Information Sciences 275 (2014) 293–313

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Anti-modularity and anti-community detecting in complex networks

Ling Chen [a,b,*], Qiang Yu [a], Bolun Chen [a]

[a] Institute of Information Science and Technology, Yangzhou University, Yangzhou, China
[b] National Key Lab of Novel Software Technology, Nanjing University, Nanjing, China

ARTICLE INFO

Article history:
Received 29 August 2012
Received in revised form 15 November 2013
Accepted 7 February 2014
Available online 19 February 2014

Keywords:
Anti-community structure
Anti-modularity
Network
Graph
Bipartite graph

ABSTRACT

Many networks of interest in sciences and social research can be divided naturally into anti-communities. The problem of detecting and characterizing such anti-community structure has attracted recent attention. In this paper, we first define the anti-modularity as a quantitative measure of anti-community partitioning on a network. We also theoretically and empirically show the reliability of anti-modularity as a measurement of the quality of an anti-community partitioning. A label propagation algorithm LPAD for anti-community detection is proposed. Experimental results on synthetic and real world networks show that our algorithm LPAD can obtain higher quality anti-community partitioning than other methods.

1. Introduction

The study of complex networks has become an important area of multidisciplinary research involving physics, mathematics, biology, social science, informatics and other theoretical and applied science. Many complex systems in the real world can be modeled as networks including social networks [4,32], information networks [7,8,14] and biological networks [17,19]. The topology structure of a network influences the behaviors of individuals in the network. For instance, topology of social networks affects the spread of information and disease, and topology of a power grid affects the robustness and stability of power transmission. These complex networks can be divided naturally into communities. Communities are of interest because they often correspond to functional units and reflect the inhomogeneity or locality of the topological relationships between the elements of the systems. Networks can have properties at the community level that are quite different from the properties from the whole network itself. For instance, in some social networks, individuals with different average numbers of contacts belong to different groups. The individuals in one group might be gregarious, having many contacts with others, while the individuals in another group might be more reticent. An example of this behavior can be seen in networks of sexual contacts, where separate communities of high- and low-activity individuals have been observed [2,16].

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 [3,6,11,24,26–28,30,31,36,38,39,41–43,45,46,50,51]. For survey and comparison, the reader can refer to [9,15,25,37]. These methods identify the community structure by finding an optimal partition of the network according to a certain criterion or definition of community.

* Corresponding author at: Institute of Information Science and Technology, Yangzhou University, Yangzhou, China. Tel.: +86 514 87870026; fax: +86 514 87887937.

E-mail addresses: yzulchen@gmail.com (L. Chen), 77724954@163.com (Q. Yu), chenbolun1986@163.com (B. Chen).

http://dx.doi.org/10.1016/j.ins.2014.02.040
0020-0255/

Those community structures studied in the past are also called ''homophily'' or ''assortative mixing'', where vertices are divided into groups such that the members of each group are mostly connected to other members of the same group [10,17]. Recently, another type of community is called ''disassortative mixing'' or ''anti-community'' has also been discussed to a less extent [13,20,31]. In an anti-community, vertices have most of their connections outside their group and have no or fewer connections with the members within the same group. A network with 3 anti-communities is shown in Fig. 1.

In a special case where a network consists of only two such anti-communities, the problem is to detect the largest bipartite subgraph in a given graph. Bipartite or approximately bipartite graphs have attracted some attention in the recent literature. In such approximately bipartite graphs or ''two-mode'' networks, two disjointed sets of nodes are related by links representing the relationship between the elements of both classes [20,22,29,44,49]. There are numerous natural systems that can be modeled as approximately bipartite networks. For instance, Kleinberg [23] has suggested that small bipartite subgraphs in the web graph may be a signature of so-called hub/authority structure within web communities. Another example of such approximately bipartite graphs is the network of English words [31] shown in Fig. 2. In this network, the vertices represent 112 commonly occurring adjectives and nouns in a particular body of text (from the novel *David Copperfield* by Charles Dickens), with edges connecting any pair of words that appear adjacent to each other at any point in the text. Because adjectives typically occur next to nouns in English, most edges connect an adjective to a noun and the network is thus approximately a bipartite or disassortative one. In Fig. 2, the two shaded groups contain adjectives and nouns respectively and the shades of the individual vertices represent the anti-communities. This approximately bipartite networks can also represent authors that cite (or are cited by) papers, people that belong to institutions, cities that have certain services, or voting results of delegates concerning certain proposals.

Holme et al. [20] pointed out several potential applications of bipartivity detecting on a graph, such as network studies of sexually transmitted diseases, trade networks of buyers and sellers, ''genealogical'' networks of disease outbreak and food webs.

Several methods for detecting the anti-community in the complex networks have been proposed recently. Newman [31] 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 and Leicht [34] 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 any prior knowledge. Wang [48] proposed a method detecting the anti-communities using the most negative eigenvalue of the Laplacian matrix of the graph.

It is obvious that the problem of detecting the maximum bipartite subgraph is similar but not equivalent to the conventional problem of searching for the maximum cut in the graph. Several approximation algorithms for max cut problem are reported. Trevisan [47] presented an approximation algorithm for detecting max-cut. The algorithm requires $O(n^2)$ time where n is the number of vertices, and achieves an approximation ratio of 0.531. On instances in which an optimal solution cuts a $1-\varepsilon$ fraction of edges, the algorithm can finds a solution that cuts a $1-4\sqrt{\varepsilon}+8\varepsilon-O(1)$ fraction of edges. Alon and Sudakov [1] proved that the smallest eigenvalue μ of any non-bipartite graph on n vertices with diameter D and maximum degree Δ satisfies $\mu \geqslant -\Delta + \frac{1}{(D+1)n}$. This improves previous estimates and tights up to a constant factor. They determinate the maximum cut algorithm of Goemans and Williamson [18] for graphs $G=(V,E)$ where the size of the maximum cut is at least $A|E|$, for all A values between 0.845 and 1. To measure the precision of the results of anti-community detecting, it is important to give a quantitative measure of bipartitivity of a given graph. Holme et al. [20], Estrada and Rodriguez-Velazquez [13] have independently devised measures of bipartitivity and used them to analyze a variety of real world networks. Estrada and Rodriguez-Velazquez [13] provided a quantitative measure of network bipartivity as a proportion of total number of closed walks in the network. All those measure are for evaluating the bipartivity of a network. However, the problem of

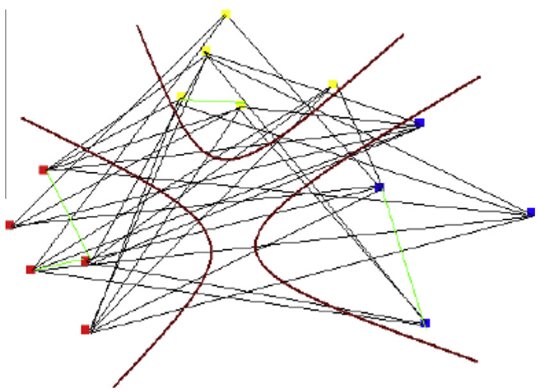

Fig. 1. A network with 3 anti-communities.