Toward Bio-Inspired Network Robustness - Step 1. Modularity

Suyong Eum, Shin'ichi Arakawa, and Masayuki Murata
Osaka University, Graduate School of Information Science and Technology
1-5 Yamadaoka, Suita, Osaka, 565-0871 Japan
{suyong, arakawa, murata}@ist.osaka-u.ac.jp

ABSTRACT

Biological systems have evolved themselves to withstand against perturbations so that a characteristic, called robustness, is the most commonly observed feature in all living organisms. To find out the secret of robustness in biological systems, many researchers have investigated the system level structure of biological organizations. One of the known structural features that enable biological systems to be robust is modularity.

In this paper we study the correlation between modularity structure and robustness in IP networks. We carry out a simulation study to observe resistibility of different topologies, which have different level of modularity structure, against a perturbation created synthetically. The numerical results show that the quantified modularity seems to be more important measure to understand robustness of IP networks than any other common properties such as clustering coefficient, degree distribution, and average path length.

Keywords

Modularity, Robustness, Random and intentional attack, cascading failure, traffic dynamic

1. INTRODUCTION

Many measures, which characterize complex networks, have been proposed and studied to predict its certain dynamic behaviors and to reveal its hidden properties. One of these hidden properties is robustness, which enables a system to withstand external and internal perturbations [5]. Since robustness is an indispensable property that guarantees certain performance of networks, researchers have tried to find a correlation between robustness of networks and measured features such as the average path length [7], clustering coefficient, and degree distribution [1].

Especially, the degree distribution of complex networks has attracted great attention from researchers after it was discovered that degree distributions of many different complex networks can be better described by a power law of the form $P(k) \sim k^{\gamma}$ rather than the conventional Poisson distribution. The power law degree distribution implies that a few nodes have extremely large number of links as well as there are large number of small degree nodes. Thus, if we assume that high degree nodes are more important than less degree nodes, the power law networks (also called scale free networks) provide clues that they are error tolerance - randomly chosen node is likely to be small degree nodes- and attack vulnerable - when high degree nodes are intentionally chosen [1]. In other words, complex networks with scale free property are robust to a random failure and fragile to an intentional attack.

It is an undeniable fact that the scale free property as well as the other measures introduced previously somehow contribute for understanding robustness of networks. The question is which one is better than the other to describe the robustness. It would be nice if we find a superior measure against the others to explain robustness of all different kinds of complex networks. However, finding such a measure is a difficult task since each complex network has heterogeneous characteristics. Therefore, rather than trying to find an universally accepted measure, categorizing networks and differentiating disturbances are easier and faster way to understand the problem.

As the title of this paper implies, we have been developing a robust network model based on biologically inspired approaches. Reviewing some biology papers, we found an interesting fact that many biological systems have a specific topological structure called modularity. For instance, Ravasz et al [8] found that 43 metabolic networks they observed were organized as highly connected structural modules. Also, the relation between robustness and modularity structure of biological system has been briefly mentioned in [11] that modularity structure localizes damages and prohibits the malicious effect from spreading to the whole system. The structural characteristic called modularity attracted our attention because the modularity structure found in biological systems may be a key element to understand robustness of any system and we may be able to include the property in our robust network design process. This is the motivation of this paper to find whether modularity really provides sufficient information about robustness of complex networks.

In this paper we study the impact of modularity property on robustness of IP (Internet Protocol) networks. It is not difficult to image that IP networks have modularity

Bionetics '07, December 10-13, 2007, Budapest, Hungary
Copyright 2007 ICST 978-963-9799-11-0.

structure due to the fact that they consist of a collection of PoPs (Point of Presence) which can be regarded as a module. The main contribution of this paper is to show the correlation between robustness and modularity property in different topologies. To demonstrate the correlation, we create a disturbance to network topologies that have the different strength of modularity, and observe the resistivity of the topologies against the perturbation.

The rest of this paper is organized as follows. In Section 2, we provide a toy example that provides an intuitive understanding of the concept of modularity, and the impact of modularity on robustness of networks. This is followed by the quantification of modularity, and we discuss why IP networks have modularity structure in Section 3. Section 4 provides the experimental setup used in our evaluation study and present results from a perturbation scenario. Finally, we conclude the paper in Section 5.

2. WHICH TOPOLOGY IS MORE ROBUST?

Understanding robustness of networks starts from understanding different metrics from the theory of complex networks. Most commonly used metrics are average path length, clustering coefficient, and degree distribution. The average path length is calculated by averaging shortest paths between any pair of nodes in the network. This metric is closely related to a cascading failure or a virus spreading in the network. For instance, a network with small average shortest paths is vulnerable to a virus spreading. Clustering coefficient of a node, the ratio between the number of actually connected neighbor nodes and the number of possible connections, shows how well the neighbor nodes are connected each other. Therefore, high clustering coefficient networks are likely (but not always) to be robust against random failure of links or nodes. The degree distribution was briefly described in the previous section. Degree of a node simply represents the number of links which are connected to neighbor nodes so that degree distribution shows the overall connectivity among nodes in the network.

Although these measures are useful information to reveal certain characteristics of networks, it is not good enough to show the structural property of a network, which is important to understand its robustness. This can be illustrated by a toy example as follow.

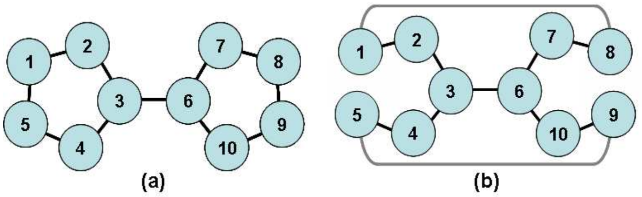


Figure 1: Toy example.

The both topologies (a) and (b) have the same average clustering coefficient 0 and average degree 2.2. The diameters, which is defined to be the maximal distance among all distances between any pair of nodes, are the same to be 5 in each topology. Also, the distribution of degree for both topologies are the same (degree exponent, $\gamma = 2.2$). The both topologies can be easily extended to be scale free networks by adding some additional links to each node. The only difference is the average path lengths which are 2.56 and 2.44 in both topologies (a) and (b) respectively.

In this toy example, seeing the visualized topologies in Fig. 1 can provide more useful information about the robustness of topologies than reading the measured metrics. For instance, the topology (a) in Fig. 1 can be divided into two groups easily by cutting the link connecting the node 3 and 6. Therefore, the topology (a) must be less robust than the one (b) for a link failure case. How about a different perturbation such as a failure cascading? Assuming that the node 1 experiences a certain failure, how much time does it take the failure to reach to the node 9, which is the farthest node from the node 1? Since the number of hops between node 1 and 9 are 5 in both topologies, it may take the same amount of time. However, since the failure has to go through the node 3 to reach to node 9 in the topology (a), the structure of topology (a) restricts the failure spreading more efficiently than the topology (b), which implies that the topology (a) is more robust than the other topology (b) against a failure spreading case. How about a traffic dynamic scenario? For example, when a link is congested, some flows on the link need to be re-routed. In that case, which topology experiences less disturbance by the re-routed flows? It is probably topology (a) because when a link in one of two isolated local networks composed of nodes (1-2-3-4-5) and (6-7-8-9-10) experiences a congestion, flows on the link tend to be re-routed inside the local network so that the disturbance caused by the re-routed flows is localized.

Since the robustness is measured in terms of the network behavior as a function of disturbance [9], the toy example clearly shows that robustness of a network needs to be understood according to a disturbance.

3. MODULARITY IN IP NETWORKS

One question we need to answer from the toy example is which structural property enables us to predict the robustness of the networks before any analytical and numerical studies. We think the answer is modularity. The property has been studied in different areas for a long time with different names such as community structure, graph partitioning in graph theory and computer science, and hierarchical clustering in sociology [6]. Modularity can be defined as a network property which shows how easily a network can be divided into groups. The topology (a) in Fig.1 can be divided into two subnetworks more easily than topology (b) so that we can say the former topology has higher modularity value than the latter.

Modularity structure is observed in many different kinds of networks from biological networks to technical networks such as IP networks. IP networks consist of point-of-presence (POPs), which can be regarded as modules. Inside a PoP, BR(Backbone Router) and ER(Edge Router) are densely connected, and PoPs are connected each other with core backbone links. As we observed in the toy example, this structural property impacts on behavior of networks against certain perturbations.

In [6], Newman et al developed an algorithm to quantify the strength of modularity. The method follows an iteration process. The first step involves finding of a link where most flows use. Such a link is called a high betweenness link. In the topology (a) in Fig. 1, the link composed of node 3 and 6 has the highest betweenness value. After the highest betweenness link is found, the link is removed, and then if the disconnection splits a network, the strength of modularity is calculated using the following equation.

$$\mathrm{Q} = \sum_i (e_{ii} - a_i^2) \qquad (1)$$

Where Q is the quantified modularity value, e is a symmetric matrix which represents the connectivity among modules. The dimension of e is the same as the number of modules in the network. Also, a_i represents a row or column sum of matrix $a_i = \sum_j e_{ij}$. As a result, the value Q shows the ratio between the number of links in modules and that of links among modules.

Newman et al stressed that a high betweenness link needs to be found again in each iteration. Since the iteration process produces a series of modularity values, the maximum value is chosen to be the modularity value of the network. In Fig. 1, the modularity values obtained using Newman's method for topologies (a) and (b) are 0.314 and 0.0165 respectively, and they are well matched to the structural observation of two networks.

4. RESULTS AND DISCUSSION

We have shown the concept of modularity and the impact of modularity structure on robustness of networks using a toy example. In the example, we guessed the behaviors of networks purely based on the structure observation in three different perturbation scenarios, namely random link failure and intentional attack, failure cascading, and traffic dynamic. In this section we carry out simulation studies whether the guess in the toy example is correct in real IP topologies. A simulation program, using the C^{++} language, was built to simulate a perturbation.

Table 1: Real ISP topologies, and topologies generated from BA and FKP models that have the same number of nodes and links with the real one. Modularity values of topologies from BA model are set to zero since they produced negative values. (ACC: Average Clustering Coefficient, AD: Average Degree, ASP: Average Shortest Paths, DE: Degree Exponent, M: Modularity).

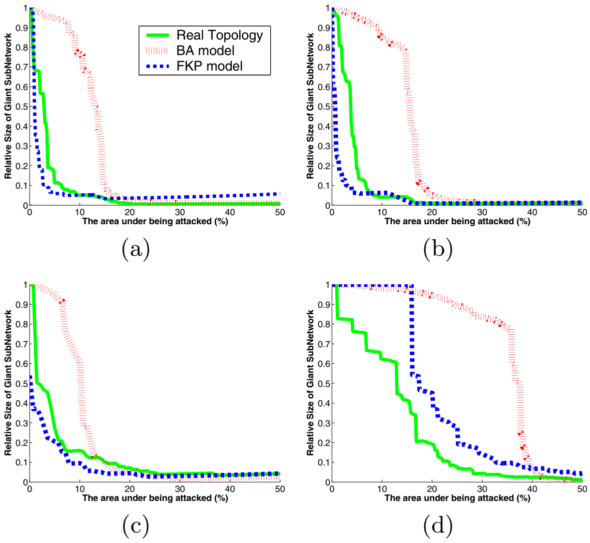
		ACC	AD	DE	ASP	M
ORG	AT&T	0.166	4.987	1.68	5.073	0.784
	Sprint	0.399	5.482	1.66	4.062	0.732
	Ebone	0.298	3.729	1.22	4.974	0.724
	Level3	0.260	17.01	1.00	3.356	0.238
BA	AT&T	0.047	4.987	2.58	3.429	0.000
	Sprint	0.063	5.482	2.75	3.338	0.000
	Ebone	0.045	3.729	2.46	3.201	0.000
	Level3	0.159	17.01	1.32	2.529	0.000
FKP	AT&T	0.835	4.987	5.36	4.677	0.832
	Sprint	0.894	5.482	6.36	4.708	0.809
	Ebone	0.221	3.729	1.27	4.580	0.538
	Level3	0.602	17.01	2.15	3.229	0.211

We use four of real ISP (Internet Service Provider) level topologies, which were obtained by Spring et al in [10]. We manipulated the topology data to extract link information for our simulation study. The four ISP level topologies are AT&T, Sprint, Ebone, and Level3.

In addition, we generated topologies from BA [2] and FKP [4] models respectively in order to generate topologies with different modularity levels. For example, BA model is based on the preferential attachment algorithm so that nodes in the topologies tend to be connected each other strongly. The strong connectivity makes it difficult to split the network, and it causes the network to have a weak modularity value. On the other hands, FKP model makes use of measures of betweenness of nodes for the network growth algorithm so that cutting some high betweenness links are likely to split the network. That is why topologies from FKP model produce relatively high modularity values. Table 1 shows topologies categorized into three main groups, which are original ISP topologies, topologies generated from BA [2] and FKP [4] models with their calculated properties.

4.1 Random and intentional attack

Random failure and intentional attack are the most well known scenarios to examine robustness of networks. Albert et al [1] explored the scenario and suggested that scale free networks are robust against random attack but fragile to intentional attack. However, if all networks we consider are scale free networks, is it possible to tell which topology is more robust than the other under the disturbance? In this section, we want to observe which property of topology provides most useful information to predict the robustness against these random and intentional attack.

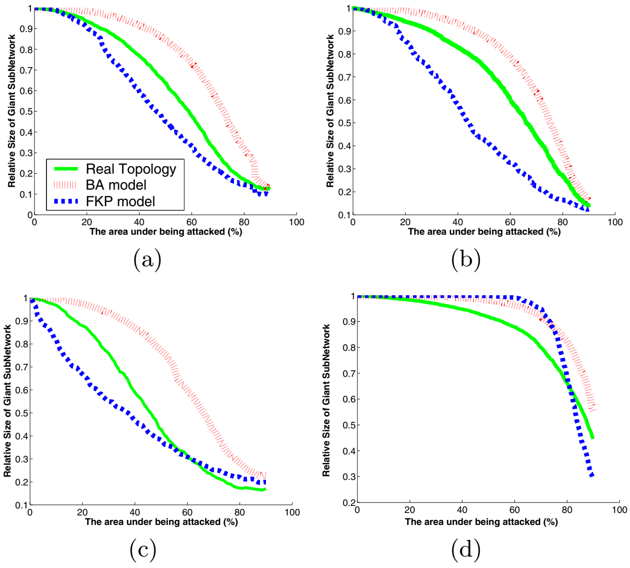


(a)
(b)
(c)
(d)

Figure 2: Intentional node removals based on betweenness centrality: (a)AT&T (b) Sprint (c) Ebone (d) Level3

Figure 2 shows the variation of diameter of networks as a node is removed in the order of highly utilized one. The highly utilized node is found by counting how many flows go through the node. In this perturbation scenario, a topology that keeps its size longer against the node removal is regarded as more robust. Most interesting observation is that maintenance of connectivity in each topology seems to be inversely proportional to the modularity value. It means that a network topology with smaller modularity value keeps its function longer than the one with higher modularity value. We can observe the correlation in Fig. 2(a), 2(b), and 2(d).

Although, Ebone topology provides a result that the original Ebone topology keeps its connectivity longer than the one from FKP model in spite of its larger modularity value, modularity values seem to provide more useful information to predict the robustness against this intentional attack scenario than the any other measures in Table 1.



(a)

(b)

(c)

(d)

Figure 3: Random node removals: (a)AT&T, (b) Sprint, (c) Ebone, (d) Level3

Figure 3 shows the case of random node attack. Since the topologies we consider are scale free networks, we observed the same result as Albert et al [1] did, which is that networks seem to be robust to a random attack but fragile to an intentional attack.

What we are interested in this result is that the resistivity of topologies against the random attack is also in inverse proportion to the calculated modularity values. Once again the modularity values well predict the robustness of topologies against the random attack scenario better than the other measures shown in Table 1.

5. CONCLUSIONS

We showed a toy example, which provides an intuitive understanding about robustness of networks in different disturbances. The robustness of each topology was explained based on its topological structure called modularity.

To extend our understanding from the toy example to real complex networks, we carried out numerical simulations to find out the correlation between modularity structure and robustness of networks in real IP networks. We simulated different perturbations and observed behavior of the networks under the synthetic perturbations. The simulation results confirmed that the modularity value of a network plays an important role to discover the robustness of networks.

In addition, topologies generated from FTK model produced more similar behavior to the original topologies than BA topologies under the synthetic perturbations. The results suggest that accuracy of network models needs to be verified using not only common metrics such as the average path length, clustering coefficient, and degree distribution but also how well the network produced by the model duplicates a certain behavior against different perturbations.

Since the correlation between modularity structure and robustness of networks is revealed satisfactorily in the simulation study, further work will focus on including this structural property into the robust network design process.

6. ACKNOWLEDGMENTS

This work was supported in part by a grant of Special Coordination Funds for Promoting Science and Technology: Yuragi Project from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

7. REFERENCES

[1] R. Albert, H. Jeong, and A. Barabsi. The Internet's Achilles' Heel: Error and attack tolerance of complex networks, 2000.

[2] A. L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, October 1999.

[3] A. Fabrikant, E. Koutsoupias, and C. Papadimitriou. Heuristically optimized tradeoffs: A new paradigm for power laws in the internet. In *Heuristically optimized tradeoffs: A new paradigm for power laws in the internet*. 34th Symposium on Theory of Computing, 2002.

[4] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On Power-law Relationships of the Internet Topology. In *SIGCOMM*, pages 251–262, 1999.

[5] H. Kitano. Biological robustness. *Nat Rev Genet*, 5(11):826–837, November 2004.

[6] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks, August 2003.

[7] A. Ng and J. Efstathiou. Structural Robustness of Complex networks. NetSci, International Workshop and Conference on Network Science, March 2006.

[8] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A. L. Barabasi. Hierarchical Organization of Modularity in Metabolic Networks. *Science*, 297:1551–1555, 2002.

[9] M. Savageau. Parameter sensitivity as a criterion for evaluation and comparing the performance of biochemical systems. *Nature*, 229:542–544, 1971.

[10] N. Spring, R. Mahajan, D. Wetherall, and T. Anderson. Measuring ISP Topologies With Rocketfuel. *IEEE/ACM Transactions on Networking.*, 12(1), February 2004.

[11] J. Zhao, H. Yu, J. Luo, Z. W. Cao, and Y.-X. Li. Complex networks theory for analyzing metabolic networks. *Chinese Science Bulletin*, 51:1529, 2006.