A New Method for Detecting Anti-community Structures in Complex Networks

Qiang Yu¹, Ling Chen¹²
1. Institute of Information Science and Technology, Yangzhou University, Yangzhou, China
2. National Key Lab of Novel Software Tech, Nanjing University, Nanjing, China
1. E-mail: 77724954@163.com
2. E-mail: yzulchen@gmail.com

Abstract. In this paper, we present a label propagation algorithm named ACD for anti-community detection. Experimental results on some real world networks show that our algorithm can obtain higher quality results than other methods.

1. Introduction
In the past decade, identification of community structure has attracted much attention in various scientific fields. Many methods have been proposed and applied successfully to some specific complex networks. Those community structures studied in the past are also called "homophily" or "assortative mixing", where vertexes are divided into groups such that the members of each group are densely connected to other members of the same group [1,2]. Recently, another type of community is called "disassortative mixing" or "anti-community" has also been discussed to a less extent [3-5]. In an anti-community, vertexes have most of their connections outside their group and has no or fewer connection with the members within the same group.

Several methods for detecting the anti-community in the complex networks have been proposed recently. Newman et. al. [6] proposed a method based on the modularity matrix to detect both communities and anti-communities. While communities are detected by the eigenvectors corresponding to the largest eigenvalue of the modularity matrix, the anti-communities can be detected by the eigenvectors corresponding to the most negative eigenvalue. Newman et. al. [7] also proposed a method based on the mixed model to detect both communities and anti-communities. The method uses the machinery of probabilistic mixture models and the expectation maximization algorithm. It can detect communities, anti-communities and other types of structure in networks without prior
knowledge of what one is looking for. Wang Fan [8] proposed a method detecting the anti-communities using the most negative eigenvalue of the Laplacian matrix of the graph.

In this paper, a label propagation algorithm ACD (Anti-Community Detection) is proposed. We test the algorithm ACD on some real networks and compare the results with other methods [7,8]. Experimental results show that our algorithm can obtain higher quality results than other methods.

2. A label propagation method for detecting the anti communities

Recently, Raghavan et al. [9] introduce a very innovative and promising algorithm called label propagation algorithm (LPA) for community detection. The most striking feature of LPA is that it can find meaningful community structure at very high speed, which is faster than what is possible so far. In this paper, we improve the LPA to make it more suitable for anti-community detection in large scale networks. Note that we define anti-community structure in networks as groups of sparsely connected nodes, in our algorithm the label should be propagated to the non-connected nodes.

In our label propagation algorithm, there is a label for each node in a network and communities are defined as the sets of nodes of the same labels. Let \( G = (V, E) \) be a graph with \( n \) vertexes. Let \( v \in V \) be a vertex in graph \( G \), we define the set \( H(v) = \{ v' | v' \in V, (v, v') \in E \} \), which is the set of vertexes having connections with \( v \), and the set \( F(v) = \{ v' | v' \in V, (v, v') \notin E \} \), which is the set of vertexes having no connection with vertex \( v \). The algorithm tries to assign each vertex \( v \) a label \( L(v) \) which indicates the class \( v \) belongs to. Initially each all the vertexes are assigned different labels: \( L(v_i) = i \). The algorithm propagates the label \( L(v_i) \) of vertex \( v \) to the vertexes in \( F(v) \). The iterative relabeling process continues until every node has a label. In this fashion, sparsely connected group of nodes can reach a consensus on a unique label and form an anti-community quickly.

If two disconnected vertexes \( v_1 \) and \( v_2 \) are both connected to a third vertex \( v \), this means they probably can be assigned an identical label. In this case, we assign them \( \min \{ L(v_1), L(v_2) \} \) as the new label. Since two connected vertexes can not be grouped to the same anti-community, they should not be assigned the same label. Therefore, when a vertex \( v \) being assigned a new label \( L \), the algorithm checks the labels of all its neighbor vertexes in \( H(v) \) to see if there is any vertex in \( H(v) \) which has the same label \( L \). Vertex \( v \) can be assigned a new label \( L \), only when there is no vertex in \( H(v) \) labeled \( L \). Such label propagation process should continue until no node in the network changes its label.

Since label of each vertex is replaced only by a smaller one, and the label of each vertex is an integer less than \( n \), which is the number of the vertexes in the graph, it takes \( O(n) \) time for a label to converge to its minimal value. Therefore, the time complexity of the algorithm is \( O(n^2) \). Our ACD algorithm can obtain the anti-communities directly, unlike the bipartition-based method, which requires a recursive procedure of bipartitioning which requires a large amount of computation time. ACD algorithm does not require a predefined number of clusters \( k \). In addition, it can obtain a result where there is no connection between the vertexes within each cluster, and maximize the number of edges between clusters while the number of clusters is minimized.

3. Experimental results

In this section, we empirically demonstrate the effectiveness of our algorithm ACD for detecting the anti-communities in networks on some real world networks. All experiments have been run on Pentium IV, Windows XP, P1.7G, and visualize the results on Matlab 6.0.
We apply our method to the real network shown in Figure 1. The network of the most frequent 50 English adjectives and 50 nouns selected from the novel “The Old Man and the Sea” with edges connecting any pair of words that appear adjacent to each other at any point of the text. Because adjectives typically occur next to nouns in English and very few pairs of nouns or adjectives appear in the text, most edges connect an adjective to a noun. The network is thus approximately bipartite or disassortative. In Figure 1 the red nodes represent the adjectives and the blue nodes represent the nouns. In the graph, most edges are observed to run between two nodes of different colors.

We first test on this network using the spectral based algorithm[8], the result is shown in Figure 2. From the figure, we can see that the network has been divided into two anti-communities, of which one consists of most of the red nodes and the other consists of most of the blue ones. But there are 6 blue nodes are assigned into the red anti-community, and 8 red nodes are assigned into blue anti-community. Therefore, the precision of the partitioning is 86%.

We also test our algorithm ACD on this network, the result is shown in Figure 3. From the figure, we can see that the network also has been divided into two anti-communities. There are 3 blue nodes are assigned into the red anti-community, and 5 red nodes are assigned into blue anti-community. Therefore, the precision of the partitioning is 92%. Therefore, ACD algorithm can obtain higher quality results than the spectral based algorithm.

To measure the quality of the anti-community partition, we use the Newman-Girvan modularity for community partition [10]. Since we detect the anti-community, smaller value of such modularity indicates higher quality of the partition result.

We test the on the networks of adjective-noun, Southern women [11], and Scotland Corporate Interlock [12] using the mixture models based algorithm [7], the spectral based algorithm [8], and algorithm ACD. The Southern women data set was collected by Davis et al. [11] in and around Natchez, Mississippi during the 1930s as part of an extensive study of class and race in the Deep South. Scotland Corporate Interlock [12] is a data set on corporate interlocks in Scotland in the early twentieth century. Modularity of the results by the algorithms are as shown in Table 1. From the table, we can see that the result of ACD has lower modularity than other methods. This indicates that algorithm ACD can get better anti-community partition than other methods.

4. Conclusions
A label propagation based algorithm ACD for anti-community detection is proposed. Our ACD algorithm can obtain the anti-communities directly in O(n^2) time. The ACD algorithm does not require a predefined number of clusters k, and can obtain a result where there is no connection between the vertexes within each cluster, and maximize the number of edges between clusters while the number of clusters is minimized. Experimental results show that our algorithm can obtain higher quality results than other methods.

Acknowledgements
This research was supported in part by the Chinese National Natural Science Foundation under grant Nos. 61070047, 61070133 and 61003180. State Key Fundamentals Research (973) Project under contract 2012CB316003, Natural Science Foundation of Jiangsu Province under contracts BK2010318, BK21010134.
References

[1] M. Girvan and M. E. J. Newman, “Community structure in social and biological networks,” Proc. Natl. Acad. Sci. U.S.A., vol. 99, no. 12, pp. 7821-7826, 2002.
[2] Danon L, Duch J, Diaz-Guilera A, Arenas A (2005) J Stat Mech, P09008.
[3] M. E. J. Newman, "Finding community structure in networks using the eigenvectors of matrices,” Phys. Rev. E, vol. 74, no. 3, pp. 036104, 2006.
[4] P. Holme, F. Liljeros, C. R. Edling, and B. J. Kim, Net-work bipartivity. Phys. Rev. E 68, 056107 (2003).
[5] Estrada E, Rodri´guez-Vela´zquez JA (2005) Phys Rev E 72:046105.
[6] M.E.J. Newman, Finding community structure in networks using the eigenvectors of matrices, Proc Natl Acad Sci USA, 2007, 104(23) : 9564-9569
[7] Newman M E, Leicht E A. Mixture Models and Exploratory Analysis in Networks. Proc Natl Acad Sci USA, 2007, 104(23) : 9564 – 9569
[8] Wang Fan, Detecting “Anti-communities” of Networks Based on Spectral Method, Thesis for the Master Degree, Huazhong University of Science and Technology Wuhan, Hubei 430074, China, June, 2008
[9] U. N. Raghavan, R. Albert, and S. Kumara, “Near linear time algorithm to detect community structures in large-scale networks,” Phys. Rev. E, vol. 76, no. 3, 036106, Sep. 2007, doi: 10.1103/PhysRevE.76.036106.
[10] M. E. J. Newman and M. Girvan, Phys. Rev. E 69, art. no. 026113 (2004).
[11] A. Davis, B. B. Gardner, and M. R. Gardner, Deep South, (University of Chicago Press, 1941).
[12] J. Scott and M. Hughes, The anatomy of Scottish capital: Scottish companies and Scottish capital, 1900–1979, (Croom Helm, London, 1980).

Table 1 Comparison of the modularity of the results by the algorithms

| Data set               | Algorithm        | ACD    | spectral based | mixture models based |
|------------------------|------------------|--------|----------------|----------------------|
| adjective-noun         | ACD              | -0.3448| -0.2988        | -0.3199              |
| Southern women         | ACD              | -0.4778| -0.4386        | -0.4561              |
| Scotland Corporate Interlock | ACD              | -0.4113| -0.3569        | -0.3815              |