IJE TRANSACTIONS A: Basics Vol. 30, No. 4, (April 2017) 486-492

International Journal of Engineering

Journal Homepage: www.ije.ir

Mining Overlapping Communities in Real-world Networks Based on Extended Modularity Gain

S. Rao Chintalapudi*, M. H. M. Krishna Prasad

Department of CSE, University College of Engineering Kakinada(A), JNTUK, Kakinada, India

PAPER INFO

Paper history:
Received 28 October 2016
Received in revised form 09 February 2017
Accepted 09 February 2017

Keywords:
Social Network Analysis
Community Detection
Overlapping Communities
Graph Mining
Modularity
Extended Modularity Gain

ABSTRACT

Detecting communities plays a vital role in studying group level patterns of a social network and it can be helpful in developing several recommendation systems such as movie recommendation, book recommendation, friend recommendation and so on. Most of the community detection algorithms can detect disjoint communities only, but in the real time scenario, a node can be a member of more than one community at the same time, that leads to overlapping communities. A novel approach is proposed to detect such overlapping communities by extending the definition of newman's modularity for overlapping communities. The proposed algorithm is tested on LFR benchmark networks with overlapping communities and on real-world networks. The performance of the algorithm is evaluated using popular metrics such as ONMI, Omega Index, F-score and Overlap modularity and the results are compared with its competent algorithms. It is observed that extended modularity gain can detect highly modular structures in complex networks with overlapping communities.

***doi**: 10.5829/idosi.ije.2017.30.04a.05*

NOMENCLATURE			
Q	Modularity	Q_{E_a}	Extended modularity for community a
A	Adjacency Matrix	ΔQ_E	Gain in Extended Modularity
M	Number of edges	Q_{ov}	Overlap Modularity
k_i	Degree of node i	C	Community
$\delta(c_i, c_j)$	Kronecker function	**Greek Symbols**	
k_i^{out}	Out-degree of node i	β_1	Actual belonging coefficient
k_j^{in}	In-degree of node j	β_2	Expected belonging coefficient
Q_d	Modularity for directed graphs	$\alpha_{i,c}$	Belonging coefficient of node i with respect to community c
Q_E	Extended modularity	μ	Mixing parameter
		Ω	Omega Index

1. INTRODUCTION

In recent times, data in many complex systems can be represented as networks such as food webs, transportation networks, co-authorship networks, social networks, communication networks, citation networks, world wide web, biological networks and so on. One of the most important properties of such networks is community structure. Communities can be formed naturally with the interaction between nodes in the network. Community is a group of densely connected nodes with sparse connections to rest of the network. It can be very helpful to explore and understand group level patterns of a given network. Communities in small real-world networks can be identified manually, whereas in large scale networks it is an extremely difficult problem. Many researchers from both computer

*Corresponding Author's Email: *srao.chintalapudi@gmail.com* (S. Rao Chintalapudi)

Please cite this article as: S. Rao Chintalapudi, M. H. M. Krishna Prasad, Mining Overlapping Communities in Real-world Networks Based on Extended Modularity Gain, International Journal of Engineering (IJE), TRANSACTIONS A: Basics Vol. 30, No. 4, (April 2017) 486-492

science and physics domains are contributed well to solve the problem of identifying communities. Community detection algorithms broadly categorized into two, based on type of community that the algorithm can detect such as disjoint or overlapping communities. In disjoint communities, a node can be a member of only one community and it can be illustrated as in Figure 1. The Figure 1 represents a sample network with 9 nodes, where sets of nodes {1,2,3,4,5} and {6,7,8,9} are its two disjoint communities. Several disjoint community detection algorithms [1-4] were proposed in the recent years such as Modularity Maximization, LPA, Infomap. However, in the real-world scenario, a node may be a member of more than one community that leads overlapping communities. The Figure 2 represents a sample network with 8 nodes, where sets of nodes {1,2,3,4,5} and {5,6,7,8} are its two overlapping communities. Here, node 5 is the overlapping node that can be shared among two communities. Detecting overlapping communities is harder than detecting disjoint communities because it has exponential number of possible solutions.

The majority of the community detection algorithms can detect only disjoint communities. In the recent years, some of community detection algorithms were proposed to detect overlapping communities and the popular among them are CPM [5], COPRA [6], SLPA [7]. The primary focus of this paper is to design an overlapping community detection algorithm by extending the definition of newman's modularity [8].

The remainder of this paper is organized as follows. In section 2, background and the related work is reviewed. In section 3, extended modularity gain is derived for overlapping communities. Extended modularity gain based overlapping community detection is presented in section 4. In section 5, experimental results are presented and comparison of the proposed algorithm with two baseline algorithms. Finally, this paper is concluded in section 6.

2. RELATED WORK

Newman et al. [8] proposed a quality function for communities called modularity.

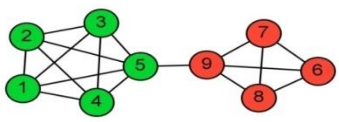

Figure 1. A sample network with two disjoint communities, each color represents one community

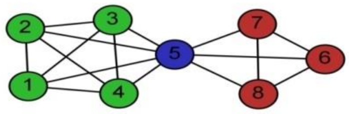

Figure 2. A sample network with two overlapping communities, Node 5 is shared among two communities

It can be defined as sum of differences between actual number of edges and expected number of edges in the network. This is one of the popular metric for evaluating quality of a community detection algorithm. It can be useful for both evaluating and generating communities in the network. Modularity of a network can be computed using Equation (1).

Q= (Number of edges within communities) - (Expected number of such edges).

$$Q = \frac{1}{2m}\sum_{i,j\in V}\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta(c_i,c_j) \quad (1)$$

where, m is the total number of edges in the network, A_{ij} are values of adjacency matrix that are defined as $A_{ij}=1$ if nodes i and j have an edge, $A_{ij}=0$ otherwise, k_i and k_j are degrees of nodes i and j respectively, $\delta(c_i,c_j)$ represents *Kronecker function* where its value is 1 if both the nodes i and j belongs to same community, 0 otherwise. This definition can be applicable to only undirected graphs and it is extended for directed graphs by Leicht et al. [9] as in Equation (2).

$$Q_d = \sum_{i,j\in V}\left[\frac{A_{ij}}{m} - \frac{k_i^{out}k_j^{in}}{m^2}\right]\delta(c_i,c_j) \quad (2)$$

where, k_i^{out} is out-degree of node i and k_j^{in} is in-degree of node j. The problem with these two definitions is that they can be applicable for disjoint communities only and not used for overlapping communities in the network. The extension of modularity gain definition for overlapping communities is discussed in section 3.

Many community detection algorithms have been proposed over recent years. Majority of them can detect disjoint communities effectively; only few can detect overlapping communities [3]. The popular among overlapping community detection algorithms are Clique Percolation Method (CPM), Community Overlap Propagation Algorithm (COPRA) and Speaker-Listener Label Propagation Algorithm (SLPA). Palla et al. [5] proposed clique percolation method that considers cliques in a graph and performs community detection by finding adjacent cliques. COPRA is an extension of label Propagation Algorithm, in which each node updates its belonging coefficients based on average of belonging coefficients of its neighbors. SLPA is a linear time algorithm for detecting overlapping communities based on speaker listener interaction rules. Meng et al. [10] also proposed an overlapping community detection algorithm based on modularity.

3. EXTENDED MODULARITY GAIN

$\lfloor \alpha_{i,1}, \alpha_{i,2}, \alpha_{i,3}, ..\alpha_{i,n} \rfloor$ If a node is a member of more than one community at the same time, then the node has different membership strength to different communities. Therefore, one needs to find out node strength in terms of different communities. Consider [alpha i,1, alpha i,2.....] are *belonging coefficients* for node *i* with respect to all communities. Here, *n* represents number of communities and the sum of all *belonging coefficients* for a node is 1. Similarly, the *belonging coefficient* of an edge can be computed as a function of belonging coefficients of nodes that forms edge as in Equation (3).

$$\beta_1 = F(\alpha_{i,c}, \alpha_{j,c}) \tag{3}$$

where, $F(\alpha_{i,c}, \alpha_{j,c})$ can be chosen as a two-dimensional function as in Equation (4)

$$F(\alpha_{i,c}, \alpha_{j,c}) = \frac{1}{\left(1+e^{-f(\alpha_i, c)}\right)\left(1+e^{-f(\alpha_j, c)}\right)} \tag{4}$$

where, $f(\alpha_{i,c})$ is a linear scaling function and it can be computed using Equation (5).

$$f(x) = 2px - p, p \in R \tag{5}$$

The expected belonging coefficient can be computed as the average of all possible belonging coefficients to the same community using Equation (6).

$$\beta_2 = \frac{\sum_{i,j \in V} F(\alpha_{i,c}, \alpha_{j,c})}{|V|} \tag{6}$$

Finally, according to Newman's modularity the extended modularity for overlapping communities in the case of undirected networks can be derived as in Equation (7).

$$Q_E = \frac{1}{2m} \sum_{c \in C} \sum_{i,j \in V} \left[\beta_1 A_{ij} - \frac{\beta_2 k_i k_j}{2m} \right] \tag{7}$$

where, β_1 and β_2 are actual and expected belonging coefficients of an edge *e* with nodes *i* and *j* belonging to community *c*.

If the combination of two communities increases Q_E value, then it means that the combined community structure is superior to the communities before combination. Hence, one can adopt the definition of Q_E for two communities *a* and *b* as follows.

$$Q_{E_a} = \frac{1}{2m} \sum_{a \in C} \sum_{i,j \in a} \left[\beta_1 A_{ij} - \frac{\beta_2 k_i k_j}{2m} \right] \tag{8}$$

$$Q_{E_b} = \frac{1}{2m} \sum_{b \in C} \sum_{i,j \in b} \left[\beta_1 A_{ij} - \frac{\beta_2 k_i k_j}{2m} \right] \tag{9}$$

If you add Q_{E_a}, Q_{E_b} values of communities a and b, then:

$$Q_{E_{a+b}} = Q_{E_a} + Q_{E_b} \tag{10}$$

A new community *c* is obtained after combining two communities *a* and *b* and the Q_E value of *c* is defined as follows.

$$Q_{E_c} = \frac{1}{2m} \sum_{c \in C} \sum_{i,j \in c} \left[\beta_1 A_{ij} - \frac{\beta_2 k_i k_j}{2m} \right] \tag{11}$$

The gain in extended modularity (ΔQ_E) can be computed as follows.

$$\Delta Q_E = Q_{E_c} - Q_{E_{a+b}} \tag{12}$$

4. EMOCD ALGORITHM

Extended Modularity gain based Overlapping Community Detection (EMOCD) algorithm assigns initially each node to different community. It also finds extended modularity for each community. If the combination of two communities increases, the extended modularity gain value, then it combines both the communities. The major steps involved in detecting overlapping communities from complex networks are as follows.

Algorithm: EMOCD
Input: Network in edge list format
Output: Overlapping community set
Step 1: Each node in the network is assigned to different community.
Step 2: For each node *i* compute the gain in extended modularity using (12) with respect to removal of node *i* and added to its neighbor's community.
Step 3: The node *i* is assigned to the community for which extended modularity gain is positive and highest and it can be done using greedy approach with maximizing extended modularity as an objective function.
Step 4: Construct a new network based on communities identified in step 3.
Step 5: Repeat Step 2 to Step 4 iteratively until extended modularity gain of every pair of adjacent communities is less than or equal to zero.

In the above algorithm, step 1 assigns each node to a different community and the initial modularity can be calculated. In step 2, for each node *i* compute extended modularity gain using (12) with the removal of node *i* and move the node from its community to its neighbor's community. It finds the community for which extended modularity gain is highest using greedy approach. Finally, place the node *i* in the community for which extended modularity gain is positive and the highest. Repeat the same process for all the nodes in the network. After this step, preliminary communities can be found. Based on these communities, a new network can be constructed in step 4, in such a way that an edge can be placed between two nodes, if there is a link between nodes in the two preliminary communities.

According to the Equation (12), if the value of ΔQ_E is higher, it means that the combine community makes more contribution to the value of Q_E. Based on this, compute ΔQ_E for every pair of adjacent communities and then combine the communities with largest ΔQ_E value. In Step 5, Iterate the above process until ΔQ_E is less than or equals to zero for every pair of adjacent communities. Therefore, the final community structure will have optimal Q_E value.

5. EXPERIMENTS AND RESULT ANALYSIS

In this section the results are reviewed by applying the algorithm to synthetic networks and real-world network datasets. Three classic algorithms CPM, COPRA and SLPA are selected to compare with proposed algorithm aiming at proving the validity and the feasibility of the algorithm. All the experiments are done on a Intel Xeon ® E5-2620 CPU at 2.10 GHz with main memory of 32 GB and the algorithm is implemented in java.

The performance of the community detection algorithm can be evaluated in two ways. One is to test overlapping community detection algorithm on synthetic networks with ground truth information. The results can be quantified using ONMI, Omega index and F-Score and remaining details of evaluation can be found in section 5.1. The second way of evaluating the algorithm is to test with real-world networks. However, the problem with this is that the overlapping community ground truth is not available with the real-world networks. So, one has to evaluate the algorithm without ground truth information. It can be possible with the metric overlap modularity and the evaluation can be explained in section 5.2.

5. 1. Tests on Synthetic Networks

To study the behavior of the algorithm, six synthetic networks are generated using LFR benchmark generator [11] by varying mixing parameter (μ) from 0.1 to 0.6. The lower μ value generates highly modular communities in the network. The other parameters used in generating LFR benchmark networks are number of nodes (N =5000), average degree ($\bar{k} = 10$), maximum degree ($K_{max}=30$), exponents of the power law distribution (t_1,t_2), minimum community size ($C_{min}=20$), maximum community size ($C_{max}=30$), number of overlapping nodes ($O_n=20$) and number of memberships of the overlapping nodes ($O_m= 2$ *to* 6). These networks are resembles real-world networks in terms of degree distribution and clustering coefficient. The ground truth is available for these networks, so one can apply the metrics such as ONMI, Omega Index and F-Score.

Normalized Mutual Information is a concept of information theory and it can be used to measure the quality of community detection algorithm. This can be applicable for disjoint community detection algorithms only. Hence, the extended NMI called Overlapping Normalized Mutual Information (ONMI) which is proposed by McDaid et al. [12] is used to evaluate the proposed algorithm. It is defined based on two normalization inequalities, such as max(H(X), H(Y)) which is denoted by $ONMI_{MAX}$ and 0.5*(H(X)+H(Y)) which is denoted by $ONMI_{SUM}$. $ONMI_{MAX}$ and $ONMI_{SUM}$ can be calculated using Equation (13) and Equation (14), respectively. The range of $ONMI_{MAX}$ and $ONMI_{SUM}$ is also in between 0 and 1, where 1 corresponds to perfect matching and 0 corresponds to no matching.

$$ONMI_{MAX} = \frac{I(X:Y)}{\max(H(X),H(Y))} \tag{13}$$

$$ONMI_{SUM} = \frac{I(X:Y)}{0.5*(H(X)+H(Y))} \tag{14}$$

where, the mutual information *I(X:Y)* can be calculated as follows.

$$I(X:Y) = \frac{1}{2}[H(X) - H(X|Y) + H(Y) - H(Y|X)] \tag{15}$$

The Figure 3 represents comparison of the quality of communities detected using EMOCD with its competent algorithms CPM, COPRA and SLPA. Here, ONMI is computed for LFR Benchmark networks by varying mixing parameter (μ) from 0.1 to 0.6. The significance of the mixing parameter in LFR Benchmark networks is that the lower values of μ creates highly modular communities in the network. The community property of a network can be lost with the larger values of μ. ONMI has two variants such as $ONMI_{MAX}$ and $ONMI_{SUM}$ but this paper adopts only $ONMI_{SUM}$ as in Figure 3. Here, the number of communities of a node can be shared (O_m) is fixed at 2. The behavior of the algorithm is studied in terms of ONMI and it is observed that the proposed approach has higher values of ONMI over its competent algorithms. Hence, the proposed algorithm EMOCD will detect good quality communities.

Another popular metric to evaluate overlapping community detection algorithm is Omega Index [13]. It is an overlapping version of the Adjusted Rand Index and it can be defined as in Equation (16).

$$\Omega(C_1,C_2)=\frac{\Omega_u(C_1,C_2)-\Omega_e(C_1,C_2)}{1-\Omega_e(C_1,C_2)} \quad (16)$$

where, unadjusted omega index $\Omega_u(C_1,C_2)$ is the fraction of pairs that occur together in the same number of communities in both partitions and it can be computed as in Equation (17).

$$\Omega_u(C_1,C_2)=\frac{1}{M}\sum_j \left|t_j(C_1)\cap t_j(C_2)\right| \quad (17)$$

where, $t_j(C)$ is the set of pairs that appear together in j communities of partition C. $\Omega_e(C_1,C_2)$ is the expected value in the null model and it can be computed as in Equation (18).

$$\Omega_e(C_1,C_2)=\frac{1}{M^2}\sum_j \left|t_j(C_1)\right|\left|t_j(C_2)\right| \quad (18)$$

The above definition of omega index is reduced to Adjusted Rand Index, if the network doesn't have overlapping communities. The comparison of EMOCD with its competent algorithms in terms of overlapping version of omega index is depicted in Figure 4 and is studied by varying mixing parameter (μ) from 0.1 to 0.6. It is observed that the value of omega index is degraded with the increment of μ and the proposed approach gives higher values of omega index than the other three algorithms.

F-score [4] is also adopted to evaluate accuracy of overlapping nodes detection and it can be defined as the harmonic mean of *precision* and *recall*. *Precision* for overlapping nodes is defined as a fraction of the number of correctly detected overlapping nodes and the total number of detected overlapping nodes and *recall* for overlapping nodes defined as a fraction of the number of correctly detected overlapping nodes and the real number of overlapping nodes. The range for F-score is from 0 to 1, and the value of F-score is higher when the accuracy of detecting overlapping nodes is higher. In Figure 5, the behavior of the algorithm is studied in terms of F-score by varying number of communities a node participates (O_m) from 2 to 6 and it is observed that the value of F-score is reduced with the increment of O_m. That means, with the increment of overlapping the quality of communities will be degraded. Here, the proposed algorithm gives good values of F-score with respect to other three algorithms. For this experiment, the mixing parameter value (μ) is fixed at 0.1.

5. 2. Tests on Real-World Networks To test the proposed algorithm, seven real-world networks are considered namely Karate, Dolphins, Polbooks, football, Jazz, Netscience and Polblogs. These network datasets are acquired from Newman's web page[2].

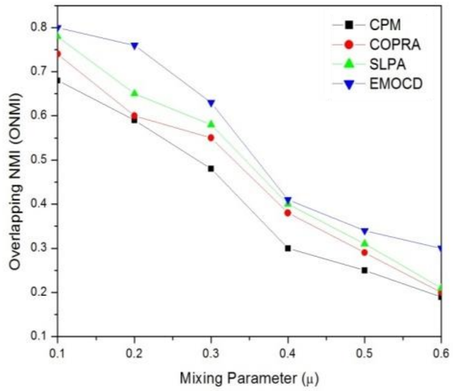

Figure 3. Comparison of EMOCD with other algorithms in terms of ONMI on LFR Benchmark networks with O_m=2

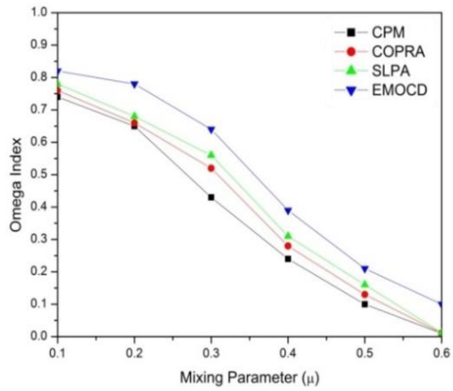

Figure 4. Comparison of EMOCD with other algorithms in terms of Omega Index on LFR Benchmark with O_m=2

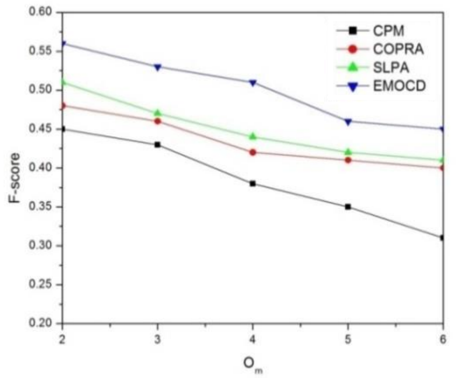

Figure 5. Comparison of EMOCD with other algorithms in terms of F-Score on LFR Benchmark networks with $\mu = 0.1$

[2] http://www-personal.umich.edu/~mejn/netdata/.

The detailed summary of these networks are listed in Table 1. In these networks, there are few networks whose community structure is already known and the communities are usually not overlapped, so it is difficult to execute accuracy analysis of the algorithm using ONMI, Omega Index and F-score. Therefore, authors adopted Overlap modularity as the evaluation criteria to assess density of intra community edges detected by the algorithm. It can be used as both evaluating and detecting overlapping communities in complex networks and can be defined as in Equation (7). The scaling function can be assumed as $f(x) = 60x - 30$, suggested by Nicosia et al. [14]. The higher values of overlap modularity indicates good community detection algorithm. Usually, the modularity above 0.3 indicates good modular structures in the network. Table 2 is the comparison of EMOCD with three algorithms CPM, SLPA and COPRA in terms of overlap modularity. It is observed that EMOCD gives good values of overlap modularity in comparison with other three algorithms except for NetScience dataset. Hence, the proposed approach finds good modular communities in the complex networks.

TABLE 1. Summary of real-world networks

Data set	Nodes	Edges	Description
Karate	34	78	Zachary's karate club
Dolphins	62	159	Dolphin social network
Polbooks	105	441	Books about US Politics
Football	115	613	American college football
Jazz	198	2742	Jazz musicians network
Netscience	379	914	Co-authorship Network
Polblogs	1490	16718	Political blogs

TABLE 2. Comparison of EMOCD with CPM, SLPA and COPRA in terms of Overlap Modularity (Q_{ov})

Data set	CPM	COPRA	SLPA	EMOCD
Karate	0.52	0.44	0.65	**0.68**
Dolphins	0.66	0.70	0.76	**0.79**
Polbooks	0.79	0.82	0.83	**0.85**
Football	0.64	0.69	0.70	**0.73**
Jazz	0.55	0.71	0.70	**0.72**
Netscience	0.61	0.82	**0.85**	0.84
Polblogs	0.51	0.47	0.49	**0.53**

6. CONCLUSION AND FUTURE SCOPE

The most popular quality function used in detection of communities is modularity. The definition of modularity is extended in order to detect overlapping communities in real-world networks. The proposed algorithm Extended Modularity gain based Overlapping Community Detection (EMOCD) uses the extended definition of modularity for overlapping communities in a greedy manner. Hence, at the end of the algorithm it gives optimal value of Extended Modularity. The communities that are having higher values of extended modularity imply that they are good modular structures in the network. The proposed algorithm is tested on LFR benchmark networks with overlapping communities and seven real-world networks. The accuracy of the algorithm is also evaluated using popular metrics such as ONMI, Omega Index, F-Score and Overlap modularity and the results are compared with its competent algorithms such as CPM, COPRA, and SLPA. It is observed that EMOCD can detect highly modular structures in complex networks. It is also suitable for vide variety of real-world network datasets. The scalability of the algorithm can be further improved using High Performance Computing paradigms in future.

7. REFERENCES

1. Harenberg, S., Bello, G., Gjeltema, L., Ranshous, S., Harlalka, J., Seay, R., Padmanabhan, K. and Samatova, N., "Community detection in large-scale networks: A survey and empirical evaluation", ***Wiley Interdisciplinary Reviews: Computational Statistics***, Vol. 6, No. 6, (2014), 426-439.
2. Chintalapudi, S.R. and Prasad, M.K., "A survey on community detection algorithms in large scale real world networks", in Computing for Sustainable Global Development (INDIACom), 2nd International Conference on, IEEE., (2015), 1323-1327.
3. Chakraborty, T., "Leveraging disjoint communities for detecting overlapping community structure", ***Journal of Statistical Mechanics: Theory and Experiment***, Vol. 2015, No. 5, (2015), P05017.
4. Xie, J., Kelley, S. and Szymanski, B.K., "Overlapping community detection in networks: The state-of-the-art and comparative study", ***Acm Computing Surveys (csur)***, Vol. 45, No. 4, (2013), 43-50.
5. Palla, G., Derényi, I., Farkas, I. and Vicsek, T., "Uncovering the overlapping community structure of complex networks in nature and society", ***Nature***, Vol. 435, No. 7043, (2005), 814-818.
6. Gregory, S., "Finding overlapping communities in networks by label propagation", ***New Journal of Physics***, Vol. 12, No. 10, (2010), 103018.
7. Xie, J. and Szymanski, B.K., "Towards linear time overlapping community detection in social networks", in Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer., (2012), 25-36.
8. Newman, M.E. and Girvan, M., "Finding and evaluating community structure in networks", ***Physical review E***, Vol. 69, No. 2, (2004), 026113.

9. Leicht, E.A. and Newman, M.E., "Community structure in directed networks", ***Physical Review Letters***, Vol. 100, No. 11, (2008), 118703.

10. Meng, K., Liu, G., Hu, Q. and Li, J., "An modularity-based overlapping community structure detecting algorithm", in Advances in Social Networks Analysis and Mining (ASONAM), IEEE/ACM International Conference on, (2014), 113-117.

11. Lancichinetti, A., Fortunato, S. and Radicchi, F., "Benchmark graphs for testing community detection algorithms", ***Physical review E***, Vol. 78, No. 4, (2008), 046110.

12. McDaid, A.F., Greene, D. and Hurley, N., "Normalized mutual information to evaluate overlapping community finding algorithms", ***arXiv preprint arXiv:1110.2515***, Vol., No., (2011).

13. Gregory, S., "Fuzzy overlapping communities in networks", ***Journal of Statistical Mechanics: Theory and Experiment***, Vol. 2011, No. 02, (2011), P02017.

14. Nicosia, V., Mangioni, G., Carchiolo, V. and Malgeri, M., "Extending the definition of modularity to directed graphs with overlapping communities", ***Journal of Statistical Mechanics: Theory and Experiment***, Vol. 2009, No. 03, (2009), P03024.

Mining Overlapping Communities in Real-world Networks Based on Extended Modularity Gain

S. Rao Chintalapudi, M. H. M. Krishna Prasad

Department of CSE, University College of Engineering Kakinada(A), JNTUK, Kakinada, India

PAPER INFO

Paper history:
Received 28 October 2016
Received in revised form 09 February 2017
Accepted 09 February 2017

Keywords:
Social Network Analysis
Community Detection
Overlapping Communities
Graph Mining
Modularity
Extended Modularity Gain

چکیده

تشخیص جوامع نقش حیاتی در مطالعه الگوهای سطح گروه یک شبکه اجتماعی بازی می کند و آن می تواند در توسعه چندین سیستم های پیشنهاد مثل پیشنهاد فیلم، پیشنهاد کتاب، پیشنهاد دوستان و غیره مفید باشد. بسیاری از الگوریتم های تشخیص جامعه می تواند جوامع مجزا را فقط شناسایی کند، اما در سناریوی زمان واقعی، یک گره می تواند یک عضو از بیش از یک جامعه در همان زمان باشد که به تداخل جوامع راهنمایی می کند. روش جدیدی برای تشخیص چنین جوامعی که با هم تداخل دارند با استفاده از گسترش تعریف پیمانه نیومن برای جوامعی که با هم تداخل دارند، پیشنهاد می شود. الگوریتم پیشنهاد شده در شبکه های معیار LFR با جوامعی که با هم تداخل دارند و در شبکه های دنیای واقعی تست شده است. عملکرد الگوریتم با استفاده از معیارهای رایج مانند ONMI ، شاخص امگا، نمره F و پیمانهای همپوشانی تخمین زده می شود و نتایج با الگوریتم درست خود مقایسه می شود. مشاهده شده است که افزایش پیمانه گسترش یافته می تواند ساختارهای بسیار مدولار در شبکه های پیچیده ای که با جوامع تداخل دارند را تشخیص دهد.

***doi**: 10.5829/idosi.ije.2017.30.04a.05*