A New Approach to Identify Influential Spreaders in Complex Networks*

Qingcheng Hu**, Yang Gao, Pengfei Ma, Yanshen Yin, Yong Zhang, and Chunxiao Xing

Research Institute of Information Technology, Tsinghua National Laboratory for Information Science and Technology, Department of Computer Science and Technology, Tsinghua University, Beijing, China
{hqc10,yang-gao10,mpf11,yys12}@mails.tsinghua.edu.cn,
{zhangyong05,xingcx}@tsinghua.edu.cn

Abstract. In the research of the propagation model of complex network, it is of theoretical and practical significance to detect the most influential spreaders. Global metrics such as degree centrality, closeness centrality, betweenness centrality and K-shell centrality can be used to identify the influential spreaders. These approaches are simple but have low accuracy. We propose K-shell and Community centrality (KSC) model. This model considers not only the internal properties of nodes but also the external properties of nodes, such as the community which these nodes belong to. The Susceptible-Infected-Recovered (SIR) model is used to evaluate the performance of KSC model. The experiment result shows that our method is better to identify the most influential nodes. This paper comes up with a new idea and method for the study in this field.

Keywords: Complex networks, Centrality measures, Influential node, SIR model.

1 Introduction

Many systems perform like complex networks, such as the Internet, social net-work, computing networks, biological network and social system. There are many researches on the topology and functionality of the complex networks. It is valuable to identify the most influential nodes [1-2]. This will help to control the disease transmission, rumors spreading [3], computer virus spreading and popularize new ideas and new products.

In this paper, we propose the KSC (K-shell and Community centrality) model. This model considers not only the internal properties of nodes but also the external properties of nodes. The internal properties mean the classic centralities such as degree, closeness, betweenness and so forth. The external properties mean the properties based on community, such as the size and closeness of the community which the node

* Supported by National Basic Research Program of China (973 Program) No.2011CB302302; Tsinghua University Initiative Scientific Research Program.
** Corresponding author.

J. Wang et al. (Eds.): WAIM 2013, LNCS 7923, pp. 99–104, 2013.

is belong to. Then we use the Susceptible-Infected-Recovered (SIR) model to evaluate the performance of our model. The experiment results shows that our method is much better to identify the most influential nodes.

2 Related Work

Identifying influential spreaders has remarkable practical value in complex networks. It can help people control the disease transmission, rumors spreading, computer virus spreading and popularize new ideas and new products.

2.1 K-Shell Decomposition Method

K-shell decomposition [4] is a well-established method for analyzing the structure of large-scale graphs, denoted by $C_{ks}(v)$. A step of decomposition is performed by repeatedly deleting all nodes with degree less than k. The K-shell value of the nodes removed in this step is k. The whole decomposition is finished when all nodes are removed. K increases from 1.

2.2 Community Detecting Algorithms

Network community structure [5] is one of the most fundamental and important topological properties of complex networks, within which information spreading is faster than between which they are quite sparse. In this paper, we implement FN algorithm [6] to detect communities.

3 KSC Centrality Model

We think that a node's influence is determined not only by the node's internal properties but also by the node's external properties. This is consistent with the philosophical internal cause and external cause. KSC is a novel idea which combines the internal properties with the external properties.

Given a complex network G=(V, E), the KSC value of node v_o is denoted by:

$$KSC(v_o) = \alpha f_{internal}(v_o) + \beta f_{external}(v_o)\text{, } \alpha+\beta=1 \tag{1}$$

$f_{internal}(v_o)$ represents the node's internal influence while $f_{external}(v_o)$ represents the node's external influence, α is the internal factor while β is the external factor, which satisfies $\alpha+\beta=1$, α and β are determined by the actual topology and functionality of the network.

$f_{internal}(v_o)$ is denoted by:

$$f_{internal}(v_o) = \frac{K(v_o)}{\max_{v \in V}(K(v))} \tag{2}$$

K(ν) is the internal property, which can be valued with degree, betweenness, closeness and K-shell. $\max_{v\in V} K(v)$ is the normalized factor.

$f_{external}(v_o)$ is denoted by:

$$f_{external}(v_o) = \frac{\sum_{c\in C} d(v_o, c)\,|c|}{\max_{v\in V}(\sum_{c\in C} d(v,c)\,|c|)} \tag{3}$$

C is the collection composed by the communities calculated by FN algorithm, $d(v_o, c)$ is the number of v_o's neighbor in community c, |c| is the size of community c, $\max_{v\in V}(\sum_{c\in C} d(v, c)|c|)$ is the normalized factor. $f_{external}(v_o)$ increases by the number of neighbors which lies in different communities. It indicates v_o's ability to propagate messages to various communities which is related to the influence of v_o.

α and β are set to different values according to different networks' topology which won't be discussed here. To simplify the experiment while ensuring high performance, in this paper, the experiments use the following configurations:

$$\alpha = \beta = 0.5,\ K(v) = C_{ks}(v) \tag{4}$$

4 Experiments

4.1 SIR Model

In social networks, SIR model has been widely used in the research of disease, information and rumors spreading. In order to verify our proposed model, we use SIR model [7] to simulate the propagation process, and compare the result with ours.

The SIR model is dynamic in three senses: susceptible, infectious and recovered. When an individual is in infectious sense, it will infect neighbor individuals in susceptible sense by the probability of β. The infected individuals will recover by the probability of γ.

4.2 Dataset

Considering the different social network types representing the different properties of the network topology, we select four real networks dataset for analyzing. (i) Blogs [8]. (ii) Netscience [9]. (iii) Router [10]. (iv) Email [11]. Our model can also be applied to other types of complex networks.

4.3 Performance of the Experiment

We implement the SIR propagation model to evaluate the actual influence of nodes. Only one node is chosen as the initial propagation node in each simulation. The propagation time t is set to 10. The influence F(t) is denoted by the number of nodes in state I and R. We take the average F(t) of 1000 simulations as the final F(t).

The above F(t) is denoted by the actual influence of a node, which is used as a reference when evaluating KSC and other four classic centralities.

Fig. 1. indicates the relation between centrality and actual influence in different networks. In Blogs, the relations between F(t) and the degree, K-shell and betweenness are very weak. For example, in the figure of betweenness, some nodes with higher betweenness are less influential than some others with lower betweenness. However, the KSC centrality is able to distinguish the most influential nodes. In Email, all centralities achieve high performance, while the closeness and KSC are better. This is determined by the topology of the network. In Netscience, betweenness has the worst performance. Closeness and K-shell has little relation with F(t). KSC performs better than degree. In Router, K-shell and KSC achieve similar results. Other centralities are hard to distinguish the influential nodes.

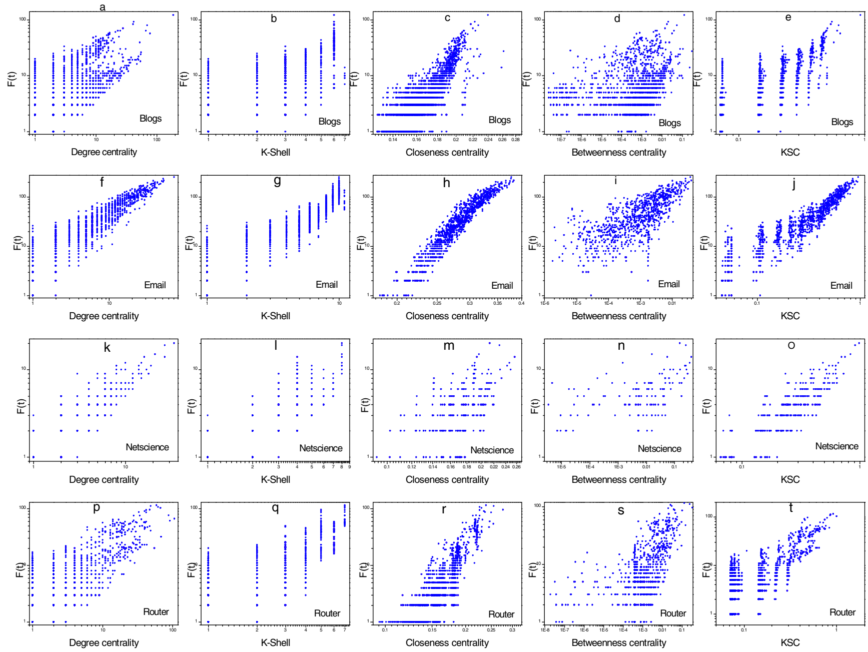

Fig. 1. Comparisons of nodes' influence

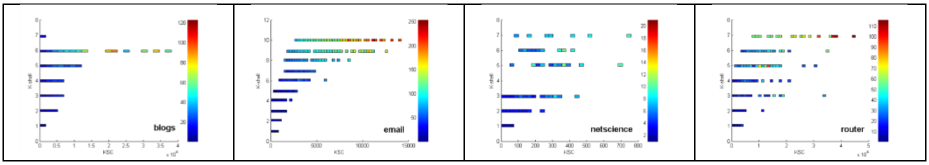

Fig. 2. The comparisons between KSC and K-shell in different complex networks. The horizontal axis represents KSC centrality while the vertical axis represents K-shell centrality. The color at point (x, y) represent the F(t) of the node with KSC=x and K-shell=y.

In one word, the classic centralities have their advantages and disadvantages. KSC is obviously the best in general. KSC is able to distinguish the most influential nodes and is suitable for more complex networks.

[12] points out that in case of single propagation source, the most influential nodes are not the nodes with the highest degree or betweenness, but the nodes with highest K-shell. KSC is based on K-shell, but we take the external properties into consideration.

Fig. 2 shows the comparison between KSC and K-shell. In Email, given K-shell=10, the actual influence F(t) is not constant and has an increasing trend with the increase of KSC. In the other side, when we fix KSC, F(t) is relatively stable, which is indicated by the small range of color change. As we can see from this point, the external properties are important factors to evaluate the most influential nodes.

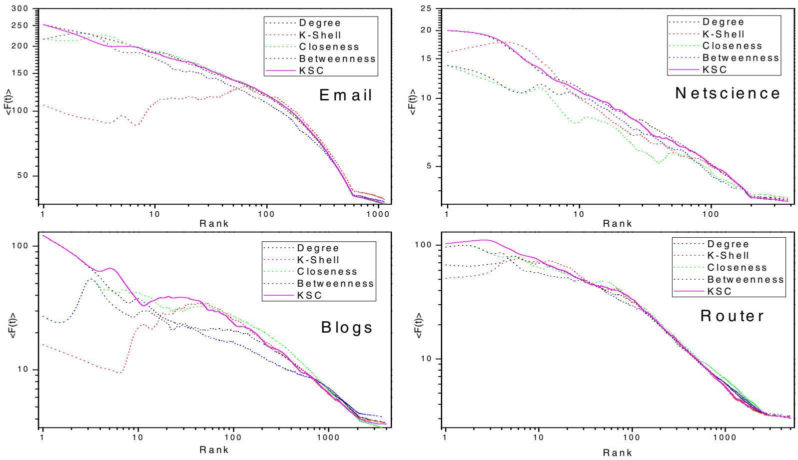

Fig. 3. The average number of F(t) (t=10) of the top-k nodes as ranked by the five centrality models. The vertical axis denotes the influence F(t). The horizontal axis denotes the different models rank. The curves of different colors denote different models.

Fig. 3 shows the average number of F(t) (t=10) of the top-k nodes as ranked by the five centrality models. For example, in Blogs of Fig. 3, (x, y) is a point (node) on the curve of KSC. The horizontal value of the point represents its influence rank (i.e. x) in KSC model. The vertical value of the point represents its influence (i.e. y) in SIR model. Theoretically, an ideal model gives a monotonically decreasing curve in the figure. If the actual influence F(t) of a node is lower, the influence rank of the node in this model will be lower too.

From the results, we can find that the four classic models fluctuate in different networks, particularly sharply in the top-10 results. In Blogs, the four models are all not accurate. The lower rank nodes are more influential than the higher rank nodes. But the KSC model we proposed almost meets the theoretical curve and the real situation.

The experiments adequately demonstrate that the influence of the nodes is not only determined by internal properties, but also closely related to external properties.

5 Conclusion

Identifying the most influential spreaders in complex networks can help us improve the propagation efficiency of new ideas and products, and develop appropriate strategies to prevent the spread of diseases and rumors. We propose the KSC centrality model. This model is used to analyze the complex networks by considering not only the internal properties of nodes but also the external properties of nodes. We chose four common complex networks, Blogs, Email, Router and Netscience as the datasets of our experiment. In the experiment, we calculate and rank the influence of all nodes in those networks by different models. After that, we use SIR model to simulate the propagation process and compare the result of different models. The experiment results show that our model is more accurate and applicable to identify the most influential spreaders than four classic models.

It is a hot but difficult research to identify the most influential spreaders in complex networks. This paper provides a new idea and method for this challenging work. We hope it can spark the future studies.

References

1. Kempe, D., Kleinberg, J.M., Tardos, É.: Maximizing the spread of influence through a social network. In: KDD 2003, pp. 137–146 (2003)
2. Weng, J., Lim, E.-P., Jiang, J., He, Q.: Twitterrank: finding topic-sensitive influential twitterers. In: WSDM, pp. 261–270 (2010)
3. Budak, C., Agrawal, D., El Abbadi, A.: Limiting the spread of misinformation in social networks. In: WWW, pp. 665–674 (2011)
4. Carmi, S., Havlin, S., Kirkpatrick, S., Shavitt, Y., Shir, E.: A model of Internet topology using k-shell decomposition. Proc. Natl. Acad. Sci. USA 104, 11150–11154 (2007)
5. Newman, M.E.J.: Detecting community structure in networks. European Physical Journal (B) 38(2), 321–330 (2004)
6. Newman, M.E.J.: Fast algorithm for detecting community structure in networks. Physical Review E 69(6), 066133 (2004)
7. Diekmann, O., Heesterbeek, J.A.P.: Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation. Wiley Series in Mathematical & Computational Biology, New York (2001)
8. Xie, N.: Social network analysis of blogs, M.Sc. Dissertation, University of Bristol (2006)
9. Newman, M.E.J.: Finding community structure in networks using the eigenvectors of matrices. Phys. Rev. E 74, 36104 (2006)
10. Spring, N., Mahajan, R., Wetherall, D., Anderson, T.: Measuring ISP topologies with rocketfuel. ACM SIGCOMM Comp. Comm. Rev. 32(4), 3–145 (2004)
11. Guimerà, R., Danon, L., Díaz-Guilera, A., Giralt, F., Arenas, A.: Self-similar community structure in a network of human interactions. Phys. Rev. E 68, 065103 (2003)
12. Kitsak, M., Gallos, L.K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H.E., Makse, H.A.: Identifying influential spreaders in complex networks. arXiv, Tech. Rep., physics and Society, 1001.5285 (2010)