Using network topology and rule-based strategy to identify community structure in social networks

Yu-Hsiang Fu^{a,b}, Chung-Yuan Huang^{b,*}, Chuen-Tsai Sun^a

^aDepartment of Computer Science, National Chiao Tung University, Hsinchu 300, Taiwan, ROC

^bDepartment of Computer Science and Information Engineering, Chang Gung University, Taoyuan 333, Taiwan, ROC

*Corresponding Author: gscott@mail.cgu.edu.tw

Abstract

Community detection is to identify groups of nodes or network partitions that the edge connectivity is tight inside communities and loose between communities. Network type variation results in different meanings of the community structure, for example, the community structure corresponds to a set of web pages on the same topic in World Wide Web, or to target groups of customers exhibiting the same purchase behaviors in a social network. Hence, many approaches like evolutionary computation, data mining, modularity optimization, density-based and topology-based methods have been proposed to address community detection in many disciplines. In this study, we propose a hierarchical arc-merging algorithm using network topology structure and rule-based arc-merging strategy to identify community structure and to reveal network hierarchy. Five well-known social networks with ground-truth-community are used to verify the correctness of the identified community structure, and two synthetic networks are used to examine the effectiveness of avoiding the resolution-limit problem. The experimental results show that the identified community of the proposed method is more close to the ground-truth-community and also overcomes the resolution-limit problem.

Keywords: complex network, social network, community detection, network hierarchy.

1. Introduction

Many real world systems can be formed as networks which consists of sets of nodes and edges⁽¹⁻³⁾. For example, in a social network, nodes represent persons and edges describe friendships between any two of them; in a collaboration network, nodes are authors and edges are the

cooperation of authors in a paper; in Word Wild Web (WWW), nodes are web pages and edges are hyper-links from one page to other pages. Networks have been investigated and found many important properties⁽¹⁾, such as “small-world effect” indicates high clustering coefficient and shorter average path length between nodes in a network, “long-tail” describes the degree distribution of a network follows the power-law which means many nodes have a small number of connections and few nodes have a lot of connections in a network, and “community structure” represents nodes tightly connected to each other with similar features in the same group and loosely connected to nodes of other groups.

The goal of community detection is to identify groups of nodes or network partitions satisfying a specific criterion such as compactness of edge connectivity^(1,4). Network type variation results in different meanings⁽¹⁻²⁾—for example, in terms of community structure corresponding to a set of web pages on the same topic in WWW; of social groups in an acquaintance network; of a circuit, pathway or motif serving a certain synthesizing or regulating function in a metabolic network; or of a target group of customers exhibiting the same purchase behaviors in a social network. Hence, the community detection approach is widely used to identify community structures or social groups in networks for many practical application, e.g. marketing and disease prevention.

Accordingly, community detection problem can be reduced to a graph partition problem. However, the reduced problem is a NP-Complete (NPC) problem⁽⁵⁾. Therefore, in many disciplines, a large number of approaches have been proposed to find approximate solutions of community detection problem. For example, in computer science, the evolutionary computation (EC) approaches such as genetic algorithm (GA)⁽⁶⁻⁷⁾, ant colony optimization (AOC)⁽⁸⁾ and

particle swarm optimization (PSO)⁽⁹⁾ and artificial intelligence (AI) approaches such as greedy algorithm⁽¹⁰⁾ and simulate annealing (SA)⁽¹¹⁾ are often used for NPC problems; in network science, tradition hierarchical clustering algorithms (e.g. GN⁽²⁾ and FN⁽¹⁰⁾) were proposed and used in the early stage, and the modularity optimization algorithms (e.g. CNM⁽¹²⁾ and BGLL⁽¹³⁾), k -medoids algorithm⁽¹⁴⁾, density-based algorithms (e.g. DenShrink⁽¹⁵⁾ and ImDS⁽¹⁶⁾) and topology-based algorithm (e.g. LPW⁽¹⁷⁾) were proposed in recent development.

Modularity^(10,18) has been widely used to evaluate the quality of community structure resulting from the above methods. Modularity evaluates the density of edges inside communities as compared to the edges between communities. Higher modularity value indicates the quality of community structure or network partition is better than lower one. Hence, modularity is used as fitness function in the evolutionary computation approaches and is used as objective function in the network science approaches for finding the best approximate solution of community detection. However, the resolution-limit⁽¹⁹⁾, a serious and inherent problem of modularity has been mentioned and investigated when the size of a community is smaller than a certain scale may not be resolved and will be merged into a bigger community. Therefore, the method uses modularity as fitness or objective function may suffering from the resolution-limit problem, but density-based algorithms are not.

In this study, we propose a hierarchical arc-merging (HAM) algorithm which takes advantages of topology-based and density-based algorithms using network topology and rule-based arc-merging strategy to identify community structure and to reveal network hierarchy. Similarity measures are used to evaluate how similar any two nodes are and to rank the edge importance. Rule-based arc-merging strategy is applied to identify communities and to reveal hierarchical network structures. In the preliminary experiment, five well-known social network datasets are used to verify the correctness of finding ground-truth-community, and the two synthetic networks are used to examine whether HAM suffers from the resolution-limit problem. According to the experimental result, the identified community structure of HAM is close to or even identical to the ground-truth-community of social networks. Also, HAM is capable of overcoming the resolution-limit problem.

2. Background

To represent a social network, let an undirected and

weighted graph $G = (V, E, S)$, where V is the node set, E the edge set and S the similarity weight of edges. $|V|$ denotes the number of nodes, $|E|$ the number of edges and $|S|$ the sum of similarity weight of edges. The network topology is represented as an adjacency matrix $A = \{a_{ij}\}$ and $a_{ij} \in R^n$, where $a_{ij} = 1$ if an edge exists between node i and j , otherwise $a_{ij} = 0$. The weight of each edge is s_{ij} , where $s_{ij} = 1$ if the similarity of node i and j is identical and $a_{ij} = 1$, otherwise $s_{ij} = 0$.

2.1 Similarity measures

Similarity measures are used to determine the weight of edges of a network⁽²⁰⁾. The most common approach of determining the similarity weight of node i and j , $s_{ij} = S_x(i, j)$, is to calculate the number of common neighbors $S_{cn}(i, j)$ as (1). High similarity weight means the two connected nodes share a lot of common neighbors. If two nodes share the same neighbors meaning that these two nodes are considered structurally equivalence. Further, $S_{cn}(i, j)$ becomes different similarity measures using various ways of normalization such as cosine similarity, Jaccard index or minimum similarity which are defined as:

$$S_{cn}(i, j) = |\Gamma(i) \cap \Gamma(j)| \quad (1)$$

$$S_{cosine}(i, j) = \frac{|\Gamma(i) \cap \Gamma(j)|}{\sqrt{|\Gamma(i)| |\Gamma(j)|}} \quad (2)$$

$$S_{Jaccard}(i, j) = \frac{|\Gamma(i) \cap \Gamma(j)|}{|\Gamma(i) \cup \Gamma(j)|} \quad (3)$$

$$S_{min}(i, j) = \frac{|\Gamma(i) \cap \Gamma(j)|}{\min(|\Gamma(i)|, |\Gamma(j)|)} \quad (4)$$

where $\Gamma(i)$ is neighbor set of node i , $\Gamma(j)$ the neighbor set of node j , $|\Gamma(i)|$ the number of neighbor set of node i , and $|\Gamma(j)|$ the number of neighbor set of node j .

2.2 Community quality measure

Since there are many possible partitions of a given network; hence, how to measure network partition quality is an important issue. Modularity^(10,18) has been widely used as fitness or objective function to evaluate community structure quality in many approaches. Regarding modularity, a meaningful network partition is one with many edges inside communities and only a few between communities. The “meaningful” means that the number of connections inside a community should larger than the expected value of randomized connections of a community with the same size

and degree sequence. Hence, the randomized network is used as the null model in modularity. For a given network with M communities, modularity Q is defined as

$$Q = \sum_{i=1}^M (e_{ii} - a_i^2) = \sum_{i=1}^M \left(\frac{l_i}{|E|} - \left(\frac{d_i}{2|E|} \right)^2 \right) \quad (5)$$

where e_{ii} is the proportion of edges with two endpoints in community i , a_i the proportion of edges with at least one endpoint in community i , l_i the number of edges with two endpoints within community i , and d_i the sum of the degree of nodes in community i .

A serious and inherent problem of modularity called resolution-limit has been mentioned⁽¹⁹⁾. In the modularity optimization algorithm, the size of some communities may not be resolved, and will be merged into bigger communities which means modularity may miss some important substructures of network. Therefore, if only using modularity to identify communities should consider how to avoid the resolution-limit problem.

3. The proposed method

In this study, we propose a hierarchical arc-merging (HAM) algorithm using network topology and rule-based arc-merging strategy to identify community structure and to reveal network hierarchy. First, edges are classified into three classes according to the similarity weight of each edge. Second, the rule-based strategy is applied to merge nodes, to find communities, and to construct hierarchical network structures.

3.1 Edge classification

Since similarity measures are used to calculate the similarity weight of edges, and then the edges are classified into three classes: weighted-edge W , bridge-edge B and sink-edge S . After that, each edge class is sorted in decreasing order by two indexes which are the weight of edges and the sum of weights between two endpoints and their neighbors. The weighted-edge is a set of edges which the weight of node i and j is larger than zero, i.e. $s_{ij} > 0$. The bridge-edge is a set of edges with the weight equals to zero, i.e. $s_{ij} = 0$, which means these edges are connecting different communities or components in a network. The sink-edge is a set of edges which either degree of node i or j equals to one, i.e. $k_i = 1$ or $k_j = 1$; additionally, in a weighted graph, the weight of node i and j also equals zero, i.e. $s_{ij} = 0$. Hence, $V = V^W \cup V^B \cup V^N$ and $E = W \cup B \cup N$, where V^W is node set of weighted-edge, V^B the node set

of bridge-edge, and V^N the node set of sink-edge.

3.2 Rule-based arc-merging strategies

The idea of arc-merging is to merge the two endpoints of edges into super-nodes according to the decreasing order of similarity of edges, and super-nodes are as seeds to attract the nearby non-merged nodes to become larger communities or substructures. Therefore, we define four arc-merging rules to achieve the idea, and these rules can be combined to build different strategies for different edge classes. For each edge (e_i, e_j) in E , the arc-merging rules are defined as

Rule-1 (R1): if node e_i and node e_j are not merged, then merge these two nodes into a super-node.

Rule-2 (R2): if node e_i is merged but node e_j is not, then merge node e_j into the super-node of node e_i .

Rule-3 (R3): if node e_i is not merged but node e_j is, then merge node e_i into the super-node of node e_j .

Rule-4 (R4): otherwise, pass.

According to edge classification, two kinds of strategies can be built using the above arc-merging rules: weighted-merging strategy and non-weighted-merging strategy, which are defined as

Weighted-merging strategy (T1): use R1, R2, R3 and R4.

Non-weighted-merging strategy (T2): use R2, R3 and R4.

First, the weighted-merging strategy consists of all four rules for merging the weighted-edge W . For each edge in W , R1 is used to merge the two endpoints of edges with high similarity into super-nodes as seeds, R2 and R3 are used to attract the nearby non-merged nodes of super-nodes, and R4 is used to handle the other cases. Second, the non-weighted strategy consists of three rules excluding R1 for merging the bridge-edge B and sink-edge S . After applying T1, all endpoints of weighted-edge W are merged into super-nodes. Hence, the major communities or substructures of network are almost found. However, some endpoints of bridge-edge B and sink-edge S are not. Therefore, the non-weighted strategy is used to process these nodes.

3.3 Hierarchical arc-merging (HAM) algorithm

In HAM, the architecture of community detection consists of two phases: the original network phase and the super-node network phase, and is shown in Fig. 1. In the original network phase, first, the similarity measure is used to calculate the similarity weight of edges. Second, the T1

Fig. 1. The architecture of community detection in HAM.

strategy is applied to process weighted-edge W . After that, the nodes of bridge-edge B and sink-edge S remain non-merged, i.e. $V - V^W = V^B + V^S$. Third, the T2 strategy is applied to process the remained nodes. Finally, every node of original network is merged to one community and the high-level network structures are constructed by super-nodes.

In the super-node network phase, first, the similarity weight of edges are given by calculating the average of similarity weight of edges between the corresponding communities. Second, classify edges into three classes, i.e. W , B and S . Third, the T1 strategy is applied to process weighted-edge W . Fourth, the T2 strategy is applied to process sink-edge S . Finally, higher-level network structures are constructed. According to the edges between super-nodes should be used to connect different communities, therefore the T2 strategy is not applied to the bridge-edge B . The procedure stops until the weighted-edge W is empty. The details of HAM is shown in algorithm 1.

4. Preliminary experimental results

In this study, we use five well-known social networks with ground-truth-community to analyze how well the identified communities match the real communities, and use two synthetic networks to test whether HMA suffers from the resolution-limit problem. The giant connected component (GCC) of these networks used in the experiment is shown in Table 1.

4.1 Real social networks

The Zachary's karate club network⁽²¹⁾, there are 34 nodes and 78 edges in this network which nodes are club members and edges are friendships between members. Due to a conflict of disagreement between administrator of the club and the club's instructor, and then the club is split into two groups. The instructor was leaving the club, taking a half of club members and creating a new club.

Algorithm 1: Hierarchical Arc-Merging (HAM)

Input: a network $G = (V, E)$.

Output: community detection results C .

```

1: assign  $level \leftarrow 0$ 
2: calculate similarity weight of edges of original
   network
3: create  $C$  to preserve community results
4:  $level \leftarrow level + 1$ 
5: while true do
6:   classify edges into  $W$ ,  $B$  and  $S$  and sort edges
7:   if  $W$  is empty:
8:     break
9:   if  $level$  is 1: // process the original network
10:    for  $(ei, ej)$  in  $W$ :
11:      apply T1 strategy, use R1, R2, R3 and R4
        rules
12:    for  $(ei, ej)$  in  $B$  and  $S$ :
13:      apply T2 strategy, use R2, R3 and R4 rules
14:    else: // level > 1, process the super-node network
15:      for  $(ei, ej)$  in  $W$ :
16:        apply T1 strategy
17:      for  $(ei, ej)$  in  $S$ :
18:        apply T2 strategy
19:    create super-node network of higher level
20:    calculate similarity weight of edges of super-node
      network
21:    preserve community results of current  $level$  in  $C$ 
22:     $level \leftarrow level + 1$ 
23: return  $C$ 

```

The Dolphins network⁽²²⁾, this network consists of 62 bottlenose dolphins living in Doubtful Sound, New Zealand, and 159 interactions of dolphin pairs occurring more often than expected by chance. The network is constructed by the observations of seven years from 1994 to 2001. This network can be divided into two groups by following the departure of a key individual of the population.

The United States college football network⁽²⁾, there are 115 teams which play 613 games during the regular fall season of 2000. Nodes are represented as teams and edges are represented a game play of two teams. The teams can be divided into 12 conferences which the teams play more games in the same conference than the games between different conferences.

The political book network⁽²³⁾, this network is constructed by users' historical purchase behaviors of buying political books on Amazon.com website. There are 105

books and 441 edges in this networks. Nodes are books and edges are co-purchasing relationships which a user bought one book and also bought another one book. The books are manually classified into three classes which are conservative, neutral and liberal.

The political blogs network⁽²⁴⁾, this network is composed of blogs which are personal or group web diaries about U.S. presidential election of 2004 and the web links between blogs. There are 1222 blog websites as nodes and 16714 edges between nodes. The blogs are manually labeled by its political orientation, and can be divided into two groups which are conservative and liberal.

4.2 Synthetic networks

Two synthetic networks⁽¹⁹⁾ are used to examine whether the proposed method suffers from the resolution-limit problem. The first synthetic network is called RING, the RING is constructed by a ring of r cliques (with r is even) connected by a single edge. Each clique is a complete graph K_p which consists of p nodes and $[p(p-1)]/2$ edges. The second synthetic network is called PAIR, the PAIR is constructed by two parts. The first part is two K_p complete graphs connected by a single edge. The second part is two complete graphs K_q also connected by a single edge. One clique of first part connects to two cliques of second part using an edge respectively. The parameters of RING are $p = 5$ and $r = 30$, and RING consists of 150 nodes and 330 edges. The parameters of PAIR are $p = 20$ and $q = 5$, and RING consists of 50 nodes and 404 edges.

4.3 Community validation

Normalized mutual information (NMI) is used to measure the quality of community structures identified by an algorithm that means calculating the degree of similarity between true partition X and identified partition Y . The NMI is defined as

$$NMI(X, Y) = \frac{2 \cdot I(X, Y)}{H(X) + H(Y)} \quad (6)$$

$$I(X, Y) = \sum_x \sum_y P(x, y) \log \frac{P(x, y)}{P(x)P(y)} \quad (7)$$

where $H(X) = \sum_{x \in X} P(x) \log P(x)$ is the Shannon entropy of true partition X and $P(x)$ the proportion of community x in partition X . $H(Y)$ is the Shannon entropy of identified partition Y and $P(y)$ the proportion of community y in partition Y . $I(X, Y)$ is the mutual information for X and Y indicating the level of correct Y identified by an

algorithm when X is given. $P(x, y)$ is the percentage of overlap between X and Y . If $NMI(X, Y) = 1$, the two partitions are identical; otherwise they are considered independent.

4.4 Experimental results

The community detection results of HAM are comparing to conventional methods (e.g. CNM⁽¹²⁾ and LGA⁽⁷⁾) and the referred values of topology-based algorithm (e.g. LPW⁽¹⁷⁾) are shown in table 1. The configuration of LGA used in the experiment is that the gene representation is locus-based representation and the genetic operations are truncation selection, uniform crossover and bit-by-bit mutation. The parameters of genetic operations are that selection rate is 0.1, crossover rate is 0.8 and mutation rate is 0.05. LGA performs 10 runs evolution, each evolution consists of 200 generations. The population size is 50.

In table 1, the community detection results include the number of identified community, number of level, modularity value and NMI value. The results of number of identified community can be compared to the number of ground-truth-community that means how many true communities can be found by the method. The results of number of level shows the best NMI value in which level and how many levels constructed by HAM. For example, the “2/3” of dolphins network means that the large NMI value is in level 2 and there are three levels constructed by HAM. In the results of modularity, the best modularity values are found by CNM or LGA, because CNM and LGA are trying to maximize the modularity.

However, HAM cannot find the best modularity value according to the mechanism of modularity optimization is absent. In the NMI results, HAM can find higher NMI value than other methods that means the identified communities of HAM are more close to the ground-truth-community. The subscripts of NMI values, e.g. “ c ”, “ j ” and “ m ”, means using cosine similarity, Jaccard index and minimum similarity in HAM respectively. Further, the RING and PAIR networks are used to examine whether the method overcomes the resolution-limit problem. For example, the method is free from resolution-limit problem in case of all ground-truth-community of PAIR (or RING) network can be identified. The results shows that HAM overcomes the resolution-limit problem as shown in Fig. 2.

The Fig. 3 shows the network hierarchy demonstration of dolphins network. The Fig. 3a is the connection structure of original network, i.e. level 0; the solid-line indicates the weighted-edge; the dash-line is represented the bridge-edge;

Table 1. The giant connected component (GCC) and community detection results of networks.

Network		karate	dolphins	football	polbooks	polblogs	RING	PAIR
# node		34	62	115	105	1222	150	50
# edge		78	159	613	441	16714	330	404
# ground-truth-community		2	2	12	3	2	30	4
# identified community	HAM	2	2	13	2	2	30	4
	CNM	3	4	7	4	11	17	3
	LGA	4	4	7	4	24	22	3
	LPW	2	3	11	3	-	-	-
# identified level	HAM	2/3	2/3	2/5	2/3	3/4	1/1	1/1
Modularity	HAM	0.3715	0.3787	0.5253	0.4569	0.4233	0.8758	0.5416
	CNM	0.3807	0.4955	0.5773	0.5020	0.4269	0.8863	0.5426
	LGA	0.4198	0.5268	0.5027	0.5265	0.3269	0.8822	0.5426
	LPW	-	-	-	-	-	-	-
NMI	HAM	<i>1.0_m</i>	<i>0.8889_m</i>	0.8403 _{c,j}	<i>0.6069_m</i>	<i>0.7011_m</i>	<i>1.0_{c,j,m}</i>	<i>1.0_{c,j,m}</i>
	CNM	0.7069	0.6020	0.7698	0.5314	0.6609	0.9074	0.9401
	LGA	0.7236	0.6132	0.6930	0.5397	0.3931	0.9441	0.9401
	LPW	<i>1.0</i>	0.7621	<i>0.9720</i>	0.5780	-	-	-

Bold, best modularity value;

Bold and Italic, best NMI value;

Subscript, similarity measures of the best NMI value.

and dot-line shows the sink-edge; according to the clarity of connection structure, the nodes are not shown. The Fig. 3b is the ground-truth-community of dolphins network. The Fig. 3c and 3d show the projection from the specific level of super-node network to the level of original network. For example, the Fig. 3d exhibits the projection from level 2 of super-node network to the level 0 of original network, and also presents that each node belongs to which high level community. Each level of super-node network exhibits various community structure, i.e. the higher level structures are constructed by lower level structures. According to only one community in level 3, therefore the visualization of level 3 is not shown.

5. Conclusion and future work

In this study, we propose a hierarchical arc-merging (HAM) algorithm using network topology structure and rule-based arc-merging strategy to identify community structure and to reveal network hierarchy. Five well-known social networks and two synthetic networks with ground-truth-community are used to analyze and to test the proposed

method. The experimental results indicates that HAM is capable of identifying the community structures with higher NMI value that the identified community result is more close to the ground-truth-community. The experimental results also show that HAM overcomes the resolution-limit problem. Although HAM shows its advantages in identifying more correct community detection results; however, we did not do the detailed comparison between HAM and other state-of-art methods such as density-based and the topology-based algorithms in this study. Also, the mechanism of modularity optimization should be taken into consideration when a given network without ground-truth-community. The issues mentioned above and the further improvement are left as future works.

Acknowledgment

The work was supported in part by a Grant from the Republic of China National Science Council (MOST-104-2221-E-182-047). The work was supported in part by the High Speed Intelligent Communication (HSIC) Research Center, Chang Gung University, Taiwan.

References

- (1) M. Newman : “Networks: an introduction”, Oxford University Press, 2010.
- (2) M. Girvan and M. E. J. Newman : “Community structure in social and biological networks”, *Proceedings of the National Academy of Sciences (PNAS)*, Vol. 99, No. 12, pp. 7821–7826, 2002.
- (3) Y.-H. Fu, C.-Y. Huang, and C.-T. Sun : “Using global diversity and local topology features to identify influential network spreaders”, *Physica A: Statistical Mechanics and its Applications*, Vol. 433, pp. 344–355, 2015.
- (4) M. E. J. Newman and M. Girvan : “Finding and evaluating community structure in networks”, *Physical Review E*, Vol. 69, No. 2, p. 026113, 2004.
- (5) S. Fortunato : “Community detection in graphs”, *Physics Reports*, Vol. 486, No. 3–5, pp. 75–174, 2010.
- (6) C. Pizzuti : “A Multi-objective Genetic Algorithm for Community Detection in Networks”, *21st International Conference on Tools with Artificial Intelligence (ICTAI '09)*, pp. 379–386, 2009.
- (7) Y.-H. Fu, C.-Y. Huang, and C.-T. Sun : “Selecting multiple network spreaders based on community structure using two-phase evolutionary framework”, *2015 IEEE Congress on Evolutionary Computation (CEC)*, pp. 2482–2489, 2015.
- (8) C. Honghao, F. Zuren, and R. Zhigang : “Community detection using Ant Colony Optimization”, *2013 IEEE Congress on Evolutionary Computation (CEC)*, pp. 3072–3078, 2013.
- (9) D. Xiaodong, W. Cunrui, L. Xiangdong, and L. Yanping : “Web community detection model using particle swarm optimization”, *2008 IEEE Congress on Evolutionary Computation (CEC)*, pp. 1074–1079, 2008.
- (10) M. E. J. Newman : “Fast algorithm for detecting community structure in networks”, *Physical Review E*, Vol. 69, No. 6, p. 066133, 2004.
- (11) R. Guimerà and L. A. Nunes Amaral : “Functional cartography of complex metabolic networks,” *Nature*, Vol. 433, No. 7028, pp. 895–900, 2005.
- (12) A. Clauset, M. E. J. Newman, and C. Moore : “Finding community structure in very large networks”, *Physical Review E*, Vol. 70, No. 6, p. 066111, 2004.
- (13) V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre : “Fast unfolding of communities in large networks”, *Journal of Statistical Mechanics: Theory and Experiment*, Vol. 2008, No. 10, p. P10008, 2008.
- (14) X. Zhang, J. Zhu, Q. Wang, and H. Zhao : “Identifying influential nodes in complex networks with community structure”, *Knowledge-Based Systems*, Vol. 42, pp. 74–84, 2013.
- (15) J. Huang, H. Sun, J. Han, and B. Feng : “Density-based shrinkage for revealing hierarchical and overlapping community structure in networks”, *Physica A: Statistical Mechanics and its Applications*, Vol. 390, No. 11, pp. 2160–2171, 2011.
- (16) J. Wu, Y. Hou, Y. Jiao, Y. Li, X. Li, and L. Jiao : “Density shrinking algorithm for community detection with path based similarity”, *Physica A: Statistical Mechanics and its Applications*, Vol. 433, pp. 218–228, 2015.
- (17) W. Liu, M. Pellegrini, and X. Wang : “Detecting Communities Based on Network Topology”, *Scientific Reports*, Vol. 4, 2014.
- (18) M. E. J. Newman : “Modularity and community structure in networks”, *Proceedings of the National Academy of Sciences (PNAS)*, Vol. 103, No. 23, pp. 8577–8582, 2006.
- (19) S. Fortunato and M. Barthélemy : “Resolution limit in community detection”, *Proceedings of the National Academy of Sciences (PNAS)*, Vol. 104, No. 1, pp. 36–41, 2007.
- (20) E. A. Leicht, P. Holme, and M. E. J. Newman : “Vertex similarity in networks”, *Physical Review E*, Vol. 73, No. 2, p. 026120, 2006.
- (21) W. W. Zachary : “An Information Flow Model for Conflict and Fission in Small Groups”, *Journal of Anthropological Research*, Vol. 33, No. 4, pp. 452–473, 1977.
- (22) D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten, and S. M. Dawson : “The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations”, *Behavioral Ecology and Sociobiology*, Vol. 54, No. 4, pp. 396–405, 2003.
- (23) V. Krebs, unpublished, <http://www.orgnet.com/>, [Accessed: 23-Nov-2015].
- (24) L. A. Adamic and N. Glance : “The Political Blogosphere and the 2004 U.S. Election: Divided They Blog”, *Proceedings of the 3rd International Workshop on Link Discovery (LinkKDD '05)*, pp. 36–43, 2005.