Twelfth International Multi-Conference on Information Processing-2016 (IMCIP-2016)

An Analysis of Overlapping Community Detection Algorithms in Social Networks

J. Chitra Devi* and E. Poovammal

SRM University, Kattankulathur, Chennai 603 203, India

Abstract

In the field of research, Social Network Analysis is prevalent domain which pulls the attention of many data mining experts. Social network analysis is the specific field of sociology and anthropology. It shares a number of characteristics common to real network. Some real networks like Facebook, Twitter exhibit the concept of community structure within the network. Social network is represented as a network graph. Detecting the communities involves finding the densely connected nodes. Overlapping communities are possible if a node is a member of more than one community. This paper discusses various modularity based approaches on detecting the overlapping communities in the social networks. This work aims in providing the characteristics and limitations of modularity based overlapping community detection algorithms.

Keywords: Community Detection; Modularity; Overlapping Community; Social Network; Social Network Mining.

1. Introduction

As real world scenario is dynamic and evolving, Social network is used intensively in wide range of applications and represented as a network graph with nodes and edges. Nodes represent the individual users/actors/items/resources whereas edge represents the link/flow of interaction/relationship among the users. The arrangement of nodes and edges in a graph is coined as topology. Some common techniques involved in social network includes Recommendation Systems, Link Analysis, Expert Identification, Influence Propagation, Trust & Distrust Relationship Prediction, Opinion Mining, Mood Analysis and Community Detection[1–3]. Recommendation systems analyze the available data and suggest something the actor might be interested in. As a result a new link is introduced in the network. Recommending friends or recommending resources happens through collaborative filtering and Content-based filtering. Each node represents entity and the analyzes of links among the nodes gives the behavior pattern of different activities. Link analysis helps in finding this behavior pattern of a social network. Strategy of finding an expert in a required domain by analyzing the social network is coined as Expert Identification. With the help of previous log information, Influence Propagation between the nodes is identified. Two models are used in Influence Propagation namely Independent Cascade Model and Linear Threshold Model. Based on attributes of a node, a relationship or a link in a social network is identified as Trustable or not. This study is termed as Prediction of Trust or Distrust relationship.

*Corresponding author. Tel.: +91 -9940638001.
 E-mail address:* chitradevi21@gmail.com

Opinion mining involves building a system to collect and categorize opinions about a product or a person. It uses ideas of machine learning, text mining and natural language processing. The levels of opinion mining are document level, sentence level and phrase level. To track the mood of public about a particular product, mood analysis is performed. Set of actors interacting with each other frequently forms a Community. Detecting the community involves clustering of similar actors in social network graph. Different types of communities are node-centric community, link-centric community, network-centric community and hierarchy-centric community. Various community detection approaches have been proposed namely graph-based approach[4], link-based approach[5], agent and dynamic based approach[25,26], fuzzy-based approach[20–22] and modularity scoring based approach[6–19,23,24]. This paper analyzes the modularity based algorithms in detail.

2. Related Works

Social network is a network graph holding nodes with edges connected to it. Community detection aims to identify the community in the graph. Community is a module containing the set of nodes with major activities/interaction/ similarity among them. Overlapping communities are possible if any of the nodes has participated in the formation of different modules. Each node's strength or membership value in different modules varies accordingly. Various metrics are utilized to measure the strength of the community. Most popular of it is modularity measure. This modularity is the strength of partition of network as a community. Section 3 explains the modularity based community detection approach in detail.

3. Modularity Scoring based Clustering Approach

A community is said to be strong if it has more internal interaction within the community. To qualify the strength of a community, parameters like quality functions, density, conductance and modularity are used. This work reviews on the adaptation of the modularity measure in a undirected or directed, weighted or unweighted graph. Modularity defines the strength of partition of network into communities. It quantifies the association between a node and a community based on some parameters. In this paper, different version of belonging factor in modularity measure is considered. Belonging factor is a measure explaining the degree to which a node u belongs to a community C.

3.1 Modularity based disjoint community detection

Newman's modularity[6] measure considers the intra connection and inter connection between the nodes. As like other properties of communities say, node degree, betweenness, centrality or clustering coefficients, modularity measures the strength of formation of communities in a social network. This modularity is defined in equation (1)

$$Q = \frac{1}{2m} \sum_{uv} \left[A_{uv} - \frac{K_u K_v}{2m} \right] \delta(C_u, C_v) \tag{1}$$

where A_{uv} represents the adjacency matrix with row and column as node u and v respectively A_{uv} value is 1, if there is an edge between u and v and 0, otherwise. K_u, K_v is the degree of node u and v respectively, m is total number of edges in the graph $\delta(C_u, C_v)$ is represented as equation (2)

$$\delta(C_u, C_v) = \begin{cases} 1, & (u, v) \in C \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

At any particular time t, it can produce only two communities in a network. Hierarchical partitioning with two sub communities and further partitioned into two smaller communities. Thus only binary partition is possible. It limits its application only to an undirected and unweighted graph in a social network.

In order to overcome the limitations of Newman's modularity, Leicht *et al.*[7] defines the modularity measure for directed graph. The extended definition is the null-model directed graph with P_{uv}. This P_{uv} is the probability of having the link starting from node u to node v. Generally in directed graph, link from node u to node v is different from the

link from node v to node u. The same concept is been applied. The P_{uv} is not equal to P_{vu}. Thus the modularity for directed graph is given in equation (3)

$$Q = \frac{1}{m} \sum_{uv} \left[A_{uv} - \frac{K_u^{\text{out}} K_v^{\text{in}}}{m} \right] \delta(C_u, C_v) \tag{3}$$

Note that K_u^{out} is the out-degree of node u and K_v^{in} is the in-degree of node v. The in-degree is the number of incoming edges to a node and out-degree is the number of outgoing edges from a node. This work can identify the community well. But Newman[6] and Leicht[7] modularity measure does not take into account the overlaps among the communities. This problem is resolved by Nicosia *et al.*[8] by introducing the concept of overlapping communities based on edge.

3.2 Modularity based overlapping community detection

In case of overlapping communities, each node may belong to one or many communities. But the strength of participation of node in all overlapped communities may not be equal. This reason leads to design of array values namely "belonging factors". This array is defined as, $C = \{a_{i,1}, a_{i,2}, \ldots, a_{i,|C|}\}$ for any node i belonging to communities $1, 2, \ldots, |C|$. In a directed graph $G(V, E)$, coefficient $a_{i,c}$ expresses how strongly node i belongs to community c. In general for any node i,

$$\sum_{c=1}^{|C|} a_{i,c} = 1.$$

Thus it introduces the belonging coefficient to the equation (3) given by Leicht[7], instead of $\delta(C_u, C_v)$. This belonging coefficient β are used to weight the probability of having a link starting at node u and link pointing to node v. Accordingly, modified modularity is given in equation (4)

$$Q = \frac{1}{m} \sum_c \sum_{uv} \left[\beta_{l(u,v),c} A_{uv} - \frac{\beta_{l(u,v),c}^{\text{out}} K_u^{\text{out}} \beta_{l(u,v),c}^{\text{in}} K_v^{\text{in}})}{m} \right] \tag{4}$$

This equation (4) has more computation overhead with belonging coefficient of link from node u to node v, out-degree belonging coefficient from node u and in-degree belonging coefficient from node v. Thus time computation to compute this belonging coefficient is more compared to modularity computation.

3.2.1 Modularity with fuzzy concept

Nepusz *et al.*[9] considered the belonging coefficient as the probability of the event that node u is in community c. The probability that node u and node v belongs to same community C is the dot product of their membership vector denoted as $S_u v = \sum_{c \in C} a_{u,c} a_{v,c}$ S_{uv} is the similarity measure between nodes i and j. Thus replacing the $\delta(C_u, C_v)$ in equation (3) with similarity measure S_{uv}, equation (5) is obtained

$$Q = \frac{1}{m} \sum_{c \in C} \sum_{uv} \left[A_{uv} - \frac{K_u K_v}{m} \right] a_{u,c} a_{v,c} \tag{5}$$

Thus the concept of overlapping community detection is defined. If the communities are disjoint, there exists only one community c for every node u with $a_{u,c}$ as 1. Then the equation (5) will get reduced to equation (2). The simple dot product of membership vector doesn't holds good. Since the membership vector ranges from 0 to 1, the dot product of decimal values may lead to a small value decreasing the modularity measure unnecessarily.

Chen *et al.*[10] gives the generalized belonging function calculating the modularity for undirected graph. It is given by equation (6)

$$Q = \frac{1}{2m} \sum_{c \in C} \sum_{uv} \left[A_{uv} - \frac{K_u K_v}{2m} \right] f(a_{u,c}, a_{v,c}) \tag{6}$$

It is same as equation (5) but only difference is how the value $f(a_{u,c}, a_{v,c})$ is computed. The belonging coefficient function $f(a_{u,c}, a_{v,c})$ can be the product or average of $a_{u,c}, a_{v,c}$. If it is a product the equation is same as equation (5). Or if it is a average, it becomes equation (7)

$$Q = \frac{1}{2m} \sum_{c \in C} \sum_{uv} \left[A_{uv} - \frac{K_u K_v}{2m} \right] [(a_{u,c} + a_{v,c})/2] \tag{7}$$

Compared to similarity measure based on dot product as in equation (5) introduced by Nepusz *et al.*[9], equation (7) with average belonging coefficient works better. But it limits the equation to be applied only to undirected graph.

Shen *et al.*[11] redefines the modularity definition in equation (5). The belonging coefficient of node u is calculated as the reciprocal of the number of communities to which the node u belongs to $a_{u,c} = \frac{1}{O_u}$ where O_u is the number of communities containing node u. Thus equation (5) becomes,

$$Q = \frac{1}{m} \sum_{c \in C} \sum_{uv} \left[A_{uv} - \frac{K_u K_v}{m} \right] \frac{1}{O_u O_v} \tag{8}$$

This computation is very simple and straight forward. The strength of node u in the number of communities O_u is considered as equal. Practically this does not sound good. The participation of node u in each of the communities is based on its importance. And it is not always same.

EAGLE algorithm[12] works with two stages. First stage generates a dendogram by combining the similar group of nodes together and second stage breaks the dendogram into communities. In first stage to compute the similarity between group of nodes or between communities, equation (9) is introduced.

$$S = \frac{1}{2m} \sum_{u \in C_1, v \in C_2, u \neq v} \left[A_{uv} - \frac{K_u K_v}{2m} \right] \tag{9}$$

While combining communities together based on similarity measure, equation (5) is used to find the strength of the newly formed communities. In second stage, overlapping community structure is considered as the cover of network instead of partition of network. Therefore to quantify the overlapping communities, the measure of cover of network is used as in equation (10). For a graph $G(V, E)$ and a partition P of the network G,

$$Q = \frac{1}{2m} \sum_{c \in P} \sum_{uv} \delta_{uc} \delta_{vc} \left[A_{uv} - \frac{k_u K_v}{2m} \right] \tag{10}$$

In equation (8), δ_{uc} denotes whether the node u belongs to community C. The value of δ_{uc} is 1 when the node u belongs to community C and 0, otherwise. This algorithm has two drawbacks. It is applicable only to undirected and unweighted graph. The parameter $\delta_{uc} \delta_{vc}$ in equation (10) may not reflect the belonging coefficient of node to more than one community. To be specific, overlapping among the communities is not support in this algorithm. In order to cover the overlapping communities as well, EAGLE algorithm is extended to introduce the belonging coefficient α_{uc}. This belonging coefficient α_{uc} reflects the value how much node u belongs to community c. With the belonging coefficient α_{uc}, the strength of community C in network is measured as,

$$Q = \frac{1}{2m} \sum_{c \in P} \sum_{uv} \alpha_{uc} \alpha_{vc} \left[A_{uv} - \frac{k_u K_v}{2m} \right] \tag{11}$$

Belonging coefficient always satisfies a normalization property as $0 \leq \alpha_{uc} \leq 1$, for every node u and for every community c and $\sum_{c \in C} \alpha_{uc} = 1$. Though the extension in EAGLE algorithm considers the belongingness of a node but still it could not solve either weighted graph or directed graph.

3.2.2 Modularity using graph structure clique

Chen *et al.*[13] proposed another extension to the modularity function. The belonging coefficient of a node is defined based on the graph structure namely clique. Clique is a strongly connected component of a graph. In this case, the belonging coefficient of node u to community c is defined as,

$$\alpha_{u,c} = \frac{1}{\alpha_u} \sum_{k \in C} \frac{M_{uk}^c}{M_{uk}} A_{uk} \tag{12}$$

M_{uk} denotes the number of maximal cliques in the network containing edge (u, k) and M_{uk}^c is the number of maximal cliques in community c that contains edge (u, k) and $a_u = \sum_{c \in C} \sum_{k \in C} \frac{M_{uk}^c}{M_{uk}} A_{uk}$ is a normalization term. Now the equation (5) on substituting equation (11) becomes,

$$Q = \frac{1}{m} \sum_{c \in C} \sum_{uv} \left[A_{uv} - \frac{K_u K_v}{m} \right] \left[\frac{1}{\alpha_u} \sum_{k \in C} \frac{M_{uk}^c}{M_{uk}} A_{uk} \right] \left[\frac{1}{\alpha_v} \sum_{k \in C} \frac{M_{vk}^c}{M_{vk}} A_{vk} \right] \tag{13}$$

The limitation of Shen *et al.* equation is the requirement of densely connected components namely clique in the social network graph. This condition may not be possible in all real time scenarios. The real time graph is highly sparse in nature. It also gives a new way to find the belonging coefficient of node i to community c in a graph based structure as in equation (12) and equation (13)

$$a_{u,c} = \frac{\Sigma_{k \in C} A_{uk}}{\Sigma_{c' \in C_i} \Sigma_{k \in c'} A_{uk}} \tag{14}$$

Where c_i is set of communities to which node i belongs consequently, extended modularity definition for overlapping community structure is given in equation (15)

$$Q = \frac{1}{m} \sum_{c \in C} \sum_{uv} \left[A_{uv} - \frac{K_u K_v}{m} \right] \left[\frac{\Sigma_{k \in C} A_{uk}}{\Sigma_{c' \in C_i} \Sigma_{k \in c'} A_{uk}} \right] \left[\frac{\Sigma_{k \in c} A_{vk}}{\Sigma_{c' \in C_i} \Sigma_{k \in c'}, A_{vk}} \right] \tag{15}$$

Densely connected component is required for this equation (15) to be applied.

3.2.3 Modularity using random walk

Feng *et al.*[14] measures the belonging coefficient of members by Random walk. This belonging coefficient describes the closeness of the nodes connected to a community. Nodes with the largest belonging coefficients correspond to core members. Transition probability of taking a random walk from node u to node v is given by $P_{uv} = \frac{A_{uv}}{d_u}$ where A_{uv} is lif there is a link between node u and node v, 0 otherwise, d_u is degree of node u. Given source node s, the probability walking from node u to node s with t steps is

$$q_s^t(u) = \sum_{v=1}^{N} q_s^{T-1}(v) P_{uv} \tag{16}$$

The belonging coefficient is the probability walking from node u to node s within T steps is in equation (17)

$$BC_s^t(u) = \sum_{r=1}^{t} q_s^t(u) \tag{17}$$

This gives how close the node i is connected to community with sources.

3.2.4 Modularity using similarity measure

Lee *et al.*[15] computes the symmetric measure given by Newman's formula for modularity computation. This symmetric measure is a distance based on a community C getting embedded in another community C'. The formula to compute the similarity measure is in equation (18)

$$\delta(C_u, C_v) = 1 - \frac{|C_u \cap C_v|}{\min(|C_u|, |C_v|)} \tag{18}$$

Extended modularity function Q on applying the equation (17) in equation (2) the equation (19) is derived

$$Q = \frac{1}{m} \sum_{c \in C} \sum_{uv} \left[A_{uv} - \frac{K_u K_v}{m} \right] \left[1 - \frac{|C_u \cap C_v|}{\min(|C_u|, |C_v|)} \right] \tag{19}$$

3.2.5 Modularity with edge types

A new formula is defined to measure the strength of community in the novel algorithm namely OCDLCE[17]. It is defined as the ratio of its internal edges and external edges. The proposed algorithm relies on both inside edge and outside edge rather than inner edges alone. The formula is defined in the equation (20)

$$M = \frac{M_{\text{in}}}{M_{\text{out}}} = \frac{\frac{1}{2} \Sigma_{uv} A_{uv} \theta(u, v)}{\Sigma_{uv} A_{uv} \lambda(u, v)} \tag{20}$$

where M_{in} represents the internal edges whose end points are both in community C and M_{out} represents the external edges which means only one of its end points belongs to community C. The value of $\theta(u, v)$ is 1 if both nodes u, v belongs to community C, 0 otherwise and $\lambda(u, v)$ is 1 if just only one of node u or v belong to community C, else 0. It traverses each edge of network and finds local communities by checking the intersection neighbour set of its endpoints. Its time complexity is O(m) where m is the number of edges.

3.2.6 Modularity based on conductance

Li *et al.*[18] uses the graph based approach with one of its components as conductance. The conductance defines the measure of how well-knit the graph is. Conductance is used to measure the belonging coefficient. It is defined in the equation (21)

$$\varphi(c) = \frac{|\theta(c)|}{\min(\text{Vol}(C), \text{Vol}(C'))} \tag{21}$$

where $\theta(c)$ is the cut size and Vol (c) is is the sum of vertex degree in the set C. This formula holds good for a undirected and unweighted graph. But it limits the size of the social graph. Finding the cut size is not always possible as social network data is large and dynamic in the real time environment. A slight modification is made to the equation of conductance to support the weighted graph is given by Lu *et al.*[19]. Lu *et al.* computed the conductance based on the weights of the cut edges of community C instead of cut size of Clique. This conductance equation (22) is applicable to the weighted graph. This is the first work in community detection which considers the weighted graph in the analysis

$$\varphi(c) = \frac{cut(C, G/C)}{w_c} \tag{22}$$

where $cut(C, G/C)$ is weights of the cut edges of C and w_c is weights of all edges in community C including the cut edges. When edge weight is more in intra cluster module, i.e w_c is too high, the conductance is low. Hence the community identified is better. Once better community is formed, belonging degree of node u in community $CB(u, C)$ is defined based on the equation (23) with K_u is the total degree of node u

$$B(u, C) = \frac{\Sigma_{v \in C} w_{uv}}{K_u} \tag{23}$$

To measure the strength of the community formed, modularity measure is computed as in equation (24)

$$Q = \frac{1}{2m} \sum_{uv} \delta(\rho_u, \rho_v) \left[A_{uv} - \frac{k_u K_v}{2m} \right] \qquad (24)$$

where $\rho_u = arg_{c \in C_u} \max B(u, C)$ and $\delta(\rho_u, \rho_v)$ is 1 if node u and v are in the same community and 0 otherwise. According to this algorithm, A neighbouring node is eligible to be added into a community only if newly formed community has a lower conductance. The time complexity to compute is $O(m^2)$ with m as number of vertices. It is applicable for the weighted graph and undirected graph. Its application in directed graph still not yet resolved. There should be a algorithm which can solve any type of graph either weighted or unweighted and directed or undirected. When the social network gets evolved in a large and dynamic real time environment, the algorithm should work without any limitations.

4. Analysis of Typical Network Graph

Consider the graph shown in the Fig. 1. It has 9 nodes connected with one another.

Based on Newman's modularity[6], the two disjoint communities are identified. The strength of partition of nodes to form the community is measured as modularity. The modularity value of the community structure formed in the Fig. 2 is 0.7567. Usually the modularity value ranges from −1 to 1. The value −1 represents the community strength is minimum and +1 represents the maximum strength of a community. The identified community are represented using large oval shapes of varying thickness.
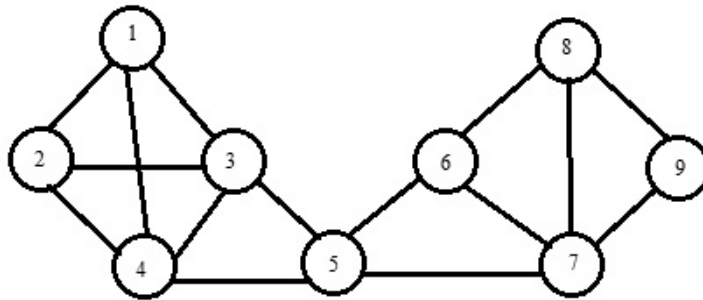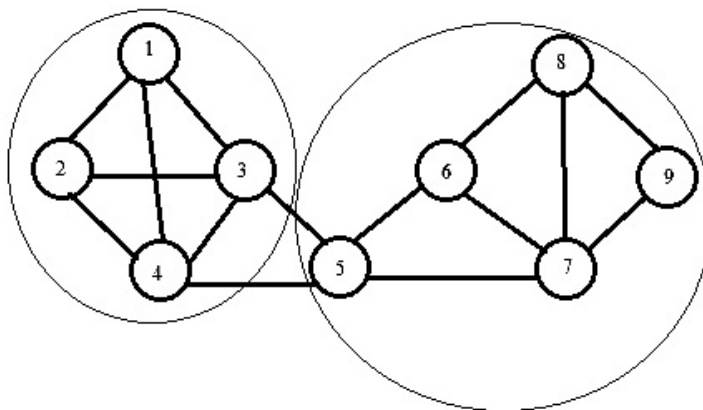

Fig. 1.   Example Graph with 9 Nodes.

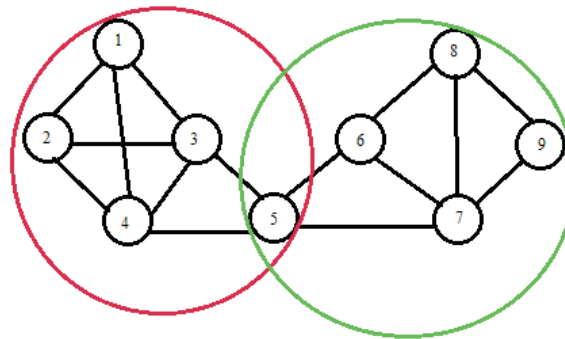Fig. 2.   Disjoint Community Structure of the Graph.

Fig. 3.   Overlapping Community Structure of the Graph.

Table 1. Summary of Modularity Score.

Author/Algorithm Name	Modularity score
Newman *et al.*	0.7567
Nepusz *et al.*	0.7457
Chen *et al.*[10]	0.7851
Shen *et al.*[11]	0.7400
EAGLE algorithm	0.7425
Chen *et al.*[13]	0.6425
Feng Jiao *et al.*[14]	0.3317
Lee *et al.*	−0.2738
OCDLCE algorithm	0.5000

The formed communities are $C1 = \{1, 2, 3, 4\}$ and $C2 = \{5, 6, 7, 8, 9\}$. Apart from Newman's approach, all other approaches[7–19] result in formation of overlapping community. It is shown in the Fig. 3. For the given sample graph in the Fig. 1, the overlapping community is detected. It is represented with different color each represent the community detected.

Figure 3 represents that the two communities $C1$ and $C2$ identified are $\{1, 2, 3, 4, 5\}$ and $\{5, 6, 7, 8, 9\}$. The common node between $C1$ and $C2$, i.e. $C1 \cap C2 = \{5\}$. The node 5 participates in both the communities $C1$ and $C2$ with varying participation strength in each community. This participation strength is measured as *belonging degree*. Its value differs according to each researcher's view and may depend on the domain/application chosen. The Modularity scores are calculated for the typical network graph considered in Fig. 1. The Modularity scores computed are listed in the Table 1.

This value is computed based on the adjacency matrix designed for the graph and the degree of each node in the graph. For example, In equation (1) Newman *et al.*, The value of adjacency matrix A_{uv} is 1 if there is an edge between the nodes and 0, otherwise K_u denotes the degree of node u. Similarly K_v represents the degree of node v. $\delta(u, v)$ is 1 if node u and node v belongs to community C and 0, otherwise. With seed elements selected, nodes are added to community at each iteration thus producing the modularity score as 0.7567. From the Table 1 it can be inferred that the modularity measure is maximum for Chen *et al.*[10] equation. It requires the belonging coefficient for each of the overlapping node to its participating community. This requirement of belonging coefficient may be an overhead compared to the work of Shen *et al.*[11]. Yet there is a trade off between modularity score and overhead. Lee *et al.* modularity equation yields a low community strength/modularity score for the same scenario. But his work can be preferred in a sparse network environment.

5. Modularity Approaches in Community Detection

Various algorithms of community detection are analyzed based on modularity. The summarized view of the algorithms which have been discussed in Section 3 is depicted in Table 2. It gives the review of modularity adaptation based community detection algorithms.

Table 2. Overview of Modularity Adaptation based Community Detection.

Author Name/Algorithm	Techniques Used	Observations
Girvan-Newman[6]	It includes the adjacency matrix and the probability of an edge between the vertices proportional to their degree.	The algorithm runs slowly, taking time complexity as $O(m^2 n)$ on a network of n vertices and m edges. It is impractical to apply for a network with more than a few thousand nodes. Not applicable to directed graph.
Leicht et al.[7]	To apply for the directed graph, the starting and end point of line is noted. The probability of link from node u to node v is different for a link from node v to node u.	Communities are identified well. But not applicable for weighted or overlapped communities.
Nicosia et al.[8]	Based on belonging factor of edges, strength of participation of node in a community changes.	This belonging factor is defined for both indegree as well as outdegree edges. Has little computational overhead because of this. Not applicable for weighted graph.
Nepusz et al.[9]	The probability that node u and node v belongs to same community C is the dot product of their belonging factor.	Simple dot product does not holds good. Since factor value ranges from 0 to 1, dot product results in minimal value. Not applicable for weighted graph.
Chen et al.[10]	Generalized the belonging function. This function later can be modified as dot product or average or anything else.	Not applicable for directed or unweighted graph.
Shen et al.[11]	Redefines the belonging function as the reciprocal of number of communities to which node u belongs to.	Used in most of the research paper. The strength of node u in the number of communities are considered equal but practical scenario has variant in intensity of node participation.
EAGLE algorithm[12]	It uses two stages of approaches namely agglomerative approach and divisive approach.	Only applied to undirected and unweighted graph.
Chen et al.[13]	Graph partitioning technique namely Clique is used to find the belonging coefficient of a node.	Sparse graph if taken for computation cannot be solved.
Feng jiao et al.[14]	Belonging coefficient is based on random walk. Probability of taking a random walk from node i to node j is considered for computation.	This algorithm gives the measure of closeness of a node in a community.
Lee et al.[15]	The distance based on a community C getting embedded in the other community C' is computed as symmetric measure. It is used in modularity function to find the strength of community.	Not applicable for weighted network.
OCDLCE[17]	Modularity is defined as the ratio of its internal edges and external edges in a community.	Its time complexity is $O(m)$ with m-number of edges.
Li et al.[18]	Similar to graph based approaches with conductance as measure. This conductance show how well-knit the graph is.	It is not applicable to weighted graph too. And it limits the size of the social graph.
Lu et al.[19]	Extended Li et al. work with inclusion of weight of a social graph in the formula of conductance.	It is the modularity approach which works for a weighted graph.

6. Conclusions

In this work, several state-of-the-art modularity based community detection algorithms with disjoint and overlapped communities are analyzed. Quantitative analysis is performed with the modularity score to infer the best available method. Chen et al.[10] algorithm proves to be the best with high modularity score. But computational overhead is high since it requires belonging coefficient of each node. Newman et al. modularity score is equivalent to Chen et al.[10] with a limitation of detecting only disjoint communities. Shen et al.[11] computation of modularity is simple compared to Chen et al.[10]. Few algorithms cannot be applied to directed graphs namely Newman et al.[6], Chen et al.[10], Shen et al.[11], EAGLE algorithm[12]. Lu et al.[19] have worked on graph with weights in it. The analysis can be applied in a dynamic

social network with machine learning techniques involved in it. Edge weights have a major role in determining the strength of node in a community. Few researches are made considering this edge weight as a key role in community detection in the social network field.

References

[1] Ting Yuan, Jian Cheng, Xi Zhang, Qingshan Liu and Hanqing Lu, How Friends Affect User Behaviors? an Exploration of Social Relation Analysis for Recommendation, *Knowledge Based Systems*, vol. 88, pp. 70–84, (2015).
[2] Malhar Anjaria, Ram Mohana and Reddy Guddeti, A Novel Sentiment Analysis of Social Networks Using Supervised Learning, *Social Network Analysis Mining*, vol. 4, no. 181, March (2014).
[3] Jing Zhang, Zhanpeng Fang, Wei Chen and Jie Tang, Diffusion of Following Links in Microblogging Network, *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 8, pp. 2093–2106, August (2015).
[4] I. Palla, Derenyi I. Farkas and T. Vicsek, Uncovering the Overlapping Community Structure of Complex Networks in Nature and Society, *Nature*, vol. 435, pp. 814–818, (2005).
[5] Y. Y. Ahn, J. P. Bagrow and S. Lehmann, Link Communities Reveal Multiscale Complexity in Networks, *Nature 466*, pp. 761–764, (2010).
[6] M. E. J. Newman and Girvan, M., Finding and Evaluating Community Structure in Networks, *Physical Review E*, vol. 69, pp. 026–113, February (2004).
[7] E. A. Leicht and M. E. J. Newman, Community Structure in Directed Networks, *Physical Review Letter*, vol. 100, no. 11, March (2008).
[8] V. Nicosia, G. Margioni, V., Carchiolo and M. Malgeri, Extending the Definition of Modularity to Directed Graphs with Overlapping Communities, *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2009, no. 3, pp. 03–204, (2009).
[9] T. Nepusz, A. Petroczi, L. Negyessy and F. Bazso, Fuzzy Communities and the Concept of Bridgeness in Complex Networks, *Physical Review E*, vol. 77, pp. 016–107, January (2008).
[10] D. Chen, M. Shang, Z. Lu and Y. Fu, Detecting Overlapping Communities of Weighted Networks via a Local Algorithm, *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 19, pp. 4177–4187, (2010).
[11] H. W. Shen, X. Q. Cheng and J. F. Guo, Quantifying and Identifying the Overlapping Community Structure in Networks, *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2009, no. 7, pp. 07–042, (2009).
[12] H. Shen, X. Cheng, K. Cai and M. B. Hu, Detect Overlapping Hierarchical Community Structure in Networks, *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 8, pp. 1706–1712, (2009).
[13] Mingming Chen, Konstantin Kuzmin and BoleSlaw K. Szymanski, IEEE/ACM 4th Social Network Analysis and Applications (SNAA) workshop at ASONAM, Beijing, China, (2014).
[14] Feng jiao Chen and Kan Li, Detecting Hierarchical Structure of Community Members in Social Networks, *Knowledge Based Systems*, vol. 87, pp. 3–15, October (2015).
[15] Conrad Lee, Fergal Reid, Aaron McDaid and Neil Hurley, Detecting Highly Overlapping Community Structure by Greedy Clique Expansion, 4th SNA-KDD Workshop, (2010).
[16] H. W. Shen, Community Structure of Complex Networks, Springer-Verlag Berlin Thesis, (2013).
[17] Yan Xing, Fanrong Meng, Yong Zhou, Ranran Zhou and Zhixiao Wang, Overlapping Community Detection by Local Community Expansion, *Journal of Information Science and Engineering*, vol. 31, pp. 1213–1232, (2015).
[18] Yixuan Li, Kun He, David Bindel and John E. Hopcroft, Uncovering the Small Community Structure in Large Networks: A Local Spectral Approach, *International World Wide Web Conference Committe*, (2015).
[19] Zong Qing Lu, Xiao Sun, Yong gang Wen, Guo hong Cao and Thomas La Porta, Algorithms and Applications for Community Detection in Weighted Networks, *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 11, November (2015).
[20] S. Zhang R. S. Wang and X. S. Zhang, Identification of Overlapping Community Structure in Complex Networks Using Fuzzy c-means Clustering, *Physica A: Statistical Mechanics and its Applications*, vol. 374, pp. 483–490, (2007).
[21] Y. Nepusz, A. Petrocz, L. Negyessy and F. Bazso, Fuzzy Communities and the Concept of Bridgeness in Complex Networks, *Physica Review E 77*, pp. 016–107, (2008).
[22] I. Psorakis, S. Roberts and Ebden Mark, Overlapping Community Detection Using Bayesian Non-Negative Matrix Factorization, *Physica Review E*, vol. 83, pp. 066–114, (2011).
[23] U. N. Raghavan, R. Albert and S. Kumara, Near Linear Time Algorithm to Detect Community Structures in Large-Scale Networks, *Physica Review E*, vol. 76, issue 3, pp. 036–106, (2007).
[24] A. Lancichinetti, S. Fortunato and J. Kertesz, Detecting the Overlapping and Hierarchical Community Structure of Complex Networks, *New Journal of Physics*, vol. 11, 033015, p. 18, (2009).
[25] Qiguo Dai, MaoZu Guo, Yang Liu, Xiaoyan Liu and Ling Chen, MLPA: Detecting Overlapping Communities by Multi-Label Proapagation Approach, *IEEE Congress on Evolutionary Computation*, June 2013.
[26] Ting Wang, Xu Qian and Xiaomeng Wang, A Hybrid Label Propagation Algorithm to Find Communities in Large-Scale Networks, *IEEE Conference on Awareness Science and Technology*, (2015).