Community detection in complex networks using flow simulation

Zheng Sun,Fuyou Miao,Yang Xu
Dept. of Computer Sci.andTech.
Univ. of Sci. and Tech. of China, Hefei,
230027, China
0086-551-63606694,
sunzhen;mfy;xuyang{@ustc.edu.cn}

Qiyu Shao
Department of Computer Science
DayanandaSagar Institutions,
Bangalore University,Bangalore,India
0091-7760297194
shaofromchina@gmail.com

WeihuaWang,Xingfu Wang, Chenxi Shao
Dept. of Computer Sci.and Tech.
Univ. of Sci. and Tech. of China, Hefei,
230027,China
0086-551-63606694
Whwang;Wangxfu;cxshao{@ustc.edu.cn}

ABSTRACT

Community detection and analysis is an important part of studying the organization of complex systems in real world, and it's extensively applied on many fields. Recently, many of existing algorithms are not effective or the results are unstable. In this paper, a new method of community testing is proposed by us based on the conception of flow field. In our approach, each node is represented as a field source and has a tendency to forward data to the connected nodes with highest field strength, after some iterations the nodes with same data information form a community. It is evaluated by us for the approach on some synthetic and real-world networks whose community structures are known. It is demonstrated that the approach performs wellin effectiveness and robustness.

CCS Concepts

• Networks →Network simulations •Computing methodologies →Model development and analysis

Keywords

Complex network, Data mining, Community detection, Flow simulation

1. INTRODUCTION

A lot of real-world systems can be modeled as complex networks[1][2], where nodes represent entities and edges stand for the relationships among entities. There are many statistical properties in complex network, such as small-world effect[3], scale-free property[4][5] and hierarchical structure[6]. Except for the statistical properties above, complex networks always present another important character, community structure where the density of edges in a community is more denser than that among different communities. Usually, a community is a basic functional module or organization unit of the network. Therefore, finding out the community structures is very helpful for better understanding the function and characteristics of the complex network.

In recent years, many efforts have been made on community detection problem. A lot of methods have been proposed based on different conceptions and hypothesis . For example, the GN algorithm[7] proposed by Newman and Girvan makes use of the conception of edge betweenness. GN algorithm iteratively removes links with the largest edge betweenness until all links in the graph have been removed. One of the drawbacks is that one should deal with the whole hierarchical tree, because GN doesn't offer a criterion to judge which partition is the best. In order to determine the best partition, Newman further proposed Modularity Q which originates the study of mixing patterns[8] in network. A larger modularity indicates better partition. Modularity[9] indicates to what extent the edges in the found community structure are more than that by random chance. Later, Clause et.al proposed a fast modularity greedy algorithm[10] by using max-heap data structure. Since then, many methods aiming at optimizing modularity were proposed[11][12][13][14].

Different from classical GN and Modularity methods, there are other novel algorithms for detecting community structure. Label Propagation Algorithm[15] (LPA) proposed by Raghvan has near linear time complexity. This method firstly assigns each node with a unique label; in each iteration, every node adopts the label which most of its neighbors own as its label until each node in the network doesn't change its label. Subeljet. al. noticed that the prior updated labels in LPA have an advantage of propagating their labels in network. This phenomenon depresses the robustness of LPA method. So they put forward the balanced propagation algorithm[16] (BPA) which assigned smaller weights to the nodes that are preferentially updated. Compared with basic LPA method, BPA enjoys higher robustness and accuracy. The WalkTrap[17] proposed by Pons et.al. introduced the random walk metric as the similarities between nodes and utilized the modularity as the stopping criterion. Motivated by the success of WalkTrap, Steinhaeuser et.al. employed real random walk and proposed the Random Walk method[18]. They presumed that a random walker with limited steps seldom crosses the boundary of two different communities and would like to stay in the community where it belongs. Random Walk approach results in a similarity matrix and with the help of agglomerative technique, it gets the final partition.

Except for the approaches mentioned above, some local community detection algorithms were also proposed, assuming that it is hard to obtain the global topology because of the dynamic feature of the network. These methods find local

ICNCC 2017, December 8–10, 2017, Kunming, China
© 2017 Association for Computing Machinery.
ACM ISBN 978-1-4503-5366-3/17/12…$15.00
DOI: https://doi.org/10.1145/3171592.3171600

community structure by optimizing some local metrics, such as local community metric L[19], local modularity R[20], modularity M[21] and so on. When we focus on the belongings of some specific nodes rather than all nodes in network, the local methods are more effective than those which need the whole topology.

Most algorithms above define the conception of community from the perspective of topology structure. In this paper, we notice that a community in complex network is also a functional module and organization unit. The information exchange is more frequent in the same community or during an outbreak of infectious disease, members in a community are more prone to get infected by those who are in the same community. The exchanged information or virus acts as some kinds of flows in complex network. If we appropriately simulate the flow process in the network and regard the nodes impacted by same flow as the members in same community, we can naturally get a novel definition of community from flow perspective and give a new community detection algorithm based on flow simulation.

This paper is organized as follows: the conception of data flow field and the accompanying community detection algorithm are detailed in Section 2. Section 3 shows experimental results and analysis on real-world and synthetic networks. In section 4, we give the conclusion.

2. METHOD

2.1 Data Flow Field and Community Structure from Flow Perspective

Many processes appearing in complex network, such as epidemic dynamics[22] and cascading failures[23], can be regarded as some kinds of flows in an abstract sense. The differences lie in that each kind of processes has different form and effect[24]. Generally, such flow process can be studied by using the differential equation below:

$$\dot{x}_i= f(x_i,x_{\Gamma i}) \qquad (1)$$

wherex_i and $\Gamma_i$are the state variable and neighbor set of node i, respectively while f is a function where the arguments are the state variables of node $i$and its neighbors. In mathematics and physics, such description method is equivalent to defining a field on the network and equation (1) is the field equation.

The statements above indicates us that there seems to exist some kinds of flows in complex network. In order to describe and make use of these flow processes, we put forward the conception of data flow field for community detection: we use "data" representing the interaction between nodes; the quantity and direction of "data" represent the interaction strength and direction, respectively; in addition, for community detection problem, we focus on the aspect of direction while the quantity of "data" is only used to decide where the "data" should go in next step; for simulating a flow process, we must predefine the direction on each node; once a piece of "data" come to one node, it is then transformed to the next node according to the predefined direction.

After defining a field on network, we must determine what a community looks like from the flow viewpoint. Considering that the spread of interaction or data between nodes in the same community is rapider than that between different communities, we regard the nodes influenced by same "data" flow belong to the same community.

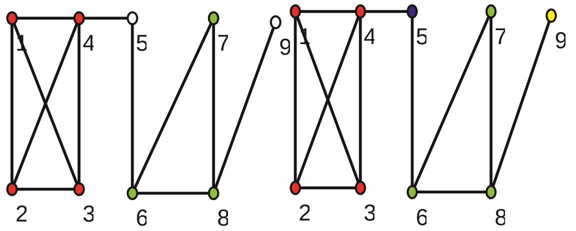
The statements above hint us that flow simulation can be used to find community structure in network as long as we use reasonable flow pattern and appropriately determine the direction predefined on each node.

One of alternative schemes is the simulation of epidemic disease. The reasons why this flow is suitable for detecting community are stated below: consider two members in the same community X and Y who have directed link between them and assume that X has been infected by some kind of virus. Then as X would infect Y through direct contact or indirect contact with their common neighbors, the probability that Y gets infected is proportional to the account of common neighbors of them. This indicates that at next moment, the virus carrier X would infect the one who has the most common neighbors with him. In this way according to the account of their common neighbors, we could determine the direction predefined on each node. As the contacts between members in the same community are more frequent than that between different communities. The virus flow prefers to stay in the same community. At last, according to the kind of virus carried by each node, we can determine the members who are in the same community.

2.2 Community Detection Method Using Information Flow Simulation

We use different labels denoting different viruses. The proposed information flow algorithm (IFA) has two phases:

1) Spread of information in network: First of all, count the numbers of common neighbors between all directed nodes. This results in a 2-tuple set, i.e., $\{((u,v),|\Gamma(u) \cap \Gamma(v)|) : (u,v) \in E,|\Gamma(u) \cap \Gamma(v)| >0\}$, where $\Gamma(u)$ is the neighbor set of u, $|A|$ is the cardinality of set A. We require that in this set for each pair of nodes, the number of common neighbors is larger than zero. Because if the number is zero, we can't decide which way the virus should go. Secondly, we sort the elements in the 2-tuple set according to the account of common neighbors in descending order. Finally, we label each pair of nodes in the set from front to back. We might encounter four kinds of situations: (ii)



(a) (b)

Figure 1. Illustration for explaining rule (iv) of IFA: (a) node 5 and 9 have not been given labels; (b) each of node 5 and 9 is given a unique label. Node colors represent the labels each node own.

A. If each node in the pair has not been labeled, then these two nodes are assigned with same unique labels. It indicates a new information appears in the network and infects the nodes in this pair at the same time;

B. If one of the nodes in a pair (u,v), assuming u, has been given a label L, but node v doesn't have a label yet. Then we give node v the same label L. It indicates the information L is transmitted from node u to v;

C. If each node in the pair has a label, we shouldn't change the labels that they own. Because in this case, each node in the pair has already been infected by certain information and it means that these nodes have found the communities where they belong, then we can't change their labels;

D. After labeling all pairs in the set, there may still exist some nodes without labels (for example, node 5 and 9 in Fig. 1(a)). We don't distinguish these cases and assign these nodes with unique labels (Fig. 1(b)).

2) Integration of quasi community: The second phase is the integration phase. As we will see later, IFA tends to find clique (complete graph) structures in network. We call this phenomenon the "Whirlpool Effect". Though clique structure is also one kind of community, these quasi communities (including cliques) are so strict that such structures prevent IFA from finding reasonable partitions. For solving this effect, we use modularity algorithm[9] to integrate these quasi communities until when combining any two communities would lead to a decrease in modularity.

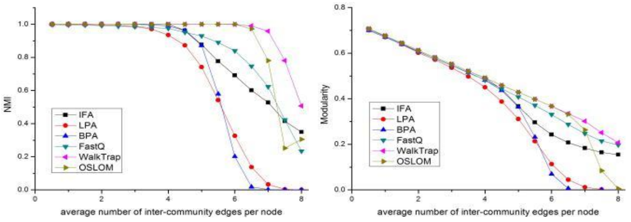
When the integration phase finished, the IFA algorithm finishes. The nodes with same labels belong to same community.

3. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we apply our method to synthetic networks and real-world to verify the rationality and effectiveness.

3.1 Experimental Results on ComputerGenerated Benchmarks

1) Results on GN benchmarks: We firstly apply our algorithm on Girvan-Newman (GN) benchmarks[2]. GN benchmark is consist of 4 planted communities and each community contains 32 nodes. The average degree of a node is fixed to 16 among which Zout edges connect to nodes in different communities and the remaining edges connect to the members in the same community. For each fixed Zout, we generate 20 GN networks and test the methods 100 times on each of them. Thus results for each parameter Zout are averaged over 2000 realizations.

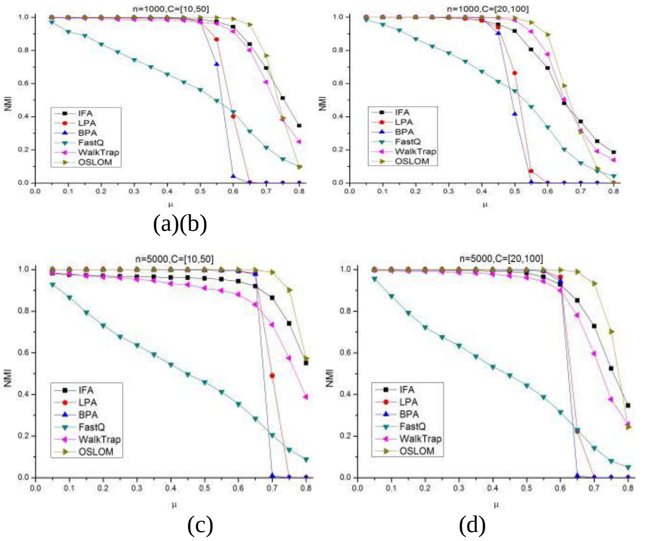


(a) (b)

Figure 2. Tests of some methods and EFA on GN benchmarks. The accuracy is measured by NMI (left) and Modularity (right) metrics.

The results are shown in Fig. 2.From the figures we can see,IFA performs better than LPA and BPA but worse than FastQ,WalkTrap which are based on modularity and OSLOM which is based on statistics. The accuracy of LPA,BPA starts to decline when $Z_{out}$is around 5 while LPA still perform good. This is because that when Z_{out}increases,the community structure would become unclear, the label in label propagation algorithm(LPA,BPA) would flow to another community. But in IFA,high density place still exist in networks with fuzzy community structure, when information flow from high density place to low density place, all the information would compete at the border of community and it's hard for them to beyond the border. On the other hand,GN network is homogeneous, so the FastQ and WalkTrap which expecting the max modularity and OSLOM which focusing to optimizing the statictics perform better than IFA.

2) Results on LFR benchmarks: The node degrees and sizes of communities in real-world networks satisfy the power-law distribution while GN benchmarks can't generate such graphs. So we further test these methods on LFR benchmarks [29]. LFR benchmark can generate testing networks satisfying the power-law distribution. LFR benchmark has eight parameters: we set the size of network to 1000; the power-law exponents of node degrees and community sizes are set to 2 and 1 respectively; average degree and maximum degree are set to 20 and 50 respectively; u is the mixing parameter which controls the significance of community structure and it ranges from 0.05 to 0.8; the ranges of community size are specified in the figures. We report average values over 100 realizations and individual runs are shown in scatter plots (top, bottom respectively in Fig. 3).



(a)(b)

(c) (d)

Figure 3. Average NMI of compared methods and IFA on LFR benchmarks. The scale of networks is 1000 and 5000 and the community size covers between [10, 50] and [20, 100] nodes (left, right top, bottom respectively).

Once again, similar results are obtained on LFR benchmarks (Fig. 3) compared with the results on GN benchmarks (Fig. 2). But the differences are more evident on LFR benchmarks.

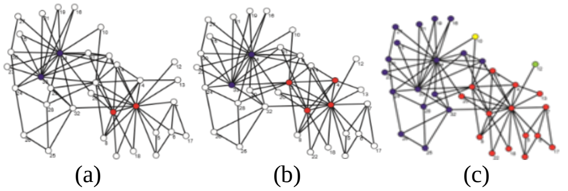
In Fig. 3(a), when the mixing parameter ranges from 0.5 to 0.65, IFA performs better than LPA and BPA; even when BPA and LPA don't perform well (u=0.6, 0.65), IFA can still find reasonable communities. The similar conclusions can be drawn from Fig. 3(b).

But when compared to OSLOM,IFA was better than statistics methods only when the structure of communities are unclear(u=0.8).We guess that when u are large, the community structures will be unclear, the LFR network will be more sensitive to random disturbance, so the local structure will be similar to global structure, then it's hard to distinguish the group that the node belongs to. As for IFA,similar to the analysis of GN,though the community structures are unclear when u is large, as long as there exists some subgraphs which are densely connected,IFA see those subgraphs as the cores of communities, so IFA could find significativecommunities.Furthermore,only when the community’s size are matched to generated network’s size(Fig. 3(c)3(d)),IFA would be better than other method.

3.2 Results on Real-World Networks

1) Case study of Zachary Karate Club: Karate[25] is a social network studied by Zachary from 1970 to 1972 and is made up of 34 members and 78 links. The links were documented by observing the interact between members outside the club. During the study, because a conflict happened between the administrator and instructor, the club was split into two. The right partition is showed in Fig.4(f) and denoted by blue and red nodes, respectively. We call these two parts red community and blue community, respectively. From the figure, we can observe that each community has its own hub, i.e., node 1 in red community and nodes 33, 34 in blue community. Intuitively, the contacts in red community are closer than that in blue community.

We illustrate the community detecting process using IFA algorithm with Fig. 4(a)-(f). Fig. (a) shows the generation of information in network. Because there are maximum common neighbors between nodes in pairs (33,34) and (1,2), this indicates that the nodes in these two pairs contact with each other closely, so these four node are infected by red and blue information firstly. Fig. 4(b) to (e) show the spread of these two information in the network. In this process, we notice that the spread is rapider in the red part than that in the blue part. This phenomenon comes from the intuitiveness that the edges in red community are denser than that in the blue community and at the same time, it also accords with the intuition that information spread rapidly in dense crowd areas. In Fig. 4(e), node 10 and 12 are colored with yellow and green according to rule (iv), respectively. Fig. 4(f) shows the result after the LPA validation phase and is just the right partition. Note that in karate the modularity integration phase is not necessary.



During the second phase of IFA, node 12 will turn red from green. The color transformation of node 10 in this phase is crucial for whether we can find the right partition or not: if node 10 turns blue, we will get the right result, but if it turns red, we then get a partition similar to the real one. However, in a complex network, there may exist overlapping communities and it is hard to decide which community some special nodes belong to. So to some extent, it is acceptable to assign node 10 to either blue community or red community.

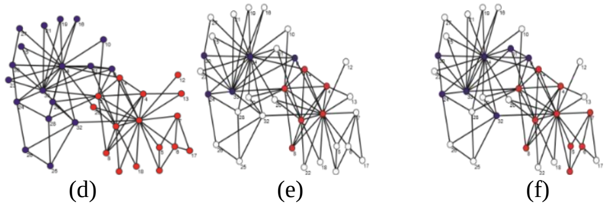


Figure 4. Information spreading process on Karate network using IFA for community detection.

2) Detailed results on real-world networks: We further give the detailed results on real-world networks including Karate[25], Football[2], Risk[26] and polBooks[27] compared with LPA and BPA methods. The statistics of these networks are shown in Table 1. The accuracy is measured by standard metrics: normalized mutual information[28](NMI), variation of information[16](VOI) and Modularity[9]. Because all these algorithms are randomized methods, every result is taken average over 10000 realizations and the final results are shown in Table 2.

From Table 2, IFA performs better than LPA and BPA measured by NMI metric except on polBooks network and the results indicate that IFA could find community structure which is similar to the grounding truth. The larger modularity of the found community by IFA shows our algorithm can find reasonable partitions while the lower VOI values indicate that our algorithm has higher robustness compared with LPA and BPA. In fact, a deterministic community detection method would has zero VOI value, but because of the existence of overlapping communities, a community detection algorithm possibly can’t find same partition all the time.

Table 1. Thestatisticsof fourreal-word networks

Algorithms	Nodes	Edges	Communit y count	Average size of communities
Karate	34	78	2	17
Football	115	613	12	9.58
Risk	42	83	6	7
polBooks	105	441	3	35

Table 2. The results measured by NMI, MODULARITY and VOIonfour real-world networks. The numbers in bracketsrepresente standard deviations. Every result is averaged over 10000 implements

Algorithms	Karate	Football	Risk	polBooks
NMI				
LPA	0.6505(0.186)	0.8882(0.027)	0.8389(0.046)	0.5245(0.026)
BPA	0.6513(0.341)	0.8805(0.024)	0.6983(0.698)	0.5613(0.037)
IFA	0.8748(0.098)	0.9233(0.007)	0.8676(0.039)	0.5423(0.017)
Modularity				
LPA	0.355(0.077)	0.587(0.015)	0.593(0.593)	0.504(0.016)
BPA	0.287(0.136)	0.599(0.065)	0.502(0.094)	0.459(0.019)
IFA	0.373(0.02)	0.601(0.0007)	0.618(0.014)	0.514(0.008)
*VOI*LP				
A	0.191(0.087)	0.071(0.036)	0.150(0.052)	0.085(0.045)
BPA	0.141(0.083)	0.068(0.033)	0.170(0.079)	0.075(0.044)
IFA	0.060(0.045)	0.010(0.011)	0.100(0.041)	0.053(0.032)

3)Whirlpool Effect: In this section, we illustrate the “Whirlpool Effect” by using a toy model with two communities to explain why IFA method is prone to find clique structures in network.

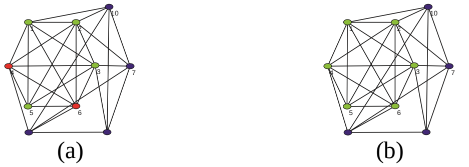


Figure 5. Toy model for explaining Whirlpool Effect: (a) the updating order of nodes is successively colored by red, green and blue; (b) the final partition found by IFA (colored by green and blue, respectively).

In reality, once an object with small inertial mass falls into a whirlpool in the water and without the help of external force, the small object is hard to escape from the whirlpool by itself. It seems that the whirlpool grasps the object tightly, keep it rotating in the vortex. Similarly, our algorithm also has such effect. Intuitively, the whirlpool effect of IFA refers to the phenomenon that if we regard the labels as the tiny object with small inertial mass, during the first phase of our algorithm, the labels seldom go across the boundary of two communities with apparent different link densities, i.e., the label is dragged by the community where it belongs and is difficult to intrude into other communities.

Table 3. Cliques found by LPA and IFA onreal-world networks

Algorithms	Karate	Risk	Football	polBooks	Total
LPA	487	1191	4179	1243	7100
IFA	324	1403	7570	171	9468

For example, the toy model (shown in Fig. 5) aiming to illustrate this effect is composed of 10 nodes. These 10 nodes can be further grouped into two cliques, i.e., {1,2,3,4,5,6}, {7,8,9,10}. We refer to these two cliques as C1 and C2, respectively. According to our algorithm, the labeling order is successively colored by red, green and blue in Fig. 5(a). In Fig. 5(a), we notice that the red and green labels don't invade C2 and it looks like that the red and green labels in C1 are constrained by attraction of C1. Likewise, the blue labels in C2 also don't invade C1. Fig. 5(b) shows the final result and indicates that the two clique structures in the toy model are separated and found completely by IFA.

In fact, during the first phase of IFA, if IFA hasn't finish labeling the nodes in denser areas, nodes in less denser areas can never be labeled. From Fig. 5, we can see that edge density in C1 are denser than that in C2. If we pay attention to the labeling order, all nodes in C1 (red, green) have been labeled before labeling the nodes in C2 (blue). This character implies that IFA separates denser areas in network while clique structure happens to be one kind of such dense areas.

To investigate to what extent our algorithm can find out clique structure, we run our algorithm without integration phase and LPA on four real-world networks 1000 times. The number of found cliques are recorded in Table 3. In this experiment we require that the size of cliques is larger than or equal to 3. As we can see from the table, our algorithm can find out more cliques than LPA in these networks. So in some kinds of networks, IFA tends to find cliques. It is crucial for IFA algorithm to integrate the quasi communities (cliques) using methods such as greedy modularity algorithm.

4. CONCLUSION

In this paper, we start from the consideration of flow process existing in the network and realize such process can be simulated through properly defining fluxion direction on each node. We proposed a new flow-based community detection algorithm, the information flow algorithm, by simulating the information spreading process in network. Experimental results on real-world and synthetic networks show that the information flow algorithm enjoys higher accuracy and robustness compared with typical algorithms. When the community structure gets vaguer, especially when LPA and BPA don't perform well, IFA method can still find some reasonable and accurate partitions. Though IFA to some extent simulates the flow propagation process in network, the rules are oversimplified. In the future, we will improve these rules for flow simulation and search for other kinds of flow patterns that can be applied to community detection problem.

5. ACKNOWLEDGMENTS

This work is supported by the Natural Science Foundation of China (NSFC) under Grant No. (61472381, 61472382, 61572454 and 61174144), NOE-Micrsoft Key Laboratory of Multimedia Computing and Communiction Foundation, Anhui Province Key Laboratory of Software in Computing and Communication.

6. REFERENCES

[1] M. E. J. Newman.2001.The structure of scientific collaboration networks, *Proc. Natl. Acad. Sci. USA.*, 98(2), 404-409.

[2] M. Girvan and M. E. J. Newman.2002.Community structure in social and biological networks, *PNAS*, 99(12), 7812-7826.

[3] D. J. Watts and S. H. Strogatz.1998.Collective dynamics of small-world networks, *nature*, 393(6684), 440-442.

[4] A. L. Barabasi and R. Albert. 1999.Emergence of scaling in random networks, *science*, 286(5439), 509-512.

[5] X. F. Wang and G. Chen. 2003.Complex networks: small-world, scale-free and beyond", *Circuits Syst. Mag.*, 3(1), 6-20.

[6] A. Clauset, C. Moore and M. E. J. Newman. 2008.Hierarchical structure and the prediction of missing links in networks, *nature*, 453(7191), 98-101.

[7] M. Grivan and M. E. J. Newman. 2002.Community structure in social and biological networks, *PNAS*, 99(12), 7821-7826.

[8] M. E. J. Newman. 2003.Mixing patterns in networks, *Phys. Rev. E*, 67(2): 026126.

[9] M. E. J. Newman.2004.Fast algorithm for detecting community structure in networks, *Phys. Rev. E*, 69(6): 066133.

[10] A. Clauset, M. E. J. Newman and C. Moore.2004.Finding community structure in very large networks, *Phys. Rev. E*, 70(6): 066111,

[11] R. Guimera, M. Sales-Pardo and L. A. N. Amaral,2004.Modularity from fluctuations in random graphs and complex networks, *Phys. Rev. E*, 70(2): 025101.

[12] S. Li, Y. Chen, H. Du and M. W. Feldman.2010.A genetic algorithm with local search strategy for improved detection of community structure, *Complexity*, 15(4): 53-60.

[13] V. D. Blondel, J. L. Guillaume, R. Lambiotte and E. Lefebvre.2006.Fast unfolding of communities in large networks, *J. Stat. Mech.*,74(3): 036104.

[14] M.E.J.Newman.2006.Finding community structure in networks using the eigenvectors of matrices.*Rev.E*,74(3): 036104.

[15] U. N. Raghavan, R. Albert and S. Kumara.2007.Near linear time algorithm to detect community structures in large-scale networks, *Phys. Rev. E*, 76(3): 036106.

[16] L.Subelj,M.Bajec.2011.Robust network community detection using balanced propagation,*Eur.Phys.J.B*,81(3): 353-362.

[17] P. Pons and M. Latapy.2005.Computing communities in large networks using random walks, *Proc. 20th Int'l Symp. Computer and Information Sciences(ISCIS'05)*, 284-293.

[18] K. Steinhaeuser and N. V. Chawla.2010.Identifying and evaluating community structure in complex networks, *Pattern Recognit. Letters*, 31(5): 413-421.

[19] J. Chen, O. R. Zaiane and R. Goebel.2009.Local community identification in social networks, in *ASONAM'09*, 237-242.

[20] A. Clauset,2005.Finding local community structure in networks, *Phys. Rev. E*, 72(2): 026132.

[21] F.Luo,J.Z.Wang, E.Promislow, 2006. Exploring local community structures in large networks,*WI'06: Proceedings of the 2006 IEEE/WIC/ACM Int.Conf.on Web Intell.*, 233-239.

[22] R. Pastor-Satorras and A. Vespignani. 2001.Epidemic spreading in scale-free networks, *Phys. Rev. E*, 86(14): 3200.

[23] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley and S. Havlin,2010. Catastrophic cascade of failures in interdependent networks, *nature*, 464(7291): 1025-1028.

[24] K. Narang,K.Lermanm,d P. Kumaraguru,2013.Network flows and the link prediction problem, *in: Proc. of the 7th workshop on social network mining and analysis. New York: ACM*, 3.

[25] W. W. Zachary,1977.An information flow model for conflict and fission in small groups, *J. Anthropological Res.*, 452-473.

[26] K.Steinhaeuser, N.V.Chawla.2010.Identifying and evaluating community structure in complex networks, *Pattern Recognit. Letters.*, 31(5): 413-421.

[27] V. Krebs.2008.A network of co-purchased books about U.S. Politics, [Online]. Available: http://www.orhnet.com/.

[28] L. Danon, A. Diaz-Guilera J. Duch and A. Arenas.2005.Comparing community structure identification, *J. Statistical Mechanics*, 2005(9): P09008.

[29] A. Lancichinetti, S. Fortunato and F. Radicchi.2008.Benchmark graphs for testing community detection algorithms, *Phys. Rev. E*, 78(4): 046110.