Community Identification in Directed Networks

Youngdo Kim, Seung-Woo Son, and Hawoong Jeong

Department of Physics, Korea Advanced Institute of Science and Technology,
373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Korea
kyd1225@kaist.ac.kr

Abstract. The most common approach to community identification of directed networks has been to ignore edge directions and apply methods developed for undirected networks. Recently, Leicht and Newman published a work on community identification of directed networks, which is a generalization of the widely used community finding technique of modularity maximization in undirected networks. However, our investigation of this method shows that the method they used does not exploit direction information as they proposed. In this work, we propose an alternative method which exploits the directional information of links properly.

Keywords: community identification method, directed networks, directional information.

1 Introduction

Community structure has been considered as an important property of complex networks, because community structure is a ubiquitous property of many real-world networks, which cannot be found in randomly generated networks. And some researches suggest that the community structure is relevant to other structural properties, such as clustering coefficient and hierarchical structure, and the dynamics in the networks. Hence, considerable attentions have been given to identify the communities of networks. Community identification problem can be considered as finding the best partitioning configuration in which nodes in the same community are more likely to interact with each other and nodes of different communities are more likely not interacting. To find this best partitioning configuration, the most common approach is maximizing a benefit function called modularity, proposed by Newman and Girvan [1]. Although a lot of optimization methods have been proposed by many researchers to maximize modularity efficiently, these methods can only be applied to undirected networks, because the modularity is only defined for undirected networks.

Despite a lot of real-world networks have directed links, in the past the common approach to community finding in directed networks was just to ignore link directions and apply those well developed methods used in undirected networks. This approach may discard useful information contained in link directions, which could be helpful or critical in detecting communities. Several recent studies [2,3,4]

J. Zhou (Ed.): Complex 2009, Part II, LNICST 5, pp. 2050–2053, 2009.

used different methods to exploit direction information, but their definitions of community do not consistent with the one in undirected networks, which is defined by the modularity.

In recent studies, Arenas *et al.* [5] proposed a generalized modularity in directed networks, and Leicht and Newman [6] utilized this generalized modularity to detect communities in direction networks, exploiting the direction information as nodes linked by unexpected direction being more likely to be divided into the same community. Unfortunately, the generalized modularity does not exploit the direction information as they expected. It will be discussed in detail at other work. In this work, we report an alternative method which exploits direction information as Leicht and Newman proposed, which is nodes linked by unexpected direction being more likely to be divided into the same community.

2 Method

Leicht and Newman [6] used the generalized modularity proposed by Arenas *et al.* [5] to find communities in directed networks. Their idea of exploiting direction information can be described as follows. Consider a pair of nodes A and B. Vertex A has high out-degree and low in-degree while B has the reverse, then a directed link connecting A and B is more likely to point from A to B than the opposite direction. Hence, if a directed link running from B to A is found in a network which we observe, it is a bigger surprise than a link from A to B and will make more contributions to the modularity, since modularity should be high for statistically surprising configurations. However, the generalized modularity does not exploit directional information as described above.

We propose a new method as follows which can exploit direction information as Leicht and Newman described. Following the discussion above, as a directed link running from B to A is a bigger surprise which is indicating A and B are more likely to be in the same community, one can consider this situation as a directed link running from B to A means stronger *relatedness* between A and B than a link running from A to B. In the introduction part, we have mentioned that the common approaches to community finding in directed networks was just to ignore link directions and transform the directed networks into undirected networks and apply those well developed methods used in undirected networks. Our idea is that, instead of simply ignoring directional information, the effect of direction can be added to the link of new undirected network, by giving higher weight to the more *surprising* link when we consider link direction.

Let's consider a link between node i and node j. The probability of this link directing from j to i, when the links are assigned randomly while keeping the degree of each node, is

$$p_{ij} = \frac{k_j^{out} k_i^{in}/2m}{k_j^{out} k_i^{in}/2m + k_i^{out} k_j^{in}/2m}, \tag{1}$$

where $k_i^{in} = \sum_j A_{ij}$ and $k_j^{out} = \sum_i A_{ij}$ are respectively the incoming and outgoing degree of node i and node j, where A_{ij} represents the link running from j to i, and the total number of links is defined as $m = \sum_i k_i^{in} = \sum_j k_j^{out} = \sum_i \sum_j A_{ij}$.

Since small p_{ij} indicates stronger *relatedness* for the direction from j to i, we can define the *relatedness* of node i and j, which is also the weight of the link between node i and node j in the transformed undirected network, as

$$w_{ij} = A_{ij}(1 - p_{ij}) + A_{ji}(1 - p_{ij}). \tag{2}$$

Thus $\{w_{ij}\}$ is a undirected network transformed from the original directed network $\{A_{ij}\}$, containing the directional information which was just ignored in other methods of the past. Then, by applying the well developed methods in undirected networks to the newly generated weighted undirected network $\{w_{ij}\}$, it is able to detect the community of original directed network without losing directional information.

3 Result

We applied our method to the first example of Leicht and Newman [6] as shown in Fig. 1, which is a computer-generated network of 32 nodes, generated by the following process. Initially every nodes are linked with each other randomly with probability p by undirected links. Then the nodes are divided into two groups of 16 nodes each and links connecting nodes of same group are assigned random direction but links connecting nodes of different groups are assigned direction with certain bias so that the direction is more likely running from nodes in group 1 to nodes in group 2 than vice versa. As there was no community structure before assigning direction for each link, no community structure can be identified by ignoring directions.

However, when our method is applied to the directed network, a significant community structure which is even recognizable by human eyes rises from the

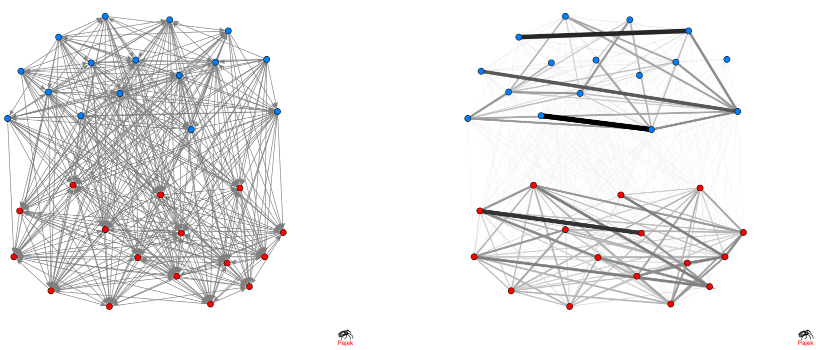

Fig. 1. Computer generated two-community random network described in the text. If we ignore the directional information, the community structure cannot be found since there is no difference in link density except link directions. On the right: Applying our algorithm, link weights are newly assigned, which is represented by the link width in this figure, and then a pre-existing algorithm applied to undirected networks. We can find two clear communities as shown on the right panel.

transformed undirected network(the right figure in Fig. 1). Optimizing modularity Q in this undirected network, also divides the network into red-node-community and blue-node-community, which perfectly matches with the network generating mechanism.

4 Summary and Remarks

We proposed a community finding method for directed networks facilitating directional information as Leicht and Newman firstly proposed. A computer generated network, in which community structure cannot be detected without direction information, was tested to confirm our method is working well, and the result shows that our method perfectly detects two communities which were designed in the generating process of the network.

We are currently applying our method to real-world examples of directed networks, in which more meaningful community structures could be detected. Furthermore, we are investigating a method which can directly detect communities in directed networks, without transforming them into weighted networks as we did in this report. This work was supported by KOSEF through the grant No. R17-2007-073-01001-0.

References

1. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. Phys. Rev. E 69, 026113 (2004)
2. Newman, M.E.J., Leicht, E.A.: Mixture models and exploratory data analysis in networks. Proc. Natl. Acad. Sci. U.S.A. 104, 9564–9569 (2007)
3. Rosvall, M., Bergstrom, C.T.: Maps of random walks on complex networks reveal community structure. Proc. Natl. Acad. Sci. U.S.A. 105, 1118–1123 (2007)
4. Guimerà, M., Sales-Pardo, M., Amaral, L.A.N.: Module identification in bipartite and directed networks. Phys. Rev. E 76, 036102 (2007)
5. Arenas, A., Duch, J., Fernández, A., Goméz, S.: Size reduction of complex networks preserving modularity. New J. Phys. 9, 176 (2007)
6. Leicht, E.A., Newman, M.E.J.: Community structure in directed networks. Phys. Rev. Lett. 100, 118703 (2008)