Volume 8, No. 4, May 2017 (Special Issue)
International Journal of Advanced Research in Computer Science
REVIEW ARTICLE
Available Online at www.ijarcs.info
ISSN No. 0976-5697

Communities in Complex Networks: A glance

Mini Singh Ahuja

Research Scholar
Punjab Technical University
Jalandhar, India
miniahuja06@gmail.com

***Abstract*—**These days we are surrounded with networks like social networks, biological networks, technological networks etc. They exist almost everywhere. Many researchers have shown their interest in these complex networks because of their wide range of applications. These complex networks have many properties like scale free networks, transitivity, presence of community structure etc. Community detection is one of the most active fields in complex networks because it has many practical applications. In this paper, we have presented all the work done till date in the field of community detection.

Keywords— Community; complex networks; community detection algorithms

I. Introduction

The analysis of complex networks is receiving a vast amount of attention from the scientific community these days as complex networks can be used in many domains, such as web, power grids, sensor networks, biological networks and social networks. In these applications, networks can be modeled as graphs where nodes represent objects and edges represent relationships between objects. These nodes can be anything: a person, an organization, a computer or a biological cell. Nodes can have different size or attributes which represent a property of real system objects. These graphs can be directed, undirected or weighted. A complex network has its roots in graph theory. Complex networks have non trival properties so they cannot be explained by uniform random, regular or complete models [4]. This has resulted in definition of set of statistics which have become fundamental properties of complex networks. These properties are now being used by many researchers for studying various phenomena's like spreading of information [5], protocol performance etc. But a major challenge in the study of complex networks is how to collect data for analysis. We cannot directly collect data from these real world complex networks to study them. So researches have to make an assumption that initially data is not fit to find the real properties but as the size of the data grows the properties become more and more stable. The research is going on this side of complex network too [3]. They are trying to find the impact of the measured procedures on the obtained data to study the induced bias [6]

II. Communities

Communities are present in real life too. Cell is a network of genes and proteins which offers a feasible strategy for showing the complexity of living things. Social networks (eg: twitter, facebook) are also one of the best examples of graphs with communities. People form groups within their work place, family and friends. These large real world networks are generally characterized by heterogeneous structures which have some particular properties. The heterogeneous distribution of the links has led to community structure [7, 8, and 9]. A community is a set of entities which are linked to all the other entities in the network. The entities in one community perform same function and share some common properties. A community structure reveals the internal organization of the nodes. Different communities combine to form a complex network. In other words, a community can be described as a collection of vertices within graph which are densely connected among themselves but are loosely connected to the rest of the graph (Newman, and Girvan, 2004). Communities can also be called as clusters, partitions,cohesive subgroups or modules which share common properties. These communities have many features. They can have hierarchal or overlapping structure inside them. Moreover these communities can be dynamic which change with time or can be multirelational (multiple relations). Many real networks such as social networks, biological networks exhibit community structure. Finding communities is crucial because these communities in a network can help to classify the nodes according to their structural position. Moreover boundary and center nodes of the communities can help in obtaining knowledge about the information flow or critical objects. This property can be used in various applications such as to study the spread of disease in social networks [10]. Web clients who have similar interests and are geographically near to each other can be clustered to improve the performance of the service providers on the World Wide Web. Each cluster can be served by a dedicated mirror server. Community Structure property reduces very large graph in to smaller ones. Various research areas related to communities are:

- Detecting communities in a network.
- Finding importance of a node in a network.
- Finding the strength of a community in a network.
- Given a node and finding the community to which it belongs.

Community detection is a NP-hard problem which means in order to find the exact community structure we have to reduce it to a non deterministic polynomial time problem. All the algorithms can detect the exact community structure in a small sized network with an acceptable time. But in case of large networks, it finds them in polynomial time which makes them slow.

III. LITERATURE REVIEW

Till now many algorithms have come up which rely on wide range of principles: hierarchical clustering, optimization methods, graph partitioning, spectral properties etc.

A Pothen (1997) [14] Early methods of community detection relied on graph partitioning. Graph partitioning divides the vertices into different groups of predefined size such that there is minimum number of edges between the groups [14]. Various measures are cut size, ratio cut, conductance, normalized cut etc. Algorithms based on this approach are not fit for community detection as they need to know information about the global structure of the network and also the number and size of the communities in advance.

Girvan and Newman (2002**)** [15] marked the beginning of a new era in the area of community detection. This algorithm is based on edge betweenness, which represents the number of shortest paths between pairs of vertices that run along an edge. Firstly the algorithm calculates betweenness scores for all edges in the network and then the edge with the highest score is removed from the network. For all the remaining edges in the network, betweenness is recalculated followed by the removal of edges with highest betweenness score. The process is repeated until no edges remain in the network. The GN algorithm was the first to recognize that the centrality score must be recalculated after each edge removal. The author conducted test on computer generated and real world graphs whose community structure was already known. But the major problem withthe algorithm of Girvan and Newman is that it is too slow. The worst-case running time of the algorithm is $O(m^2n)$, or $O(n^3)$ on a sparse graph.

Radicchi et al. (2004)[11] proposed an algorithm that is similar to that of Girvan and Newman (2002). It is based on iterative removal of edges. The author used a local measure to recalculate edge clustering coefficient (defined as triangles) every time when an edge is removed. The measure is local, so it can be calculated quickly and takes less execution time than that of Girvan and Newman's algorithm. The edges with low values of clustering coefficients are iteratively removed during every iteration. But this algorithm too has many drawbacks. It relies in the presence of triangles in the network, but real-world networks have few triangles, so algorithm fails to find communities in real world networks. It is good only for social networks and not for other type of networks. The algorithm runs in time $O(m^4/n^2)$ on a graph with m edges and n vertices, or $O(n^{2)}$ on a sparse graph.

Newman and Girvan (2004) [12] proposed modularity a measure for measuring the overall quality of a graph partition. Modularity measures internal connectivity, with reference to a randomized null model. Now modularity is the most widely used objective function for partitioning. The modularity is a numerical index of how good a particular division is and has been very influential in community detection literature. Initially, all the nodes in the network are considered to be in a community of its own. Then, the pairs of communities are merged which will give a greatest increase or smallest decrease to the modularity score of the network. later, the order of merging of communities is used to build the dendrogram. The total running time is O (mn), or $O(n^2)$ on a sparse graph. This algorithm wastes time and memory space as adjacency matrix contains mostly 0 for sparse networks. Speed is the main advantage of this algorithm so it can be used for analyzing large networks.

Clauset et al. (2004) [13] proposed a new hierarchical agglomerative (approximation algorithm) method based on greedy optimization technique. He used more sophisticated data structures such as max-heaps for sparse matrices instead of adjacency matrix to increase the speed of Newman's algorithm. Its running time on a network is O(mdlogn) where m is number of vertices and n number of edges , d is the depth of the dendrogram which describes the community structure. The author tested the algorithm on recommender network of books from the online bookseller Amazon.com which included more than 400000vertices and two million edges. The running time of this algorithm is O(nlogn) on sparse graphs.

Latapy and Pons (2005) [16] proposed a walktrap algorithm based on a distance measure called as random walk which calculates the distance between vertices (and between sets of vertices) in order to capture structural similarities between them. Walktrap algorithm is based on the assumption that while performing random walks the virtual surfer is trapped in the high density regions of the graph (i.e the communities). Using this assumption the surfer does a random walk such that at each time step, a walker is on a node and then moves to another neighbouring node uniformly and randomly. This distance must be large if the two vertices are in different communities. WalkTrap has complexity of $O(mn^2)$

Balakrishnan and Deo (2006) [17] proposed an algorithm for detecting communities in real world random networks using bibliographic metrics. The algorithm makes use of the local properties of the graph. The algorithm computes similarity between two nodes in a graph which is based on bibliometric similarity i.e. similarity based on number of common neighbors these nodes share. More the number of common neighbors more the similarity. The author conducted experiment on computer generated networks and real world networks whose community structure is already known. They concluded that the algorithms based on local properties of the graph produce better communities than those algorithms which use global properties of the graph.

Wakita and Tsurumi (2007) [18] conducted the study on clauset et al (CNM) algorithm. They modified the algorithm of clauset so that it can work efficiently in large scale networks. They noticed that fast algorithm by Clauset et al. is inefficient, because itmerges communities in unbalanced manner**.** They introduced three versions of the CNM algorithm using a metric called consolidation ratio to balance the sizes of the communities being merged. The author replaced balanced binary trees and max heaps data structures which CNM algorithm used by a doubly-linked list which stored the ordered community ID. HE algorithm measures the community size in terms of its degree. HE' algorithm ignores the size of a community and so it behaves like CNM algorithm.HN algorithm measures the size of community in terms of the number of its members The author ran four flavors of the CNM algorithm(original, HE, HE', and HN) on large data set which ranged from 1 million nodes to 5.5 million nodes and found that HN is the fastest.

Raghavan et al. (2007) [19] proposed a label propagation algorithm for community detection in large networks which is based on diffusion. It uses only the network structure to guide its process. The algorithm is mainly appreciated for its near-linear time complexity. In this method, all the nodes are assigned a unique label at the initial step. Then, every node in the network is considered in a random order and it is assigned the label which majority of its neighbors has. This process is repeated until all the nodes in the network get a label which at least half of its neighbor's have. This algorithm doesn't require any external parameter. But this algorithm suffers from unnecessary updates duringevery iteration and does not produce any unique solution. The author has also conducted experiments to verify the accuracy of the algorithm by using real world data. It takes a near-linear time [O(m+n)where m no of edges and n no of nodes] for the algorithm to run to its completion. The proposed label propagation process uses only the network structure to guide its progress and requires no external parameter settings.

Blondel et al. (2008)[20] proposed a heuristic method for detecting community structure based on modularity maximization. The author designed Louvin method which is a greedy optimization method that attempts to optimize the modularity of a partition of the network. The optimization is performed in two steps. In the first step, the method looks for small communities by optimizing modularity locally. In the second step, it aggregates nodes belonging to the same community and builds a new network whose nodes are the communities. These steps are repeated iteratively until a maximum of modularity is attained and a hierarchy of communities is produced. It merges the idea of optimization and modularity with multi level hierarchical scheme. The algorithm is extremely fast. As we known modularity optimization fails to identify communities smaller than a certain scale, which leads to resolution limit problem. But there is no resolution limit problem with this algorithm due to the multi-level nature of the algorithm. This algorithm does not need any community number, as many other optimization algorithms require. Moreover it is easy to implement and it is very fast. The complexity is linear for almost any type of data. The author conducted various tests and claimed that it outperforms all the other methods to which it is compared. The main disadvantage if this algorithm is the problem of storage capacity for very large networks.

M. Rosvall and C. T. Bergstrom (2007)[21] proposed an algorithm based on compression. They stated that the modular structure of a graph can be considered as a compressed description of the graph to approximate the whole information contained in its adjacency matrix. Rosvall and Bergstrom(IND) designed a communication process in which a partition of a graph in communities represents a synthesis of the full structure that a signaler sends to a receiver, who tries to infer the original graph topology from it. The goal was to optimally compress the information needed to describe the process of information diffusion across the graph.

Pizzuti (2008**)** [22] has conducted a study based on Genetic Algorithm for Community Detection in Social Networks. In this study the author has suggested a new algorithm termed GA-Net to invent communities in network by engaging genetic algorithms. The algorithm rearranges an easy but successful function of fitness capable to recognize densely connected node groups with sparse links between groups. The method is effective because the operators of variation are changed to take into consideration only the actual correlations among nodes thus reducing the research space of feasible solutions sensibly. The author introduces the concept of community score, and searches for optimal partitions of the network by maximizing the community score The author has also conducted experiments in this study on real life and synthetic networks which reveals the method's capability to detect the structure of network successfully.

Huang, Sun and Han (2010)[23] conducted a study on a structural clustering algorithm for detecting hierarchical communities in networks. The author has proposed a parameter free hierarchical network algorithm of clustering referred as SHRINK by combining the benefits of modularity optimization and density based clustering methods. Based on the information of structural connectivity the suggested algorithm can efficiently show the embedded structure of hierarchical community with multi-resolution in big scale weighted networks and recognize outliers and hubs. The author has conducted experiments to illustrate this methodology with both synthetic and real world sets of data for detection of community and compare it with several baseline processes. The outcome of the study reveals that SHRINK accomplishes good performance with consistent developments.

Lancichinetti and Fortunato (2010)[24] conducted a study on community detection algorithms. Here author compared several existing community detection algorithms. According to author, uncovering the community structure is most difficult process in the complex networks. In this study, author tested many methods against recently introduced benchmark graphs with heterogeneous distributions of community size and degree. Apart from these, methods are also tested against the benchmark provided by Girvan and Newman on random graphs. The result of the study shows that, algorithms introduced by Rosvall and Bergstrom provides excellent performance, and the algorithms introduced by Blondel et al provides additional benefit of low computational complexity.

De Meo et al (2011)[25] conducted a study based on Generalized Louvain method for community detection in large networks. In this study the author has proposed a novel strategy by inventing the networks community structure. This approach is based on well known network modularity optimization concept. The proposed algorithm used a novel edge centrality evaluation based on k paths. This technique permits to evaluate edge ranking in large networks mainly in closed linear time. Once the ranking of centrality is estimated the algorithm evaluates the proximity of pair wise between network nodes. The algorithm can be applied on unweighted networks as well and, it uses both global and local information. The computational cost of this algorithm is near linear.

Jaewon Yang, Julian McAuley and Jure Leskovec (2014)[26] conducted a study on community detection in the networks with respect to node attributes. Community detection algorithms are most important tools that help to uncover the principles in networks. In the complex networks, community detection algorithms focus only on network structure, while at the same time clustering algorithms focus on the node attributes. It combines the information from the node attributes as well as the network. In this study, author developed CESNA (communities from edge structure and node attributes) which is a scalable and accurate algorithm for detecting communities in the networks with respect to node attributes. Communities from edge structure and node attribute help to detect communities by identifying the relevant node attributes for the community. Apart from these, CESNA also helps to identify the interaction between the node attributes and network structure which in turn detects the community accurately. CESNA takes linear time in terms of number of edges and attributes.

Agarwal (2011)[27] In this study the author has suggested a bi-objective genetic algorithm for detection of community which expands community and modularity rank. The purpose of community detection in graphs is to recognize the modules by using data encoded in topology of network. The outcomes acquired for both real life and benchmark sets of data are contrasted with other algorithms using NMI and other modularity metrics for performance. The outcomes of this study reveal that bi-objective algorithm is capable of detecting structure of community successfully in both synthetic and real life sets of data.

Cai et al (2011)[28] In this study the author has suggested a novel algorithm to invent communities overlapping with link clustering termed as GaoCD (i.e. Genetic algorithm for overlapping community detection). Varied from conventional algorithms based on clustering of node, their algorithm is based on edge clustering. A scalable encoding schema is configured and the several communities are determined automatically in this study. The author has experimented on both real networks and artificial networks to validate the efficiency and effectiveness of the algorithm. This study has revealed that genetic algorithm for overlapping community algorithm accomplished greater partition density and predicts denser communities easily.

Xie, Kelley and Szymanski (2013)[29] proposed an article reviewing the state of art in quality measures, overlapping algorithms of community detection and benchmarks. Modular or community structure is regarded to be an essential real world social networks property as it always reports for the system functionality. In the community definition despite the ambiguity, several technologies have been developed for both effective and efficient detection of community. The author has provided a thorough difference of varied algorithms and in addition to evaluation of community level, he has suggested a structure for estimating the ability of algorithms to predict nodes of overlapping which supports to assess under detection and over detection. After examining the performance of community level detection estimated by normalized mutual data of Omega Index and the performance of node level detection estimated by F-score, the author reached to the final conclusion.

T. Ma *et al* (2016) [31] have proposed an efficient overlapping community detection algorithm named LED (Loop Edges Delete). LED algorithm is based on Structural Clustering, which converts structural similarity between vertices to weights of network. The evaluations of the LED algorithm are conducted both from classical networks and C-DBLP (which is a huge and real-life co-author social network in China). The results show LED is superior to other approaches in terms of accuracy, efficiency, comparing with Fast Modularity and GN algorithm.

Z. Li *et* al (2016) [32] have proposed a new community detection method utilizing multi-swarm fruit fly optimization algorithm (CDMFOA). CDMFOA is found to be more efficient since it needs only a few parameters and has a simple computational process. In order to resolve the premature convergence and to improve the local search ability of CDMFOA, they have adopted the multi-swarm fruit fly strategy and hill-climbing method in community detection algorithm. From the experimental results obtained on synthetic and real-world networks, CDMFOA is found to effectively detect community structure in complex networks.

A. Mahmood *et* al (2016**)** [33]have presented a fundamentally different community detection algorithm based on the fact that each network community spans a different subspace in the geodesic space. Hence, according to this approach, each node can only be efficiently represented as a linear combination of nodes spanning the same subspace. To make the process of community detection more robust, sparse linear coding with l_1 norm constraint has been employed. In this approach, in order to find a community label for each node, sparse spectral clustering algorithm is used. Here the proposed community detection technique is compared with more than ten state of the art methods on two benchmark networks (with known clusters) using normalized mutual information criterion. The proposed algorithm outperformed existing algorithms with a significant margin. The proposed algorithm has also shown excellent performance on three real-world networks.

J. Whang et al (2016) [34] have proposed an efficient overlapping community detection algorithm using a seed expansion approach. The key idea of this algorithm is to find good seeds, and then greedily expand these seeds based on a community metric. Within this seed expansion method, we investigate the problem of how to determine good seed nodes in a graph. A new seeding strategy has been developed for a personalized Page Rank clustering scheme that optimizes the conductance community score. An important step in this approach is the neighborhood inflation step where seeds are modified to represent their entire vertex neighborhood. Experimental results show that seed expansion algorithm outperforms other state-of-the-art overlapping community detection methods in terms of producing cohesive clusters and identifying ground-truth communities. The new seeding strategies is found to be better than existing strategies, and are thus effective in finding good overlapping communities in real-world networks.

F. Zhang *et al* (2016**)** [35]have used a social network analysis to produce the behavior features and transform these features into fuzzy rules which can represent the detection rules. They then optimized the fuzzy rules by genetic algorithms to build the auction fraud detection model. For implementation, real auction data were collected from the online auction site (i.e. http://www.ruten.com.tw, which is the most popular auction site in Taiwan). Finally they have detected the fraudster accounts. They hope this approach can be employed in helping the website administrators to detect the possible collusive fraud groups easier in online auction.

IV. Applications of Community Detection

The study of detecting communities in complex networks has many practical applications. Its use has benefited several application fields such as sociology, communication, computer science, biology, physics etc.

- Detected communities are useful in the study of topology analysis, functional analysis and behavioural analysis of complex networks.
- Communities in biological networks can help in understanding basic mechanisms which control normal cellular processes and diseases pathologies.
- Clusters of customers with similar interests in the network can be used to make recommender systems for viral marketing to enhance the business.
- Adhoc networks don't have any centrally maintained routing tables which can give information about communication between nodes. Nodes in these types of networks can be divided into communities which can help in generating compact routing tables.
- Community detection can help in easy visualization of complex graphs.
- Clusters of large graphs can be used to make large data structures to store huge graph data efficiently and to easily solve navigational queries related to that graph such as path search.
- Community discovery in World Wide Web can help in detecting link farms. A linkfarm is any group of web sites which hyperlink to every other site in the group.
- Detecting communities of tasks in parallel computing can help in knowing the best way of allocating tasks to processors so that the inter process communication can be minimized and better speed can be achieved for parallel programs. Tasks belonging to same community should be allocated to a single processor in the computer cluster.
- Identification of influential nodes of sub communities within large communities can help in predicting churns in telecommunication network.

V. CHARACTERISTICS OF A GOOD COMMUNITY DETECTION ALGORITHM

A good community detection algorithm should satisfy the following properties:

- Able to detect non overlapping/ overlapping communities accurately
- Able to handle the network growth.
- Easy to interpret the detected communities.

But almost all community detection algorithms don't satisfy all the properties. So it required to see which algorithm satisfies all the three above properties.

VI. TESTING OF COMMUNITY DETECTION ALGORITHMS

Community detection algorithms can be tested in following ways:

- Real world networks.
- Artificial networks & Benchmarks

Community detection algorithms are generally designed in order to study real world systems. Using community detection algorithms in the real world networks is always an issue because identification of community structure strongly implies expert human intervention which makes them relatively small and/or rare. Moreover complex networks have many properties such as average degree, shortest path, degree distribution etc which are very difficult to be controlled in real world networks. This makes artificial networks to act as an alternative as artificial networks can be generated in large amounts. They are widely used to compare the performance of different community detection algorithms. We can easily generate artificial networks with desired properties using generative models. But these cannot be substitute to real world data, but can act as complement. The first benchmarks for testing these algorithms were developed by Girvan and Newman called as GN benchmarks. GN benchmarks are very simple to use. Many algorithms give good result with GN benchmarks as all communities identified by them are identical in size. GN benchmarks produce networks with possion distribution but real world networks follow power law distribution. So GN benchmarks are not so fruitful in comparing community detection algorithms. Now days LRF benchmarks proposed by Lancichinetti et al [30] have replaced the GN benchmarks. These benchmarks can generate undirected and unweighted networks with mutually exclusive communities.

VII. FUTURE IN COMMUNITY DETECTIONALGORITHMS

Detecting clusters or communities in real world network is a problem of considerable practical interest. The community detection problem has plenty of challenges as it is highly related to the problem of clustering large heterogeneous datasets. Till date many researchers have proposed number of algorithms, but all the community detection algorithms are different from each other and are not clearly defined[24, 36].

So heterogeneity of different algorithms poses a challenge to community detection. Different networks (biological, social etc) have their own properties. This difference in properties as led to the unsolved question: which algorithm is suitable for which type of network?

Moreover these algorithms don't detect the same communities. So the problem is how to compare the performance of these algorithms. Actually the researchers are interested in following information.

- What type of information is used by the algorithm? A network can have different type of data: link attributes(weights, directions) node attributes, different types of links.
- What type of community produced (partition, overlapped)
- The nature of communities the algorithm identifies.

VIII. CONCLUSION

Community structure plays a key role in the function and formation of several systems and so several papers are published on the topic every year. Community detection is one of the fields of complex network which has gained a lot of attention in today's world. Although, several authors have proposed and described about community detection algorithms on the artificial generated networks, still there are some issues related to the performance and quality of communities detected through these community detection algorithms

REFERENCES

[1] C. Gkantsidis, M. Mihail, and E. Zegura, "Spectral analysis of internet topologies," in *IEEE INFOCOM*, 2003.

[2] P. Boldi and S. Vigna, "The webgraph framework i: compression techniques," in *WWW*, 2004

[3] S. Voulgaris, A.-M. Kermarrec, L. Massouli´e, and M. van Steen,"Exploiting semantic proximity in peer-to-peer content searching," in *IEEE FTDCS*, 2004

[4] R. Albert and A. Barab_asi, Statistical mechanics of complex networks, Reviews of Modern Physics, vol. 74,no. 1, pp. 47{97, 2002.

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

[6] Z. Li and P. Mohapatra, "Impact of topology on overlay routing service," in *IEEE INFOCOM*, 2004.

[7] S. Fortunato. Community detection in graphs. Physics Reports, 486(3-5):75 { 174, 2010.

[8] Richardt J and Bornholdt S, 2006, Statistical mechanics of community detection, Physical Review E74, 016110

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

[10] Goh K-Il, Cusick ME, Valle D, Childs B, Vidal M, et al. 2007. The human disease network. Proc Natl Acad Sci USA

[11] Radicchi, F., Castellano, C., Cecconi, F., Loreto, V. and Parisi, D.(2004) 'Defining and identifying communities in networks'. PNAS USA 101(9):2658-2663.

[12] M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical Review E*, vol. 69, no. 2, 2004.

[13] A. Clauset, M. E. J. Newman, , and C. Moore. Finding community structure in very large networks. Physical Review E, pages 1– 6, 2004

[14] A. Pothen. Graph partitioning algorithms with applications to scientific computing. Technical report, Norfolk, VA, USA, 1997.

[15] M. Girvan, and M. E. J. Newman, "Community structure in social and biological networks," *Proc. of Natl. Acad. Sci. USA, vol. 99, no. 12, pp. 7821-7826, June, 2002*

[16] Pons, P., Latapy, M., "Computing communities in large networks using random walks", *arXiv*, physics/0512106, (2005).

[17] H. Balakrishnan, and N. Deo, "Discovering communities in complex networks," in Proc. of 44th ACMSE. *pp. 280 - 285, 2006.*

[18] K. Wakita and T. Tsurumi. Finding community structure in mega-scale social networks: [extended abstract]. In Proceedings of the 16th International Conference on World Wide Web, WWW '07, pages 1275–1276, New York, NY, USA, 2007. ACM

[19] U. N. Raghavan, R. Albert, and S. Kumara, "Near linear time algorithm to detect community structures in large-scale networks," *Phys. Rev. E, vol. 76, pp. 036106, 2007*

[20] Blondel, V. D., Guillaume, J. L., Lambiotte, R.. Lefebvre, E., "Fast unfolding of communities in large networks", *Journal of Statistical Mechanics-Theory and Experiment, -, Oct (2008)*

[21] M. Rosvall and C. T. Bergstrom. An information-theoretic framework for resolving community structure **in** complex networks. Proceeding in National Academy of Sciences, 104(18):7327–7331, 2007.

[22] Pizzuti C (2008), GA-Net: A Genetic Algorithm for Community Detection in Social Networks, Parallel Problem Solving from Nature – PPSN X, Lecture Notes in Computer Science, Volume 5199, pp 1081-1090.

[23]. Huang, J., Sun, H., Han, J., Deng, H., Sun, Y and Liu, Y. (2010), SHRINK: a structural clustering algorithm for detecting hierarchical communities in networks, *In Proceedings of the 19th ACM international conference on Information and knowledge management, pp. 219-228*

[24]Andrea Lancichinetti, and Santo Fortunato (2010): Community detection algorithms: a comparative analysis

[25] De Meo, P., Ferrara, E., Fiumara, G and Provetti, A. (2011), Generalized louvain method for community detection in large networks. In*Intelligent Systems Design and Applications (ISDA), 2011 11th International Conference*, pp. 88-93.

[26]. Yang, Jaewon, Julian McAuley, and Jure Leskovec (2014): "Community Detection in Networks with Node Attributes." arXiv preprint arXiv:1401.7267 (2014).

[27] Agrawal, R. (2011), Bi-objective community detection (bocd) in networks using genetic algorithm. In *Contemporary Computing* (pp. 5-15). Springer Berlin Heidelberg.

[28] Cai, Y., Shi, C., Dong, Y., Ke, Q and Wu, B (2011), A novel genetic algorithm for overlapping community detection. In *Advanced Data Mining and Applications*, Springer Berlin Heidelberg, pp. 97-108.

[29] Xie, J., Kelley, S., & Szymanski, B. K. (2013), Overlapping community detection in networks: The state-of-the-art and comparative study. *ACM Computing Surveys (CSUR)*, *45*(4), pp 43.

[30] Lancichinetti A, Fortunato S and Radicchi F (2008) Benchmark graphs for testing community detection algorithms Phys Rev E 78

[31]T. Ma, Y. Wang, M. Tang, J. Cao, Y. Tian, A. Al-Dhelaan and M. Al-Rodhaan, "LED: A fast overlapping communities detection algorithm based on structural clustering", Neurocomputing, 2016

[32]Z. Li and J. Liu, "A multi-agent genetic algorithm for community detection in complex networks", Physica A: Statistical Mechanics and its Applications, vol. 449, pp. 336-347, 2016.46110

[33]A. Mahmood and M. Small, "Subspace Based Network Community Detection Using Sparse Linear Coding", IEEE Trans. Knowl. Data Eng., vol. 28, no. 3, pp. 801-812, 2016.

[34]J. Whang, D. Gleich and I. Dhillon, "Overlapping Community Detection Using Neighborhood-Inflated Seed Expansion", IEEE Trans.Knowl. Data Eng., vol. 28, no. 5, pp. 1272-1284, 2016.

[35]F. Zhang, J. Li, F. Li, M. Xu, R. Xu and X. He, "Community Detection Based on Links and Node Features in Social Networks", MultiMedia Modeling, pp. 418-429, 2015.

[36] Hubert L and Arabie P 1985 Comparing partitions Journal of Classification 2 193-218