\sigma_i \sigma_j}$ is 1 if i, j have the same spin (belong to the same community) or 0 otherwise. The above equation reflects the two competing forces in our system; the first term of the summand favours a homogeneous distribution of spins (minimum for all i, j in the same community) while the second term favours a uniform distribution of spins across nodes. To find the ground state (minimum) of the above system a Monte Carlo single spin flip heat-bath algorithm along with simulated annealing. The method is fast and provides very competitive results for most real-world and artificial datasets.

III. Conclusion & Future work

Since the social networks and community detection problem in particular are subjected to divers' field, so many approaches can easily be applied to the said problems. This review can be extended for community detection problem in using other data mining approaches, game theory & soft computing techniques.

REFERENCES

[1]. Fortunato, S. Community detection in graphs [Journal Article]. Physics Reports, 2010. 486(3-5): p. 75-174.

[2]. Donetti, L. and M.A. Munoz. Detecting network communities: a new systematic and efficient algorithm [Journal Article]. Journal of Statistical Mechanics: Theory and Experiment, 2004. 2004: p. P10012.

[3]. Radicchi, F., C. Castellano, et al. Defining and identifying communities in networks [Journal Article]. Proceedings of the National Academy of Sciences of the United States of America, 2004. 101(9): p. 2658.

[4]. Broder, A., R. Kumar, et al. Graph structure in the web [Journal Article]. Computer networks, 2000. 33(1-6): p. 309-320.

[5]. Moore, C. and M.E.J. Newman. Epidemicsand percolation in small-world networks [Journal Article]. Physical Review E, 2000. 61(5): p. 5678.

[6]. Pastor-Satorras, R. and A. Vespignani. Epidemic spreading in scale-free networks [Journal Article]. Physical review letters, 2001. 86(14): p. 3200-3203.

[7]. Jeong, H., B. Tombor, et al. The large-scale organization of metabolic networks [Journal Article]. Nature, 2000. 407(6804): p. 651-654.

[8]. Ravasz, E., A.L. Somera, et al. Hierarchical organization of modularity in metabolic networks [Journal Article]. Science, 2002. 297(5586): p. 1551.

[9]. Montoya, J.M. and R.V. Sol?. Small world patterns in food webs [Journal Article]. Journal of theoretical biology, 2002. 214(3): p. 405-412.

[10]. Garlaschelli, D., G. Caldarelli, and L. Pietronero. Universal scaling relations in food webs [Journal Article]. Nature, 2003. 423(6936): p. 165-168.

[11]. Redner, S. How popular is your paper? An empirical study of the citation distribution [Journal Article]. The European Physical Journal B-Condensed Matter and Complex Systems, 1998. 4(2): p. 131-134.

[12]. Newman, M.E.J. The structure of scientific collaboration networks [Journal Article]. Proceedings of the National Academy of Sciences of the United States of America, 2001. 98(2): p. 404.

[13]. Lancichinetti, A., S. Fortunato, and J. Kert?sz. Detecting the overlapping and hierarchical community structure in complex networks [Journal Article]. New Journal of Physics, 2009. 11: p. 033015.

[14]. Rosvall, M. and C.T. Bergstrom. Maps of random walks on complex networks reveal community structure [Journal Article]. Proceedings of the National Academy of Sciences, 2008. 105(4): p. 1118.

[15].Gulbahce, N. and S. Lehmann. The art of community detection [Journal Article]. Bioessays, 2008. 30(10): p. 934-938.

[16]. Palla, G., I. Der?nyi, et al. Uncovering the overlapping community structure of complex networks in nature and society [Journal Article]. Nature, 2005. 435(7043): p. 814-818.

[17]. Capocci, A., V.D.P. Servedio, et al. Detecting communities in large networks [Journal Article]. Physica A: Statistical Mechanics and its Applications, 2005. 352(2-4): p. 669-676.

[18].Santo Fortunato and Claudio Castellano - Community Structure in Graphs,chapter of Springer's Encyclopedia of Complexity and System Science (2008)

[19]. Hui Bo. Research on community detection and its applications [Thesis].2009, Beijing Jiaotong University.

[20]. Gibson, D., J. Kleinberg, and P. Raghavan. Inferring web communities from link topology [Conference Proceedings]. 1998: ACM.

[21].Loca Donetti and Miguel A Munoz - Detecting network communities: a new systematic and efficient algorithm, Journal of Statistical Mechanics: Theory and Experiment (2004)

[22]. Adamic, L.A. and E. Adar. Friends and neighbors on the web [Journal Article]. Social Networks, 2003. 25(3): p. 211-230.

[23]. Cappocci A, Servedio VDP, Caldarelli G, Colaiori F - Detecting communities in large networks, Physica A. Vol 352, No 2-4, pp 669-676 .

[24]. L.C. Freeman - A set of measures of centrality based upon betweeness, Sociometry,40 (1977), pp. 35-41.

[25]. Holme, P., M. Huss, and H. Jeong. Subnetwork hierarchies of biochemical pathways [Journal Article]. Bioinformatics, 2003. 19(4): p. 532.

[26]. Girvan, M. and M.E.J. Newman. Community structure in social and biological networks [Journal Article]. Proceedings of the National Academy of Sciences of the United States of America, 2002. 99(12): p. 7821.

[27]. Fiedler, M. Algebraic connectivity of graphs [Journal Article]. Czechoslovak Mathematical Journal, 1973. 23(2): p. 298-305.

[28]. Pothen, A., H.D. Simon, and K.P. Liou. Partitioning sparse matrices with eigenvectors of graphs [Journal Article]. SIAM J. MATRIX ANAL. APPLIC., 1990. 11(3): p. 430-452.

[29]. Palla, G., I.J. Farkas, et al. Directed network modules [Journal Article]. New Journal of Physics, 2007. 9: p. 186.

[30]. Wu, F. and B.A. Huberman. Finding communities in linear time: a physics approach [Journal Article]. The European Physical

Journal B-Condensed Matter and Complex Systems, 2004. 38(2): p. 331-338.
[31]. Tyler, J.R., D.M.Wilkinson, and B.A. Huberman. Email as spectroscopy: Automated discovery of community structure within organizations [Journal Article]. Arxiv preprint cond-mat/0303264, 2003.
[32]. Newman, M.E.J. Fast algorithm for detecting community structure in networks [Journal Article]. Physical Review E, 2004. 69(6): p. 066133.
[33]. Clauset, A., M.E.J. Newman, and C. Moore. Finding community structure in very large networks [Journal Article]. Physical Review E, 2004. 70(6): p. 066111.
[34]. Wang, X., G. Chen, and H. Lu. A very fast algorithm for detecting community structures in complex networks [Journal Article]. Physica A: Statistical Mechanics and its Applications, 2007. 384(2): p. 667-674.
[35].Jorg Reichardt and Stefan Bornholdt - Detecting Fuzzy Community Structures in Complex Networks with a Potts Model, Physical Review Letters Volume 93, Number 21 (2004)
[36]. Medus, A. and C. Dorso. Alternative approach to community detection in networks [Journal Article]. Physical Review E, 2009. 79(6): p. 066111.
[37]. Shen, H., X. Cheng, et al. Detect overlapping and hierarchical community structure in networks [Journal Article]. Physica A: Statistical Mechanics and its Applications, 2009. 388(8): p. 1706-1712.
[38]. Gregory, S. A fast algorithm to find overlapping communities in networks [Journal Article]. Machine Learning and Knowledge Discovery in Databases, 2008: p. 408-423.
[39]. Fortunato, S. and M. Barth?lemy. Resolution limit in community detection [Journal Article]. Proceedings of the National Academy of Sciences, 2007. 104(1): p. 36.
[40]. Ioannis Psorakis, Stephen Roberts, and Mark Ebden, Overlapping Community Detection using Bayesian Nonnegative Matrix Factorization[Journal Article]. Physical Review E,2011-APS.
[41]. Sales-Pardo, M., R. Guimera, et al. Extracting the hierarchical organization of complex systems [Journal Article]. Proceedings of the National Academy of Sciences, 2007. 104(39): p. 15224.
[42]. Psorakis ioannis, Stephen Roberts, Ben Sheldon," Community Detection Algorithms: a comparative evaluation on artificial and real-world networks" [Thesis]. 2010, University of Oxford.
[43].M.E.J Newman and M. Girvan - Finding and evaluating community structure in networks, Phys. Rev. E 69, 026113 (2004).
[44]. Jordi Durch and Alex Arenas Community detection in complex networks using extremal optimization. PHYSICAL REVIEW E72, 027104 (2005).
[45].M.E.J. Newman and M. Girvan, Phys. Rev. E, 69, 026113(2004).
[46]. V. Tan and C. Fevotte, in SPARS09 - Signal Processing with Adaptive Sparse Structured Representations (2009) pp. 1–19.
[47]. R.-S. Wang, S. Zhang, Y. Wang, X.-S. Zhang, and L. Chen,Neurocomputing, **72**, 134 (2008), ISSN 0925-2312, machine Learning for Signal Processing (MLSP 2006) / Life System Modelling, Simulation, and Bio-inspired Computing (LSMS 2007).
[48]. S. Zhang, R.-S. Wang and X.-S. Zhang, Phys. Rev. E, 76, 046103 (2007).
[49]. D. J. C. MacKay, Network: Computation in Neural Systems, 6, 469 (1995).
[50]. S. Fortunato and M. Barthelemy, Proceedings of the National Academy of Sciences, 104, 36 (2007).
[51]. M. E. Tipping and C.M. Bishop, Journal of the Royal Statistical Society. Series B (Statistical Methodology), 61, 611 (1999).
[52]. R. A. Choudrey and S. J. Roberts, Neural Computation, 15, 213 (2003).
[53]. S. J. Roberts and R. A. Choudrey, Pattern Recognition, 36, 1813 (2003), ISSN 0031-3203.
[54]. J. M. Bernardo and A. F. M. Smith, *Bayesian Theory* (JohnWiley, 1994).
[55]. D. D. Lee and H. S. Seung, Nature, 401, 788 (1999).
[56]. M. W. Berry, M. Browne, A. N. Langville, V. P. Pauca, and R. J. Plemmons, Computational Statistics and Data Analysis, 52, 155 (2007).