ITM Web of Conferences 7, 01008 (2016) DOI: 10.1051/itmconf/20160701008
ITA 2016

A Method about the Propagation of Faults in Network

Yao-Nan LI[1,a] and Shi-Hai WANG[1,b]

[1] *Dept. Reliability and System Science, Beihang University, Beijing, China.*
Corresponding author: [a] q2511757@126.com, [b] wangshihai@buaa.edu.cn

Abstract. The study of network has been developed for several years. The research usually focuses on the attributes of the network, including the heterogeneity [1], the topology [2] [3]and the communities[4][5] of the network. About the propagation dynamics, reference [6][7] laid emphasis on the mathematical analysis of the propagation of networks. Although these researches are important, with the developing of computer, how to simulate the propagation weighs more and more. Abstraction of ways to propagate is limited to virus model. There are two virus propagation models. One is about bio-virus model. Although these 2 respects have been developed very well and there are lots of different attributes within, the faults propagation varies a lot from these two. For example, other forms of network like SOA and supply chain vary a lot in ways to propagate. This paper focuses on another useful way to simulate the propagation of faults within scale-free directed networks. Based on the fundamental analysis of propagation and the different situations of simulating complexity and amount of information, this paper gives different methods to simulate the propagation of faults in the whole network. This paper invents an ingenious way with pseudo-nodes to deal with nodes of small in-degree. This methods balance the performance of the simulation with the complexity of the simulation. Then this paper simply analyzes an example with two different methods to test their validation. Of course, there are several questions remaining in this method, like how to give the probability of propagation based on different exclusiveness. These are shown at the end of this paper.

1 Introduction

Not far from today, random graph theory dominated the research of complex network [8]. After reconnection of a few edges in regular network, Watts and Strogatz found the small world model, in which The Clustering Coefficient is relatively high, at the same time, the average path length is short.

After the attributes of small world in real world, Barabasi and Albert did deep research in World Wide Web [10]. Barabasi and Albert surprisingly found that, the distribution of World Wide Web was not like what random graph theory had forecasted, Poisson distribution. Instead, it conform the power law distribution, A.K.A. $P(k) = C \cdot k^{-\gamma}$ (C and γ is a constant number). From there on, the complex network has been widely studied.

Attributes of the network, including the heterogeneity [1], the topology [2][3]and the communities[4][5] of the network have been fully researched. About the propagation dynamics, reference [6][7] based on the mathematical analysis of the propagation of networks have the attention of lots of peers. Although these researches are important, with the developing of computer, simulating the propagation becomes very important, too. Also, the notion of propagation of virus developed, including not only the bio-virus but the computer virus.

Apparently there are more than 1 methods of propagation of failure, fault and virus in network. Actually, according to the different attributes of the network and the mechanism of propagation with virus, the methods should vary. Reference [11][9]give some hypotheses about the propagation mechanism including that "if there are more than one neighbors, the virus and fault can only choose one of them to propagate with probability" and "an infected node can't propagate its neighbors simultaneously " These are something apparently can be adapted to other kinds of networks. Noting that the way of infection between bio-virus and computer virus is different. The infection of bio-virus depends on the contact and the moving of the population. However, the computer virus propagates through the network directly. The bio-virus can lead to the death of the host, nevertheless the computer network will not loss a node in the network because of computer virus. The hosts of bio-virus vary a little. The key nodes like the server in the computer network weigh more than the PCs in the network. After the infection, the antibody plays a very important role in the next phases. Although the computer network can also be repaired, the whole process is of passiveness instead of activeness.

Building on the knowledge of virus propagation and the differences between propagating ways of bio-virus and computer virus, the propagation of fault is something special that we are interested. There are some the fundamental attributes within the faults propagation. To simplify the model, there are 2 aspects we are concerning here. One is the network where the faults may propagate is static. The other is the calling frequency is relevant to the node itself.

With the developing of the computer, the propagation simulation of faults can be done by the computer based on the mechanism of propagation and different amount of information people already have. According to aforementioned two points, studying the mechanism and the method of propagation in detail is not only prospective but necessary. This paper analyzes the fundamental things about propagation of failure in scale-free directed network and contributes an ingenious way to simulate the propagation under different amount of information.

2 Fundamental Analysis of Faults Propagation

In [9]the hypothesis that the virus can only choose one of those to propagate with probability can be easily overridden here. Suppose that node C is upper software, A, B is called at the same time by C. Similarly, node D calls node A and B when D is running (as shown in Figure 1). If node A failed, and if C and D run at the same time, node C and D definitely will fail. A naturally idea is come up with, which is to separate this graph into 2 simpler ones ideologically (as shown in Figure 2 Figure 3).

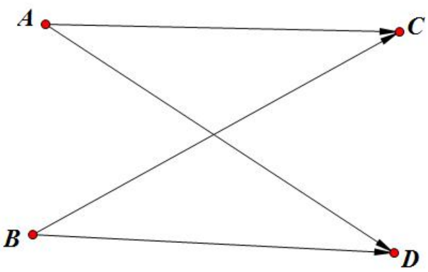

Figure 1.The Simlicity of Calling Structure

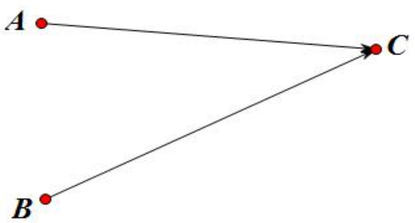

Figure 2. The Simlicity of Calling Structure (A)

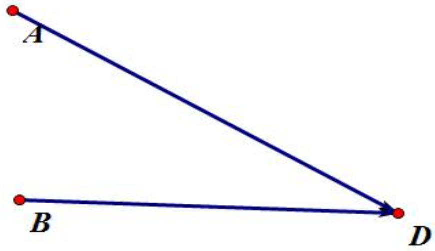

Figure 3.The Simlicity of Calling Structure (B)

Note that single propagation step can be separated and simplified with n front nodes $X_1, X_2 ... X_n$ and only one rear node O (As shown in Figure 4). The calling situation is that every member of subset of $\{X_1, X_2 ... X_n\}$ including empty set could be called at the same time. When all members in that subset are called at the same time, let's say the subset $\{X_{i_1} X_{i_2} ... X_{i_k}\}$ is called. The subset $\{X_{i_1} X_{i_2} ... X_{i_k}\}$ failed, when there is any member in that subset failed.

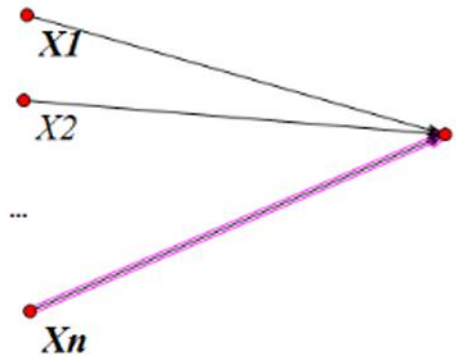

Figure 4.Details About Calling Structure

3 Different Amount of Information

The amount of information contributes to the complexity of the algorithm and simulation. As always, the more information you use, the more precise the answer is, the longer the time you spend. It is not usually necessary to use the whole information to get the precise situation ignoring the huge complexity and the time of simulation. This explains why we differentiate the methods when applying to different amount of information.

3.1. Full info situation

Full info means how often every subset is called has been known. In that way based on the frequency, the probability of every subset being called can be induced. Thus, suppose that $X_{i_1}, X_{i_2} ... X_{i_k}$ failed, by adding the probability of all subsets including $X_{i_1}, X_{i_2} ... X_{i_k}$, the probability of rear node O being failed is clear.

3.2 Little info situation

Albeit the situation of having full info is clear and easy to be accepted, the information usually won't be that much. Therefore, it is necessary to take a consideration of little information. Little information doesn't mean we know nothing but only the structure of the network. It means there are experts who can estimate the exclusiveness of the whole network. If the front nodes do not exclude others at all, it means for every calling, every front node has 50% chances to be called. If the nodes exclude others totally (It means there is only one front node can be called by the rear node O at a time), then the probability should be $\frac{1}{K}$ (K is the in-degree of rear node.) Building on exclusiveness from the experts, the probability should range from $\frac{1}{K}$ to 50%.

3.3 Situation with middle amount of info

Middle information situation is the most difficult one. The first problem is to what extent the amount of information you already have. Lots of situations should be considered. Only one special situation is going to be well developed.

The safety of scale-free network is different from evenly distributed network. Research has shown that scale-free network is robust to random attack and vulnerable to deliberate attack [12][13]. The phenomena aforementioned result from the existence of Hub nodes.

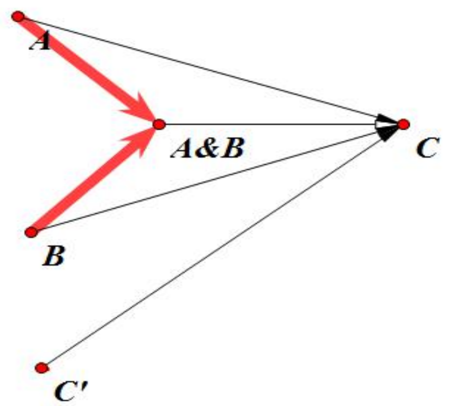

Figure 5.Pseudo-Nodes

A possible way to simulate the propagation situation is to add pseudo-nodes into the graph Figure 5. That node C calls node C' means node C calls itself. When node C calls A&B, it means node C calls A and B at the same time. In the graph, node C' and node A&B are called pseudo-nodes, but pseudo-nodes like A&B are alpha pseudo-nodes. Pseudo-nodes like C' are beta pseudo-nodes. When propagation starts, suppose node A is failed node, alpha pseudo-nodes get propagated firstly with a certain probability. Beta pseudo-nodes like C' act like normal nodes always have same state with their fathers.

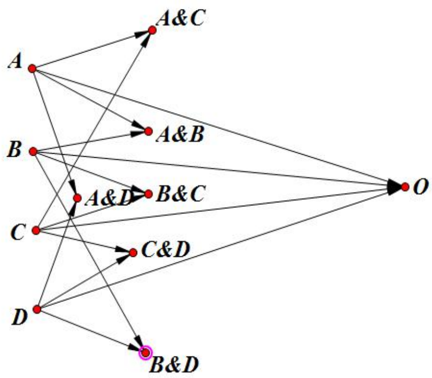

Figure 6. The Exploding Complexity With The Nodes

As can be seen in Figure 6, the more the front nodes are, the more complex the graph is. Apparently, this method can't be applied to all nodes in the graph. With the increasing of nodes, the complexity will go beyond our ability to handle. However, fortunately, in scale-free network what really matters is Hub nodes. When there are n front nodes to the Hub node and number n is big enough, the effect of a single front node will submerge to it of others. In addition, the complexity of pseudo-nodes method will increase with n dramatically. Therefore, using this method when n is big enough is not worthy Figure 6. It's better to apply this method only to Hub nodes with a few front nodes, while applying the low level information method to the rest nodes. By doing so, this method avoids big mistake with little increasing of complexity. Great balance between performance and complexity finally comes.

4 Examples and Analysis

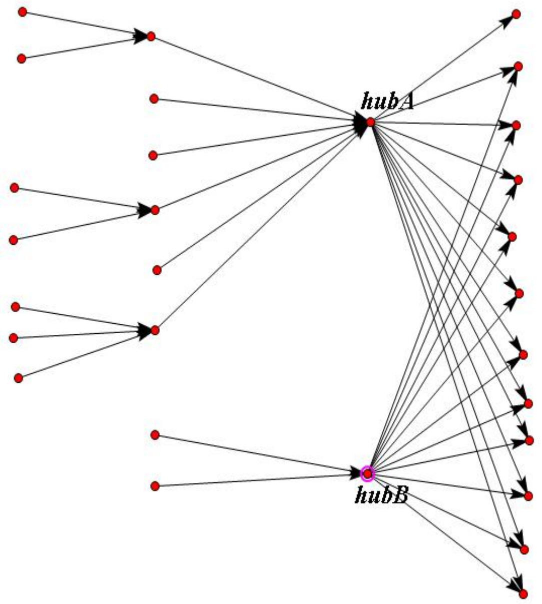

Figure 7.A Simple Example of Scale Free Network

We apply different methods to simulate and analyze faults propagation in Figure 7. The first method is computer virus propagation method. The computer virus propagates according to $\frac{1}{k}$ (k is the out degree of faulted node). The second method is the middle amount of info method mentioned in this paper. In detail, the propagation probability is given by these methods, and then the Monte Carlo simulation is used to test whether the methods are valid. Apparently, there are 2 hub nodes in the network, but the hubA is with too many front nodes. As a result, we only apply pseudo-nodes method to hub. On average, the $\frac{1}{k}$ propagating method has a probability of 23% to be exactly the same with the reality. However our method has a probability of 46 % to be exactly the same with the reality. The outcome is quite clear. The second method is much closer to the reality. The reason why this happened is the second method takes the calling frequency into consideration with the key nodes.

5 Questions Remaining

When there is little information to use, how to give the probability of propagation based on different exclusiveness is still unknown. The application of this method to free-scale network is meaningful. However, what is the complexity of this algorithm. Also, is it possible to apply this method to general network instead of Scale-free network? We are looking forward to the readers to give some clues of questions above.

Acknowledgement

Thanks to Feilong Wang, it is him who inspired me to do research about failure propagation in network. Also, during the cooperation with institute 716, we knew the application of propagation of failure in network. Without the application, we have no ways to know the flaws of the old model.

References

[1] Tan Yuejin, Wu jun. The entropy of network structure and its application in scale-free network. System engineering theory and application, 2004,24(6):1-3
[2] Faloutsos M, Faloutsos P, et al. On power-law relationships of the Internet topology.
[3] Chen Q, Chang H, et al. The origin of power-laws in internet topologies revisited. http://topology.eecs.umich.edu/data.html
[4] Newman M E J, Girvan M. Finding and evaluating community structure in networks. Phys. Rev.E, 2004, 69:026113
[5] Reichardt J, Bornholdt S. Detecting fuzzy community structures in complex networks with a Potts model. Phys. Rev. L, 2004,93(21):218701
[6] Pastor-Satorras R, Vespignani A. Epidemic dynamics and endemic states in complex networks. Phys. Rev. E, 2001, 63:066117
[7] Pastor-Satorras R, Vespignani A. Epidemic spreading in scale-free networks. Phys. Rev. Lett., 2001, 86: 3200-3203
[8] Albert R, Barabasi A L. statistical mechanics of complex networks. Reviews of modern physics 2002, 74(1): 47-97
[9] Wang Lin, Dai Guanzhong. The scale- free attributes in complex network, scale-free phenomena and its control. 2009 Science Press, 157
[10] [Barabasi A L, Albert R et al. Power-law distribution of the World Wide Web. Science, 2000, 287: 2115a]
[11] Chengxu, Wangyi, Zhu Huaihong. the probability model of computer network virus propagating. Microcomputer Developing, 14(12):204
[12] Albert R, Jeong H, et al. Eror and attack tolerance of complex networks. Nature, 2000,406:378-382
[13] Holme P, Kim B J, et al. Attack vulnerability of complex network. Phys. Rev. E, 2002, 65:056109/1-056109/5