Tsinghua Science and Technology

Volume 20 | Issue 6 Article 9

2015

Tolerance Granulation Based Community Detection Algorithm

Shu Zhao
the Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Center of Information Support & Assurance Technology, School of Computer Science and Technology, Anhui University, Hefei 230601, China.

Wang Ke
the Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Center of Information Support & Assurance Technology, School of Computer Science and Technology, Anhui University, Hefei 230601, China.

Jie Chen
the Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Center of Information Support & Assurance Technology, School of Computer Science and Technology, Anhui University, Hefei 230601, China.

Feng Liu
the Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Center of Information Support & Assurance Technology, School of Computer Science and Technology, Anhui University, Hefei 230601, China.

Menghan Huang
the Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Center of Information Support & Assurance Technology, School of Computer Science and Technology, Anhui University, Hefei 230601, China.

Recommended Citation

Shu Zhao, Wang Ke, Jie Chen et al. Tolerance Granulation Based Community Detection Algorithm. Tsinghua Science and Technology 2015, 20(6): 620-626.

Tolerance Granulation Based Community Detection Algorithm

Authors

Shu Zhao, Wang Ke, Jie Chen, Feng Liu, Menghan Huang, Yanping Zhang, and Jie Tang

TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 09/13 pp620-626
Volume 20, Number 6, December 2015

Tolerance Granulation Based Community Detection Algorithm

Shu Zhao, Wang Ke, Jie Chen*, Feng Liu, Menghan Huang, Yanping Zhang, and Jie Tang

Abstract: Community structure is one of the most important features in real networks and reveals the internal organization of the vertices. Uncovering accurate community structure is effective for understanding and exploiting networks. Tolerance Granulation based Community Detection Algorithm (TGCDA) is proposed in this paper, which uses tolerance relation (namely tolerance granulation) to granulate a network hierarchically. Firstly, TGCDA relies on the tolerance relation among vertices to form an initial granule set. Then granules in this set which satisfied granulation coefficient are hierarchically merged by tolerance granulation operation. The process is finished till the granule set includes one granule. Finally, select a granule set with maximum granulation criterion to handle overlapping vertices among some granules. The overlapping vertices are merged into corresponding granules based on their degrees of affiliation to realize the community partition of complex networks. The final granules are regarded as communities so that the granulation for a network is actually the community partition of the network. Experiments on several datasets show our algorithm is effective and it can identify the community structure more accurately. On real world networks, TGCDA achieves Normalized Mutual Information (NMI) accuracy 17.55% higher than NFA averagely and on synthetic random networks, the NMI accuracy is also improved. For some networks which have a clear community structure, TGCDA is more effective and can detect more accurate community structure than other algorithms.

Key words: tolerance relation; community; tolerance granulation; Normalized Mutual Information (NMI) accuracy

1 Introduction

Many complex networks in society, nature, and technology display a common feature, called community structure[1]. Such feature presents a nontrivial internal organization of the network and allows us to infer special relationships among the vertices that may not be easily accessible from direct empirical tests[2]. Communities are groups of vertices, many links connect vertices of the same group and comparatively few links join vertices of different groups[1,3]. Communities can be crucial to reveal abundant hidden information and help us to understand the functional properties of the networks[4,5]. Many community detection methods have been proposed over the last few years, within different scientific disciplines such as physics, computer, biology, and social sciences. The competition towards the ideal method aims at two main goals, namely reducing the computational complexity of the algorithm and improving the accuracy in uncovering community structure. The former is a well defined objective: the complexity of an algorithm can be computed analytically in many cases. The main problem is the latter, and that is to estimate the accuracy of a method and to compare it with other methods. This

• Shu Zhao, Wang Ke, Jie Chen, Feng Liu, Menghan Huang, and Yanping Zhang are with the Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, Center of Information Support & Assurance Technology, School of Computer Science and Technology, Anhui University, Hefei 230601, China. E-mail: chenjie200398@163.com.
• Jie Tang is with the School of Computer Science and Technology, Tsinghua University, Beijing 100084, China.
∗ To whom correspondence should be addressed.
Manuscript received: 2015-03-27; revised: 2015-07-15; accepted: 2015-07-22

issue of testing is as crucial as devising powerful algorithms, but until now it has not received the attention it deserves. For some social networks that have community structure, communities may be group of related individuals[6,7] and discovering their communities accurately will help us to understand and exploit networks more effectively. Therefore, an efficient and sound approach that can accurately detect the community structure of networks is needed.

How to accurately identify the community structure is a big challenge. And Normalized Mutual Information (NMI) accuracy[8] is a measure of similarity between real partition and obtained partition. In the past decade, community structure has attracted much research attention from various scientific fields. Many methods have been devised to detect the communities of complex networks, such as spectral bisection methods which draw support from the eigenvectors of Laplace matrix[9], GN algorithm[1] which uses betweenness to measure the importance of each edge and Newman Fast Algorithm (NFA)[10] which depends on modularity to get the optimization. These methods are classical and suitable for dividing various networks, and most of them can achieve decent performance. However, their accuracy of community partition does not have superiority.

In real life, the relation between network vertices is mainly tolerance relation[11], which doesn't have the characteristic of transitivity. These vertices can constitute relevant maximal tolerance granules. The link-density of a maximal tolerance granule is the highest among all kinds of vertex subsets of a network, so a dense-linked community usually contains a large maximal tolerance granule at least.

The authors presented an algorithm called Community Detection Algorithm based on Clustering Granulation (CGCDA) in Ref. [12]. Although the algorithm has lower time complexity and higher modularity, its accuracy is not ideal. In this paper, focusing on how to detect community structure accurately, we further improve CGCDA and propose Tolerance Granulation based Community Detection Algorithm (TGCDA). The algorithm utilizes tolerance relation among vertices to realize granulation of a network initially, namely tolerance granulation, then relies on the maximum granulation coefficient to hierarchically conduct the tolerance granulation operations.

The rest of the paper is organized as follows. The preliminary works are presented in Section 2. Section 3 describes TGCDA and algorithm analysis. In Section 4, we conduct experiments on synthetic and real world networks and analyze the results of experiments. Finally, Section 5 is the conclusion.

2 Preliminary Works

Given a social network, it can be modeled as a graph $G = (V, E)$, where V is the set of $|V| = n$ entities and $E \subseteq V \times V$ is the set of $|E| = m$ undirected links of entities. Assume that the community set is $\mathrm{Gr} = \{\mathrm{Gr}_1, \mathrm{Gr}_2, \cdots, \mathrm{Gr}_l\}, l \leqslant n$, where $\mathrm{Gr}_1 \cup \mathrm{Gr}_2 \cup \cdots \mathrm{Gr}_l = V$ and $\forall i, j$ such that $\mathrm{Gr}_i \cap \mathrm{Gr}_j = \varnothing$. The purpose of community detection is to detect the community set Gr.

In order to describe this problem better, we give the following definitions.

Definition 1 Maximal Tolerance Granule (MTG)

A set of vertices $\mathrm{TG} \subseteq V$ is a tolerance granule in a graph G, if each pair of vertices in TG is connected by an edge in E.

A set of vertices $\mathrm{MTG} \subseteq V$ is a maximal tolerance granule in the graph G, if (1) MTG is a tolerance granule in G and (2) there is no vertex $v \in V \backslash \mathrm{MTG}$ such that $\mathrm{MTG} \cup \{v\}$ is a tolerance granule in G.

We circle two maximal tolerance granules in Fig. 1, and Table 1 displays all maximal tolerance granules and their numbers.

Definition 2 Granulation Coefficient

Assume that Gr_i^m and Gr_j^m denote two arbitrary granules in the granule collection Gr^m of the m-th layer, $\mathrm{GC}\left(\mathrm{Gr}_i^m, \mathrm{Gr}_j^m\right)$ denotes the granulation coefficient between Gr_i^m and Gr_j^m, then $\mathrm{GC}\left(\mathrm{Gr}_i^m, \mathrm{Gr}_j^m\right)$ is defined as

$$\mathrm{GC}\left(\mathrm{Gr}_i^m, \mathrm{Gr}_j^m\right) = \frac{\left|\mathrm{Gr}_i^m \cap \mathrm{Gr}_j^m\right|}{\left|\mathrm{Gr}_i^m \cup \mathrm{Gr}_j^m\right|}.$$

Definition 3 Tolerance Granulation Operations

If the granulation coefficient of a granule pair is the maximum, we will conduct the following tolerance

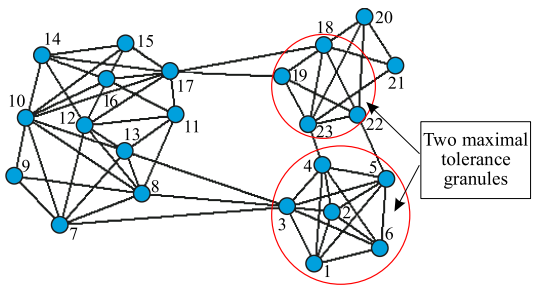


Fig. 1 The schematic network with 23 vertices and 64 edges.

Table 1 All maximal tolerance granules.

MTG	Vertics					
Gr_1	3	7	8	13		
Gr_2	7	8	9	10		
Gr_3	7	8	10	13		
Gr_4	7	8	12	13		
Gr_5	8	11	12	13		
Gr_6	11	12	16	17		
Gr_7	12	14	16	17		
Gr_8	18	19	22	23		
Gr_9	18	20	21	23		
Gr_{10}	18	20	22	23		
Gr_{11}	10	14	15	16	17	
Gr_{12}	17	18	19			
Gr_{13}	4	23				
Gr_{14}	5	22				
Gr_{15}	1	2	3	4	5	6

granulation operations TGO $\left(\mathrm{Gr}_i^m, \mathrm{Gr}_j^m\right)$:

$$\mathrm{Gr}_i^{m+1} \leftarrow \mathrm{Gr}_i^m \cup \mathrm{Gr}_j^m;$$
$$\mathrm{Gr}^{m+1} \leftarrow \mathrm{Gr}^m + \mathrm{Gr}_i^{m+1} - \mathrm{Gr}_i^m - \mathrm{Gr}_j^m.$$

Particularly, if two granule pairs have the maximum granulation coefficient and both possess a common granule, the three granules will be merged together. Similarly, the rest can be done in the same manner.

Definition 4 Granulation Criterion

Considering that there may exist several granules that have some common nodes, we use the extended modularity as the granulation criterion. The extended modularity was proposed by Shen et al.[13], which based on Q function to quantify the strength of overlapping community. It can be defined as

$$\mathrm{GQ} = \frac{1}{2m}\sum_{\mathrm{Gr}}\sum_{i,j\in\mathrm{Gr}}\frac{1}{o_i o_j}\left(A_{ij} - \frac{k_i k_j}{2m}\right),$$

where i and j are two arbitrary vertices, o_i and o_j are the total numbers of granules to which i and j belong respectively, (A_{ij}) is the adjacency matrix, m is the total number of edges. k_i is the degree of vertex i and $k_i = \sum_i A_{ij}$.

Figure 2 shows the granules which can conduct granulation operation and the GQ value at each layer. At the fourth layer, the GQ has a maximum value and three overlapping communities {[3, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17]; [1, 2, 3, 4, 5, 6]; [17, 18, 19, 20, 21, 22, 23]} are obtained.

3 TGCDA

A community can be regarded as a vertex set within which the vertices are more likely connected to each other than to the rest of the network[14]. This indicates that a community usually has relatively high link-density. Generally, the link-density of a maximal tolerance granule is the highest among all kinds of vertex subsets of a network. A dense-linked community usually contains a large maximal tolerance granule at least. Based on this observation, the algorithm TGCDA is proposed as an agglomerative and hierarchical clustering algorithm to identify the community structure, as shown in Algorithm 1.

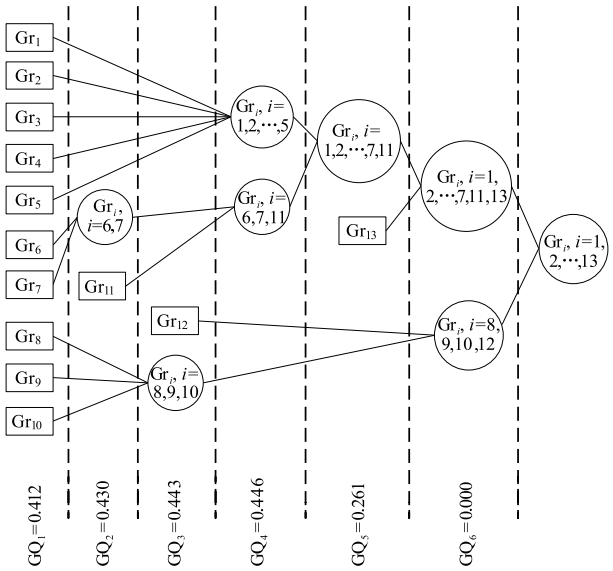


Fig. 2 The process of granulation operation about the schematic network.

Consider an undirected network G with n vertices

Algorithm 1 Tolerance Granulation Based Community Detection

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

Output: Community set $\mathrm{Gr}^s = \{\mathrm{Gr}_1^s, \mathrm{Gr}_2^s, \cdots, \mathrm{Gr}_l^s\}$.

$\mathrm{Gr}^0 \leftarrow \{\mathrm{MTG}_1, \mathrm{MTG}_2, \cdots, \mathrm{MTG}_k\}$, $m \leftarrow 0$ // initialization; k is the number of MTGs, and all vertices of each MTG don't completely belong to the others, m is the number of layers.

Repeat

  for $\forall \mathrm{Gr}_i, \mathrm{Gr}_j \in \mathrm{Gr}^m$ **do**

    calculate $\mathrm{GC}(\mathrm{Gr}_i, \mathrm{Gr}_j)$

  end for

  if $\exists \mathrm{GC}(\mathrm{Gr}_i, \mathrm{Gr}_j)$ is maximum

    $\mathrm{TGO}(\mathrm{Gr}_i, \mathrm{Gr}_j)$

  end if

  calculate GQ;

  $m \leftarrow m + 1$;

Until Gr^m includes one granule

//Select a granule set Gr^s with maximum GQ, s is the layer of maximum GQ.

for $\forall \mathrm{Gr}_i, \mathrm{Gr}_j \in \mathrm{Gr}^s$ **do**

  $\mathrm{Gr}_i' \leftarrow \mathrm{Gr}_i - \mathrm{Gr}_i \cap \mathrm{Gr}_j$, $\mathrm{Gr}_j{}' \leftarrow \mathrm{Gr}_j - \mathrm{Gr}_i \cap \mathrm{Gr}_j$

  $\mathrm{Gr}^s \leftarrow \mathrm{Gr}^s - \{\mathrm{Gr}_i, \mathrm{Gr}_j\} + \{\mathrm{Gr}_i{}', \mathrm{Gr}_j{}'\}$

  $V' \leftarrow \mathrm{Gr}_i \cap \mathrm{Gr}_j$

end for

for $\forall\ v \in V'$ and Gr_i^s **do**

  if edges between v and Gr_i^s is maximum

    $\mathrm{Gr}_i^s \leftarrow \mathrm{Gr}_i^s + \{v\}$

  end if

end for

and m edges. The time complexity of the algorithm is composed of three parts. Firstly, the initialization method for enumerating all maximal tolerance granules was proposed in Ref. [15], which employed a high efficient algorithm to reduce the complexity bound considerably. The complexity of Peamc algorithm runs with $\Delta \times M_{\mathrm{MTG}} \times \mathrm{Tri}^2$ time delay, which is similar to $O(n^2)$, Δ is the maximum degree of G, M_{MTG} represents the size of the maximum tolerance granule, and Tri denotes the number of all triangles in G, respectively. Secondly, the complexity of computing granulation coefficient and conducting tolerance granulation operations is an iterative and hierarchical process. The complexity is close to $O(n^2)$. Thirdly, the time of voting every overlapping vertex is linear, so num overlapping vertices need num $\times O(n)$. In summary, the whole algorithm's time complexity is approximately equal to $O(n^2)$.

4 Experiments

In order to evaluate the performance of the proposed algorithm, we do tests on the datasets of synthetic random networks and real world networks respectively. Meanwhile, we compare it with NFA which was used commonly and CGCDA which was also based on hierarchical clustering.

4.1 Synthetic random networks

The random networks generated from GN-benchmark model[1] were constructed with 128 vertices divided into four communities of 32 vertices each respectively. Edges were placed independently at random between vertex pairs with probability P_{in} for an edge to fall between vertices in the same community and P_{out} to fall between vertices in different communities. The values of P_{in} and P_{out} were chosen to make the expected degree of each vertex equal to 16. For each different P_{in}, we choose 100 synthetic random networks to compute mean Q-value[16] which is used to measure the goodness of community partition and average NMI accuracy[8] which is used to verify accuracy of an algorithm. Higher NMI accuracy indicates the algorithm can detect more accurate community structure. The definition of NMI accuracy is

$$\mathrm{NMI} = \frac{\sum_{i=1}^{|\mathrm{Gr_a}|}\sum_{j=1}^{|\mathrm{Gr_b}|}\frac{n_{i,j}}{n}\log\left(\frac{n_{i,j}}{n}\Big/\left(\frac{n_i^{\mathrm{a}}}{n}\frac{n_j^{\mathrm{b}}}{n}\right)\right)}{\sqrt{\sum_{i=1}^{|\mathrm{Gr_a}|}\frac{n_i^{\mathrm{a}}}{n}\log\left(\frac{n_i^{\mathrm{a}}}{n}\right)\times\sum_{j=1}^{|\mathrm{Gr_b}|}\frac{n_j^{\mathrm{b}}}{n}\log\left(\frac{n_j^{\mathrm{b}}}{n}\right)}},$$

where $\mathrm{Gr_a}$ and $\mathrm{Gr_b}$ are the community sets of networks, $\mathrm{Gr_a}$ is the result of algorithms, and $\mathrm{Gr_b}$ is the real community sets. n_i^{a} and n_j^{b} are the numbers of vertices in the i-th and j-th communities of $\mathrm{Gr_a}$ and $\mathrm{Gr_b}$, respectively, $n_{i,j}$ is the number of vertices both in the i-th community of $\mathrm{Gr_a}$ and the j-th community of $\mathrm{Gr_b}$. And when $\mathrm{Gr_a}$ and $\mathrm{Gr_b}$ are identical, the value of NMI takes its maximum value of 1. Otherwise the NMI is close to 0.

On synthetic random networks, the community partition results of TGCDA and the other compared algorithms are shown in Figs. 3 and 4. And from these figures, we can obtain the following results.

(1) When the probability P_{in} grows gradually with an increment, obviously the Q-value and NMI accuracy of

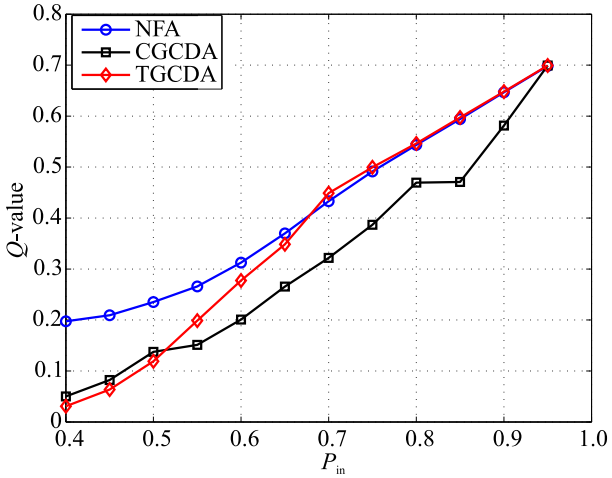


Fig. 3 Comparison of community partition results in terms of Q-value by different algorithms on synthetic random networks.

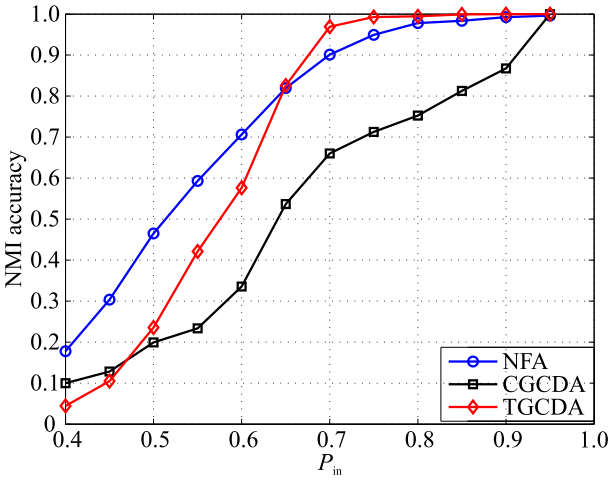


Fig. 4 Comparison of community partition results in terms of NMI accuracy by different algorithms on synthetic random networks.

the three algorithms also increase with different speeds and the increasing speed of TGCDA is higher than the other two algorithms.

(2) When the network has a clear community structure ($P_{\text{in}} \geqslant 0.7$), all considered algorithms get better Q-value and NMI accuracy, and our algorithm generally outperforms NFA and CGCDA. TGCDA can achieve NMI accuracy 1.13% higher than NFA averagely.

Therefore, the experimental results of synthetic random networks mean that our algorithm is feasible and suitable for the networks which own a clear community structure and it can gain more accurate community structure than the other compared algorithms.

4.2 Real world networks

In real world networks, the communities are formed by some certain relationship. Their topology structures are different from those synthetic random networks generated by computer. Here, we use the Zachary's karate club network[17], the bottlenose dolphin network[18], and American college football network[1] to test TGCDA. Table 2 displays their vertices and edges of the real world networks.

Zachary's karate club network is a widely used benchmark to test community detection algorithms. The network contains 34 vertices which represent the members in club and 78 edges which represent relationship between members. Due to the conflict of club's administrator and the club's instructor, the members split into two different groups. The bottlenose dolphin network represents the associations between 62 dolphins living in Doubtful Sound, New Zealand. Links between dolphins represent the statistically significant frequent associations. The network was divided into two groups because a crucial dolphin left. American college football network has 115 vertices which denote the football teams and 613 edges which denote games between two teams. Usually, 8 to 12 teams form a federation in network and each federation is a community.

We use these real world networks to test TGCDA and the other compared algorithms. The NMI accuracies of the three algorithms are illustrated in Table 3. From Table 3, we can see that TGCDA always obtains the largest NMI accuracy on each network. Meanwhile, our algorithm can achieve NMI accuracy 17.55% higher than NFA averagely. Clearly, TGCDA is superior to the other two compared algorithms. This means TGCDA has a better search ability for detecting a more accurate partition than the other two algorithms on the real world networks. Figure 5 plots the results of Q-value about the three algorithms and the true Q-value on three real world networks. Only on the bottlenose dolphin network, the Q-value of our algorithm is not the highest. Based on the true communities of three real world networks, we know that both communities' members on dolphins are not considerable and the others are adverse. From the above analysis, we can find that our algorithm is good at dealing with real world networks that their communities' members are considerable.

Table 2 Real world networks.

Networks	Number of vertices	Number of edges
Karate	34	78
Dolphins	62	159
Football	115	613

5 Conclusions

In this paper, we improve CGCDA to propose TGCDA. The algorithm introduces the tolerance relation among vertices and utilizes the thought of hierarchical granulation to realize graining of

Table 3 Comparison of NMI accuracy by different algorithms on real world networks.

Networks	TGCDA	NFA	CGCDA
Karate	**0.6995**	0.6925	0.6766
Dolphins	**0.8888**	0.5727	0.5057
Football	**0.7191**	0.6977	0.7121

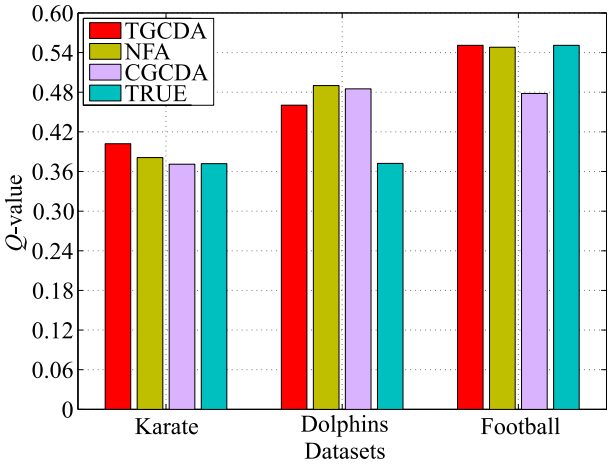


Fig. 5 Comparison of Q-value by different algorithms on real world networks.

networks. Relied on tolerance granulation of the network, the proposed algorithm can more accurately identify the community structure. TGCDA has tested and compared with other community detection algorithms. On real world networks, our algorithm can achieve NMI accuracy 17.55% higher than NFA averagely and on synthetic random networks, the NMI accuracy is also improved. Aiming at networks with a clear community structure, TGCDA can detect more accurate community structure than other algorithms. In the future, we will identify the hierarchical community structure based on multi-granularity.

Acknowledgements

This work is partially supported by the National High-Tech Research and Development (863) Program of China (No. 2015AA124102), the National Natural Science Foundation of China (Nos. 61402006 and 61175046), the Provincial Natural Science Research Program of Higher Education Institutions of Anhui Province (No. KJ2013A016), the Provincial Natural Science Foundation of Anhui Province (No. 1508085MF113), the College Students National Innovation & Entrepreneurship Training program of Anhui University (No. 201410357041), and the Recruitment Project of Anhui University for Academic and Technology Leader.

References

[1] M. Girvan and M. E. J. Newman, Community structure in social and biological networks, *Proceedings of the National Academy of Sciences*, vol. 99, no. 12, pp. 7821–7826, 2002.

[2] A. Lancichinetti, S. Fortunato, and F. Radicchi, Benchmark graphs for testing community detection algorithms, *Physical Review E*, vol. 78, no. 4, p. 046110, 2008.

[3] M. E. J. Newman, Detecting community structure in networks, *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 38, no. 2, pp. 321–330, 2004.

[4] A. Lancichinetti and S. Fortunato, Community detection algorithms: A comparative analysis, *Physical Review E*, vol. 80, no. 5, p. 056117, 2009.

[5] S. Fortunato, Community detection in graphs, *Physics Reports*, vol. 486, no. 3, pp. 75–174, 2010.

[6] W. Liu and L. Chen, Community detection in disease-gene network based on principal component analysis, *Tsinghua Science and Technology*, vol. 18, no. 5, pp. 454–461, 2013.

[7] F. L. Qian, Y. P Zhang, Z. Duan, and Y. Zhang, Community-based user domain model collaborative recommendation algorithm, *Tsinghua Science and Technology*, vol. 18, no. 4, pp. 353–359, 2013.

[8] L. I. Kuncheva, Using diversity in cluster ensembles, in *Systems, Man and Cybernetics, 2004 IEEE International Conference*, 2014, vol. 2, pp. 1214–1219.

[9] J. M. Kleinberg, Authoritative sources in a hyperlinked environment, *Journal of the ACM (JACM)*, vol. 46, no. 5, pp. 604–632, 1999.

[10] M. E. J. Newman, Fast algorithm for detecting community structure in networks, *Physical Review E*, vol. 69, no. 6, p. 066133, 2004.

[11] L. Zhang and B. Zhang, Fuzzy tolerance quotient spaces and fuzzy subsets, *Science China Information Sciences*, vol. 53, no. 4, pp. 704–714, 2010.

[12] S. Zhao, W. Ke, J. Chen, and Y. P. Zhang, Community detection algorithm based on clustering granulation, *Journal of Computer Applications*, vol. 34, no. 10, pp. 2812–2815, 2014.

[13] H. W. Shen, X. Q. Cheng, K. Cai, and M. B. Hu, Detect overlapping and hierarchical community structure in networks, *Physica A: Statistical Mechanics and Its Applications*, vol. 388, no. 8, pp. 1706–1712, 2009.

[14] T. Wu, Y. Guo, L. T. Chen, and Y. B. Liu, Fast overlapping and hierarchical community detection via local dynamic interaction, *arXiv preprint arXiv*, vol. 388, no. 8, pp. 1706–1712, 2009.

[15] N. Du, B. Wu, L. Xu, and B. Wang, A parallel algorithm for enumerating all maximal cliques in complex network, in *Data Mining Workshops, ICDM Workshops,* Hong Kong, China, 2006, pp. 320–324.

[16] M. Girvan and M. E. J. Newman, Finding and evaluating community structure in networks, *Physical Review E*, vol. 69, no. 2, p. 026113, 2004.

[17] 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.

[18] D. Lusseau, The emergent properties of a dolphin social network, *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 270, no. Suppl 2, pp. S186–S188, 2003.

Shu Zhao received her PhD degree in computer science from Anhui University in 2007. She is now an associate professor in the School of Computer Science and Technology, Anhui University. Her current research interests include quotient space theory, granular computing, and machine learning.

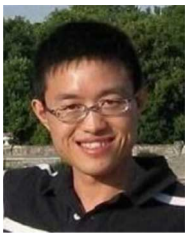
Wang Ke is a master student in the School of Computer Science and Technology, Anhui University. She received her bachelor degree from Tongling University in 2013. Her current research interests include granular computing and intelligent computing.

Jie Chen received her PhD degree in computer science from Anhui University in 2014. She is now a lecturer in the School of Computer Science and Technology, Anhui University. Her current research interests include quotient space theory and granular computing.

Feng Liu received his master degree from Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences. He is now a lecturer in the School of Computer Science and Technology, Anhui University. His current research interest includes intelligent computing.

Menghan Huang is an undergraduate in the School of Computer Science and Technology, Anhui University. His current research interest includes intelligent computing.

Yanping Zhang is a professor in the School of Computer Science and Technology, Anhui University. She received her PhD degree from Anhui University in 2005. Her current research interests include quotient space, granular computing, and intelligent computing.



Jie Tang received his PhD degree in computer science and technology from Tsinghua University in 2006. He is now an associate professor in Tsinghua University. His current research interests include social network mining, social influence analysis, and data mining.