THINK INDIA JOURNAL

ISSN: 0971-1260
Vol-22-Issue-16-August-2019

A Review on Community Detection Algorithms In Social Networks

Puneet Kumar (puneetlpu1@gmail.com)
Assistant Professor, School of Computer Science & Engineering
Lovely Professional University, Jalandhar Punjab

Abstract— **There are many real networks that are closely formed according to a community network. Many researchers have devoted their research efforts to develop algorithms and methods that can efficiently find the hidden pattern in the network which is called community. As there are different kinds of networks having different level of complexity and each may contain different properties, each algorithm in literature holds some of these properties and has different definition of community. As per the definitions provided and algorithms used, only part of the features of real communities gets extracted. Objective of this paper is to find different community detection techniques that are already available, analyze them and find the problem in these algorithms. Objective of paper is to categorize the different community discovery methods on the basis of properties these algorithms contain. The classification is also helpful in further research in community detection.**

Keywords— community, modularity, Social Networks, similarity, betweenness.

I. INTRODUCTION

With the growth of technology application of graph data structure is also grow and now a day there is no any area in computer science where graphs are not used .Graphs are widely used in data mining, social networking sites, scheduling the processes by operating system, in networking for finding shortest path and lot. These graphs contain vertices that belong to a user and edges shows relationship between the different users. These nodes combine with other nodes to form a community. So, it can be called that these networks have large number of communities. These social networks may be sliced into different parts of communities that have higher links inside each part or group, but lesser links to outside of the other part. The inference of communities in these large social networks is a problem of discussion in this area. Therefore a large number of researchers have proposed techniques to detect communities in the related work.

Most of these Community inference techniques are static and run on static networks. These networks are not always static because they evolve very frequently. New users may join the network or the old one may get removed from the network or graph within seconds. In other words, old connections in the map may be removed or new connections may be inserted to the graph with time. These types of networks may call as highly dynamic in nature. For example, social media websites like Twitter, Facebook and LinkedIn are highly dynamic in nature. Insertion or removal of a connection or a user from a graph or social network which has billions of connections may seem insignificant, but when insertion or removal of a connection or a node happen very fast, then difference in existing community structure of the graph become very important. This modification in the network structure increases the need of again running the community detection methods in the network to detect newly formed communities. This requirement arises fast and again and again and creates problem in the detection of communities. So this problem needs a fast solution for detecting the communities in networks which is dynamic.

The solution that we think for this Detection of communities in dynamic and very large networks problem is to do processing of static community inference techniques defined in related work many times by the different researchers to find new community structure when a change occur in the network. But, this solution will take huge time as we have to run the algorithm frequently whenever a change occur in the network i.e. whenever a new node get added or the old user leave the network. A more efficient, less complex and less time taken solution is to execute the technique for community detection not from the starting point but from the point where we have stored in a log file after the previous execution, in which information about the previous community results stored. Changing old discovered community structure on the place of trying to find new communities from starting every time the network get changed consumes very less time as compare to the previous one and algorithm become more efficient.

In this paper we have surveyed different algorithm already proposed by different researchers, how these algorithms are different from each other and how these algorithms can be used on dynamic social network. Algorithms detect the community on the basis of density, Vertex similarity, actions of the user and influence propagation. These all algorithms detect the communities by considering different kind of parameters. In this paper we described all the proposed algorithms on the basis of community detection basis and parameters used by the algorithm.

II. RELATED WORK

There are many algorithms proposed by different researchers for community detection. Some researchers proposed core methods which work on quality functions which divide network into communities such as Newman's technique, clustering algorithms, and many more. While, it is also possible that node in social network can belong to more than one community. That's why many researchers have moved their study path to designing different techniques of overlapping community inference. With the growth of social networks, the networks are getting becoming more complex. Previous techniques of community detection are depending on dividing the network into multiple parts called groups which costs more when graph is large and complex. So researchers moved to do research on local community detection. In this, we will try to introduce algorithms related to community detection. The algorithms are classified into different classes which show the evolution of the research on community detection.

A. Community Detection Algorithms for static network

Community Detection on the basis of partitioning

Partitioning based techniques have used in the past by the researchers to divide the network into groups in such a way that these groups have less connections. The Kernighan Lin technique for network dividing was earliest technique to partition a network. It divides the nodes of a network into different subsets of provided sizes to minimize the cost on edge cuts. A limitation of this technique is before execution number of groups needs to define. The technique is good as worst case time complexity is $O(n^2)$. Newman decreases the mostly used maximum likelihood method to detect community to a group of different techniques, where every technique is providing solution to partitioning problem [47]. It explains that two major needed community detection techniques or its degree corrected variant may be same on various different problems of graphs partitioning. This is shown by using the Laplacian spectral partitioning method to find Community detection [26].

Clustering based Community Detection

Community detection algorithms are mainly used for finding the communities. Clustering is foundation of community detection algorithms. Girvan and Newman had played a main role in clustering based community detection techniques. Girvan and Newman proposed a technique which depends on betweenness of edge for a network with a connection having no direction and weight. The technique based on communities which have most connections, and communities are constructed by deleting these connections from the network [18].

The Girvan Newman technique is extended by many researchers and used on various different graphs. Rattigan et al. proposed the indexing technique to improve complexity of Girvan Newman technique [46]. Chen et al. improved the Girvan Newman technique to partition network which is weighted [13]. Pinney et al. proposed a technique that acquire the Girvan Newman technique for the dividing the graph [58].

Newman and Girvan given a parameter called modularity to check the quality of partitions or groups. The parameter of modularity accepted and examined by authors to check the goodness of Community detection technique where high modularity means that better community.

Newman has worked on maximizing of modularity so that forming different communities tends to highest modularity. The change after joining two communities in modularity defined as $Q = e_{ij} + e_{ji} - 2a_ia_j = 2(e_{ij} - a_ia_j)$ and it can be find in fixed time and therefore it is faster to execute as compare to other techniques like Girvan Newman technique.

Greedy optimization technique of modularity is used by Clauset et al. to find groups from large networks. From a network with k connections and r nodes, the technique has complexity of (km log r) , where 'm' represents the depth of the dendrogram [14]. For sparse networks, the complexity of algorithm is ($r \log_2 r$).

An iterative algorithm introduced by Blondel et al. called Louvain method. In this algorithm two phases exists. In first phase, vertices are assigned in different groups, and then, modularity gain of node moving from one community or group to other evaluated[12]. A node is assigned to a particular community having maximum modularity gain in this movement. In the second part communities originated in the first part are act as node or user, and the weight of connection is calculated. The technique reduces the time complexity of the Girvan Newman method. Time complexity of this is O(m) [12]. Simulated annealing is used by Guimera et al. for optimizing the modularity and he shown that determining the energy of a spin system at ground state and calculating the modularity of a network is same. Additionally, the researcher showed that modular network get rise due to the stochastic network model and fluctuations [59]. To improve modularity Zhou et al. attempted simulated tempering which introducing the new method of inter connection and intra connection.

Table 1 Shows different clustering based algorithms form community detection with detail of Reference Number, approach used, parameters used to check goodness, Tools used and limitation of algorithm.

TABLE 1. COMMUNITY DETECTION ON THE BASIS OF CLUSTERING.

Reference Number	Approach	Parameters	Tools Used	Complexity
[35]-[38]	Maximization of modularity	Eigen Value, Modularity and Eigen vector	MapReduce	$O(n^2 \log n)$
[18]	Divisive clustering	Edge betweenness	MapReduce	$O(n^3)$
[14]	Modularity based on Greedy optimization	Vertices, Edges, modularity	C Compiler with MakeFile	$O(n \log^2 n)$
[12]	Hierarchical clustering	Edges, Nodes, modularity	Gephi / NetworkX	O(n)
[59]	Optimizing Modularity by simulated annealing	No. of partitions, modules, modularity	Matlab	$O(n^2 \log^2 n)$
[60]	Optimizing Modularity by external optimization	Links, No. of nodes, modularity, degree,	Radatools	$O(n^2 \log^2 n)$
[39]	Clustering based on Entropy centrality	Markov process of transition probability matrix	C++ Compiler	$O(nk^2)$
[61]	Random walk and Consensus clustering	length of random walks, similarity matrix	MapReduce	$O(n^2 \log n)$

Detection of Communities in Dynamic Networks

Dynamic networks like networks on social networking sites evolve exponentially. There are very less algorithms or research done on this field as compare to static network.

Bansal et al. have proposed two approaches for the data which depends on online community detection which changes very fast with time and on the community in which changes are not very fast and known before called offline community detection [62]. Depending on social interactions between the different nodes in the network Wolf et al. have proposed computational and mathematical formulation to analyze the dynamic communities [9]. Tantipathananandh et al. has formulated three functions of cost namely, g-cost, i-cost and c-cost and made assumptions about the relationships and behavior of an individual in the network and also deployed a heuristics based and Graph coloring based approach [63].

Lin et al. proposed a framework named as FacetNet that search the dynamic communities which are evolving in the network and covers to a minimal solution [27]. Palla et al. have proposed CPM algorithm which he have proposed after doing an experiment on two and large networks of phone call and collaboration network to find the dependencies [41]. Auto correlation function and a stationary parameter is used by them to find

overlapping between different stages of a community.

Representation of the dynamic graph as aggregation graph of time is done by Greene et al. and they have given a heuristic algorithm to discover the communities from the dynamic network. These time step communities act as the dynamic communities for a particular time. The algorithm given by them is first of all try to discover static communities from the network and in subsequent iteration it search for dynamic communities. They used standard dataset to do this experiment [19].

Bansal et al. gave the algorithm which is based on deletion and insertion of connections in the graph [62]. Clauset et al. proposed a community detection algorithm which worked on modularity based technique of greedy agglomerative [14]. Louvain method is improved by this technique to detect the community in dynamic networks [21]. Main functionality of algorithm is it uses communities of t-1 time to detect the communities of time t. Power law distribution is used by an approximation algorithm proposed by Dinh et.al. which named as Adaptive Approximation Algorithm for community detection [15].Disjoint community structure in dynamic social networks identified by Nguyen et al. Quick Community Adaptation (QCA), framework based on adaptive modularity is proposed. The growth of dynamic graph is traced by this technique [15].A two steps community detection approach has been proposed by Takaffoli et al. In first step using weighted bipartite graph different communities are matched which are extracted on different instances. Thenin next step a meta community derived from the similarity based pattern find in first step [64].

TABLE 2. COMMUNITY DETECTION IN DYNAMIC NETWORK

Reference Number	Approach	Parameters	Tool Used
[9]	Mathematical framework	meta group records	C Compiler
[62]	Greedy agglomerative	Edges, Nodes	C Compiler
[63]	Heuristics,Coloring problem,	C-cost, individual cost, group cost,	MatLab
[27]	Iterative algorithm	Temporal cost and Snapshot cost	FacetNet
[41]	Joint graphs	Stationary parameter, Auto-correlation	GA-Net
[19]	Step communities	Time t	MapReduce
[21]	Louvainalgorithm with dynamicity	Time t	C++ Compiler
[15]	Maximizing the modularity fordynamic networks	Modification in snapshot of graph	MapReduce
[15]	Adapting quick community	Edges, Nodes	GA-Net+
[64]	Events like split, survive, dissolve,combine, and form	Similarity in community	MatLab
[65]	Quasi clique by clique, Nano communities	snapshot cost ,Temporal cost	MOGA-Net
[66]	Spectral clustering	Snapshot cost, Temporal cost	RanCon
[67]	Multi-objective genetic algorithm	NMI, Community score	ACSM
[65]	Two-stage clustering	General similarity, Cosine similarity	MatLab

III. APPLICATIONS OF COMMUNITY DETECTION

With the exponential growth in social networking Sites and networks it becomes very difficult to understand and imagine the network. Communities can act as description of the whole network. Detection of these communities can play a vital role in the various fields as described below:

- *Social Media Communities*

With the growth in social networking sites and users on the sites focus of the site is expanding and modifying. Community detection techniques are used to do the sentiment analysis of users on a particular tweet or re-tweet. These algorithms are used to analyze the communities to do marketing of the products to a particular community which is more interested in that particular product. These algorithms are used to find the behavior of users or interest of user in different fields.

- *Analysis in Citation Network*

Citation networks are created by researchers which show relation between a research publication and different researchers which belong to that paper or topic. In this network paper represent a single topic on which all the researchers are working. Detection of these communities helps in finding research topics and number of researchers which are working on that topic.

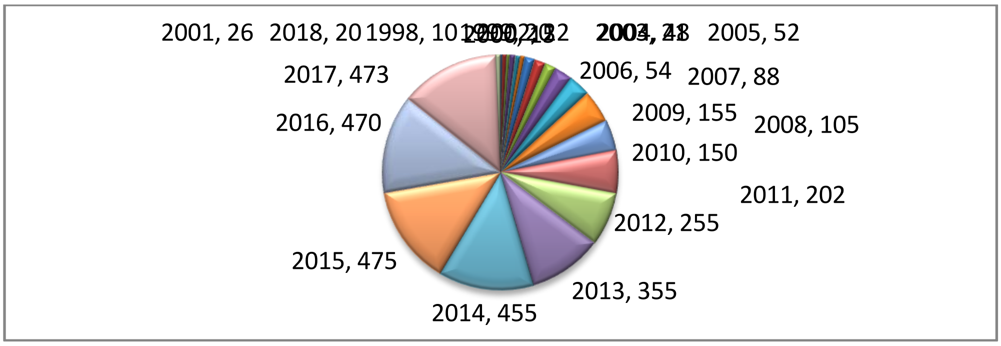

Fig. 1. Shows distribution of articles published on Community Detection in last 21 Years.

IVCONCLUSION

With the growth of Social Networks this area become area having very vast potential of research to find more better and fast algorithms for community detection. In this paper basic partition based, clustering based algorithms for static networks and some algorithms available in literature for detecting the communities in dynamic network are presented. Use of community detection algorithms provides the need of research in this field. Information provided by community detection may be helpful in development, commercialization and in education. Detail of different parameters and measures used by different algorithms is presented in this paper.

REFERENCES

[1] R. Aktunc and I. H. Toroslu, “A Dynamic Modularity Based Community Detection Algorithm for Large-scale Networks : DSLM,” pp. 1177–1183, 2015.
[2] A. Altaher and A. H. Osman, “An intelligent approach for predicting social media impact on brand building,” *J. Theor. Appl. Inf. Technol.*, vol. 95, no. 17, pp. 4097–4102, 2017.
[3] H. Alvari, A. Hajibagheri, G. Sukthankar, and K. Lakkaraju, “Identifying community structures in dynamic networks,” *Soc. Netw. Anal. Min.*, vol. 6, no. 1, 2016.
[4] M. Arab and M. Afsharchi, “Journal of Network and Computer Applications Community detection in social networks using hybrid merging of sub-communities,” *J. Netw. Comput. Appl.*, vol. 40, pp. 73–84, 2014.
[5] A. Arenas and L. Danon, “P HYSICAL J OURNAL B,” vol. 380, pp. 373–380, 2004.
[6] K. Asmi, D. Lotfi, and M. El Marraki, “Large-scale community detection based on a new dissimilarity measure,” *Soc. Netw. Anal. Min.*, vol. 7, no. 1, pp. 1–10, 2017.
[7] T. Aynaud, E. Fleury, J. Guillaume, and Q. Wang, *Communities in Evolving Networks : Definitions , Detection , and Analysis Techniques*, vol. 2. 2013.
[8] P. Bedi and C. Sharma, “Community detection in social networks,” *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 6, no. 3, pp. 115–135, 2016.
[9] T. Y. Berger-wolf and J. Saia, “A Framework for Analysis of Dynamic Social Networks,” pp. 523–528.
[10] T. Berger-wolf and D. Kempe, “A Framework For Community Identification in Dynamic Social Networks,” pp. 717–726.
[11] S. Y. Bhat and M. Abulaish, “Analysis and mining of online social networks: Emerging trends and challenges,” *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 3, no. 6, pp. 408–444, 2013.
[12] V. D. Blondel, J. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks,” vol. 10008, 2008.
[13] J. Chen and B. Y. Ã, “Systems biology Detecting functional modules in the yeast protein – protein interaction network,” vol. 22, no. 18, pp. 2283–2290, 2006.
[14] A. Clauset, M. E. J. Newman, and C. Moore, “Finding community structure in very large networks,” vol. 66111, pp. 1–6, 2004.
[15] T. N. Dinh, N. P. Nguyen, and M. T. Thai, “An Adaptive Approximation Algorithm for Community Detection in Dynamic Scale-free Networks,” pp. 55–59, 2013.
[16] F. Esposito, S. Ferilli, T. M. a. Basile, and N. Di Mauro, “Social networks and statistical relational learning: a survey,” *Int. J. Soc. Netw. Min.*, vol. 1, no. 2, p. 185, 2012.
[17] M. Fiedler, “Terms of use :,” vol. 23, no. 2, pp. 298–305, 1973.
[18] M. Girvan and M. E. J. Newman, “Community structure in social and biological networks,” vol. 99, no. 12, 2002.
[19] D. Greene, “Tracking the Evolution of Communities in Dynamic Social Networks,” 2010.
[20] R. Guimerà, M. Sales-pardo, and L. A. N. Amaral, “ﬁdlr,” vol. 25101, pp. 1–4, 2004.
[21] J. He and D. Chen, “A fast algorithm for community detection in temporal network,” *Physica A*, no. xxxx, 2015.
[22] A. Hemmati and K. S. K. Chung, “Quality of life: a social network’s perspective,” *Soc. Netw. Anal. Min.*, vol. 6, no. 1, 2016.
[23] C. Huang, D. Wang, S. Zhu, and B. Mann, “Toward local family relationship discovery in location-based social network,” *Soc. Netw. Anal. Min.*, vol. 7, no. 1, 2017.
[24] S. Kamal and M. S. Arefin, “Impact analysis of facebook in family bonding,” *Soc. Netw. Anal. Min.*, vol. 6, no. 1, pp. 1–14, 2016.

ISSN: 0971-1260
Vol-22-Issue-16-August-2019

[25] N. Kapucu, F. Yuldashev, F. Demiroz, and T. Arslan, “Social Network Analysis (SNA) Applications in Evaluating MPA Classes.,” *J. Public Aff. Educ.*, vol. 16, no. 4, pp. 541–563, 2010.

[26] B. Karrer and M. E. J. Newman, “Stochastic blockmodels and community structure in networks,” vol. 16107, pp. 1–10, 2011.

[27] Y. Lin, “FacetNet : A Framework for Analyzing Communities and Their Evolutions in Dynamic Networks,” pp. 685–694, 2008.

[28] T. H. McCormick, R. Ferrell, A. F. Karr, and P. B. Ryan, “Complex Networks as a Unified Framework for Descriptive Analysis and Predictive Modeling in Climate Science,” *Sci. Technol.*, vol. 4, no. 5, pp. 497–511, 2010.

[29] F. Medicine and S. Paulo, “schistosomiasis,” *Lancet Infect. Dis.*, vol. 14, no. 7, pp. 553–554, 2014.

[30] N. Meghanathan, “Randomness Index for complex network analysis,” *Soc. Netw. Anal. Min.*, vol. 7, no. 1, 2017.

[31] S. Moon, J. Lee, M. Kang, M. Choy, and J. Lee, “Parallel community detection on large graphs with MapReduce and GraphChi,” *DATAK*, pp. 1–15, 2015.

[32] N. Natarajan, P. Sen, and V. Chaoji, “Community detection in content-sharing social networks,” *Proc. 2013 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Min. - ASONAM ’13*, pp. 82–89, 2013.

[33] D. F. Nettleton, “A synthetic data generator for online social network graphs,” *Soc. Netw. Anal. Min.*, vol. 6, no. 1, 2016.

[34] S. Network and D. Analytics, *No Title*. .

[35] M. E. J. Newman, “Fast algorithm for detecting community structure in networks,” vol. 66133, no. September 2003, pp. 1–5, 2004.

[36] M. E. J. Newman, “Analysis of weighted networks,” vol. 56131, pp. 1–9, 2004.

[37] M. E. J. Newman, “Modularity and community structure in networks,” vol. 103, no. 23, pp. 8577–8582, 2006.

[38] M. E. J. Newman and M. Girvan, “Finding and evaluating community structure in networks,” pp. 1–15, 2004.

[39] A. G. Nikolaev, R. Razib, and A. Kucheriya, “On efficient use of entropy centrality for social network analysis and community detection,” *Soc. Networks*, vol. 40, pp. 154–162, 2015.

[40] M. Oakleaf, “Writing information literacy assessment plans: A guide to best practice,” *Commun. Inf. Lit.*, vol. 3, no. 2, pp. 80–90, 2009.

[41] G. Palla, “Quantifying social group evolution,” vol. 446, no. April, pp. 3–6, 2007.

[42] S. Papadopoulos, Y. Kompatsiaris, A. Vakali, and P. Spyridonos, “Performance and application considerations,” pp. 515–554, 2012.

[43] S. Parthasarathy, *Social Network Data Analytics*. 2011.

[44] A. Pothenf and H. D. Simon, “WITH EIGENVECTORS OF GRAPHS *,” vol. 11, no. 3, pp. 430–452, 1990.

[45] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, “Defining and identifying communities in networks,” 2004.

[46] M. J. Rattigan, M. Maier, and D. Jensen, “Graph Clustering with Network Structure Indices,” pp. 783–790, 2006.

[47] H. Search, C. Journals, A. Contact, M. Iopscience, and I. P. Address, “Community detection and graph partitioning,” vol. 28003.

[48] R. O. Sinnott and W. Wang, “Estimating micro-populations through social media analytics,” *Soc. Netw. Anal. Min.*, vol. 7, no. 1, 2017.

[49] M. Tsvetovat and A. Kouznetsov, *Social Network Analysis for Startups*. 2011.

[50] C. Wang, W. Tang, B. Sun, J. Fang, and Y. Wang, “Review on Community Detection Algorithms in Social Networks,” *2015 IEEE Int. Conf. Prog. Informatics Comput.*, pp. 551–555, 2015.

[51] P. G. Wasan and N. Jain, “Customizing content for rural mobile phones: a study to understand the user needs of rural India,” *Soc. Netw. Anal. Min.*, vol. 7, no. 1, pp. 1–13, 2017.

[52] A. Zakrzewska and D. a Bader, “A Dynamic Algorithm for Local Community Detection in Graphs,” *2015 IEEE/ACM Int. Conf. Adv. Soc. Networks Anal. Min.*, no. 5, pp. 559–564, 2015.

[53] A. Zakrzewska and D. A. Bader, “Tracking local communities in streaming graphs with a dynamic algorithm,” *Soc. Netw. Anal. Min.*, vol. 6, no. 1, 2016.

[54] N. Zalmout, “Multidimensional Community Detection in Twitter,” pp. 83–88, 2013.

[55] P. Zhang, C. Moore, and M. E. J. Newman, “Community detection in networks with unequal groups,” *Phys. Rev. E*, vol. 93, no. 1, pp. 1–12, 2016.

[56] Z. Zhang, Q. Li, and D. Zeng, “Extracting Evolutionary Communities in Community Question Answering,” vol. 65, no. 6, pp. 1170–1186, 2014.

[57] “Social Network Analysis And Its Applications -A Review From Business Perspective,” vol. 2, no. 9, pp. 3006–3013, 2015.

[58] Pinney JW, Westhead DR. “Betweenness-based decomposition methods for social and biological networks,” *Interdisciplinary Statistics and Bioinformatics. UK: Leeds University Press;* 2006, 87–90.

[59] Guimera R, Danon L, Diaz-Guilera A, Giralt F, Arenas A. Self-similar community structure in a network of human interactions. Phys Rev E 2003, 68:065103.

[60] Duch J, Arenas A. Community detection in complex networks using external optimization. Phys Rev E 2005, 72:027104.

[61] Steinhaeuser K, Chawla NV. Identifying and evaluating community structure in complex networks. Pattern Recogn Lett 2010, 31:413–421.

[62] Bansal S, Bhowmick S, Paymal P. Fast community detection for dynamic complex networks. In: Complex Networks. Berlin Heidelberg: Springer-Verlag; 2011, CCIS 116:196–207.

[63] Tantipathananandh C, Berger-Wolf T, Kempe D. A framework for community identification in dynamic social networks. In: Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2007, 717–726.

[64] Takaffoli M, Sangi F, Fagnan J, Zäıane OR. Community evolution mining in dynamic social networks. Procedia Soc Behav Sci 2011, 22:49–58.

[65] Kim M-S, Han J. A particle-and-density based evolutionary clustering method for dynamic networks. Proc VLDB Endow 2009, 2:622–633.

[66] Chi Y, Song X, Zhou D, Hino K, Tseng BL. On evolutionary spectral clustering. ACM Trans Knowl Discov Data 2009, 3:17.

[67] Folino F, Pizzuti C. An evolutionary multiobjective approach for community discovery in dynamic networks. IEEE Trans Knowl Data Eng 2014, 26:1838–1852.