Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

# Range-based attacks on links in random scale-free networks

**Baihua Gong**[1,2], **Jun Liu**[2], **Liang Huang**[1,2], **Kongqing Yang**[1,3] **and Lei Yang**[1,2,4]

[1] Institute of Modern Physics, Chinese Academy of Science, Lanzhou 730000, People's Republic of China
[2] Department of Physics, Lanzhou University, Lanzhou 730000, People's Republic of China
[3] Institute of Applied Physics, Jimei University, Xiamen 361021, People's Republic of China
E-mail: gongbh03@st.lzu.edu.cn, liujun03@st.lzu.edu.cn, liang.huang@asu.edu, yangkq@jmu.edu.cn and lyang@impcas.ac.cn

Received 23 August 2007
Accepted 18 January 2008
Published 14 February 2008

Online at stacks.iop.org/JSTAT/2008/P02008
doi:10.1088/1742-5468/2008/02/P02008

**Abstract.** *Range* and *load* play key roles in the problem of attacks on links in random scale-free (RSF) networks. In this paper we obtain the approximate relation between *range* and *load* in RSF networks by the generating function theory, and then give an estimation about the impact of attacks on the *efficiency* of the network. The results show that short-range attacks are more destructive for RSF networks, and are confirmed numerically.

**Keywords:** exact results, network dynamics, random graphs, networks



[4] Author to whom any correspondence should be addressed.


1742-5468/08/P02008+6$30.00

Attacks on complex networks, especially in the context of the Internet and biological networks, have been an interesting issue, and different aspects of attacking have been analyzed recently [1]. Many works focus on attacks on nodes, and the strategies provided include random attacks, degree-based attacks, etc [2]. Also, some works consider attacks on links, and the strategies include range-based attacks, load-based attacks, etc [3].

Motter *et al* [4] studied attacks on links in scale-free networks basing on *range*. *Range* is introduced by Watts [5] to characterize different types of links in networks: the range of a link $l_{ij}$ connecting nodes $i$ and $j$ is defined as the length of the shortest path between the nodes $i$ and $j$ in the absence of $l_{ij}$. The small-world model introduced by Watts and Strogatz [6] (WS model) is more sensitive to attacks on long-range links connecting nodes that would otherwise be separated by a long distance. It is not true for many scale-free networks, though most of them also have a short average path length like the WS model. Motter *et al* found that short-range links rather than long-range ones are vital for efficient communication between nodes in these networks. They argued that the average shortest path is a global quantity which is mainly determined by links with large *load*, where the *load* of a link is defined as the number of shortest paths passing through this link [7, 8]. And for scale-free networks, with exponent in a finite interval around 3, due to the heterogeneous degree distribution, the *load* is on average larger for links with shorter *range*, making the short-range attacks more destructive.

In this paper, employing the generating function theory, we first derive an approximate relation between $R(k_1, k_2)$ and $L(k_1, k_2)$ for RSF networks analytically, where $R(k_1, k_2)$ and $L(k_1, k_2)$ are defined as the expected value of *range* and *load* respectively for links between nodes with given degree $k_1$ and $k_2$. We then give an estimation about the decrement of *efficiency* as a function of $R(k_1, k_2)$ and $L(k_1, k_2)$, showing that short-range attacks are more destructive for RSF networks. Numerical simulations are also performed to confirm our analytical results.

To study range-based attacks on links in RSF networks, we measure the *efficiency* of the network as each link is removed. The *efficiency* of a network with size $N$ is defined as [9]

$$E = \frac{2}{N(N-1)} \sum \frac{1}{d_{ij}}, \tag{1}$$

where $d_{ij}$ denotes the length of the shortest path between the node-pair $(i, j)$; the sum is over all pairs of nodes in the network. The *efficiency* defined above has a finite value even for disconnected networks, and larger values of $E$ correspond to more efficient networks.

When a link is removed from the network, the *efficiency* of the network generally decreases. The decrement of *efficiency* involves two quantities: (1) the number of node-pairs whose geodesic lengths increase; (2) the average increment of the geodesic lengths of these node-pairs. The first quantity is related to the *load* of the removed link, and the second quantity is related immediately to the *range* of the removed link.

For RSF networks, the expected value of the geodesic length of node-pairs with given degree $k_1$ and $k_2$ is

$$d(k_1,k_2) = \sum_{i=1} ip^i\,(k_1,k_2)\,,$$



J. Stat. Mech. (2008) P02008

where $p^i(k_1,k_2)$ is the probability that the node-pair with given degree $k_1$ and $k_2$ has a geodesic length $i$. For RSF networks, we have[5]

$$p^1(k_1,k_2) \approx \frac{k_1 k_2}{2N z_1}, \tag{2}$$

where $N$ is the number of nodes in the network, and $z_1$ is the average number of first neighbors. By the generating function formalism, we can obtain [10]

$$d(k_1,k_2) \approx 1 + \frac{\ln(N \cdot z_1/(k_1 \cdot k_2))}{\ln(z_2/z_1)}, \tag{3}$$

where $z_2$ is the average number of second neighbors. Accordingly the expected diameter of RSF networks is [10]

$$D \approx 1 + \frac{\ln(N/z_1)}{\ln(z_2/z_1)}. \tag{4}$$

Since the RSF network is totally random in all aspects other than the degree distribution, $R(k_1, k_2)$ is thus equal to the expected value of the geodesic length of nonadjacent node-pairs with given degree $k_1 - 1$ and $k_2 - 1$, that is

$$R(k_1,k_2) = \sum_{i=2} i \frac{p^i\,(k_1-1,k_2-1)}{1-p^1(k_1-1,k_2-1)},$$

i.e.,

$$R(k_1,k_2) = \frac{d(k_1-1,k_2-1) - p^1(k_1-1,k_2-1)}{1-p^1(k_1-1,k_2-1)}. \tag{5}$$

Combining equations (2), (3) and (5), we can obtain

$$R[(k_1-1)(k_2-1)] \approx 1 + \frac{\ln(Nz_1/((k_1-1)(k_2-1)))/\ln(z_2/z_1)}{1-(k_1-1)(k_2-1)/2Nz_1}. \tag{6}$$

Furthermore, we assume that the network is spare, and can be seen as a tree with expected diameter $D$. Consider a link $l_{ij}$ connecting node $i$ and $j$, where $i$ has a degree $k_1$ and $j$ has a degree $k_2$. When removing $l_{ij}$, the network can be regarded as a tree $T_i$ rooted as $i$ or $T_j$ rooted as $j$, both of which have a depth of $D - 1$. Staring from the root $i$, the first layer has $k_1 - 1$ nodes, the second layer has $z_1(k_1 - 1)$, and the $m$th $(0 < m < D)$ layer has $z_1^{m-1}(k_1 - 1)$ nodes. Similarly, the $m$th layer of $T_j$ has $z_1^{m-1}(k_2 - 1)$ nodes. The geodesic path from nodes in the $d_1$th $(d_1 < D - 1)$ layer in $T_i$ to nodes in the $d_2$th $(d_2 <= D - 1 - d_1)$ layer in $T_j$ is expected to pass $l_{ij}$, which has a contribution of 1 to

[5] The RSF network considered here has in total $3Nz_1$ half-links. Pairs of half-links are chosen and connected totally randomly; thus the probability of a pair of nodes with degree $k_1$ and $k_2$ connected directly should be equal to $p^1(k_1,k_2) \approx k_1k_2/2Nz_1$.



J. Stat. Mech. (2008) P02008

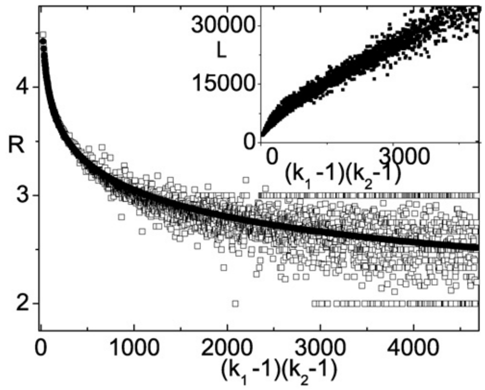


**Figure 1.** Average *range* as a function of the product $(k_1-1)(k_2-1)$ in RSF networks with $N=10^4$, the exponent of the degree distribution $\lambda=3.5$, the minimal degree $m_0=6$, and the maximal degree $m_{\max}=500$. The solid line is the theoretical curve and the hollow squares are simulation results. Inset: average *load* as a function of the product $(k_1-1)(k_2-1)$. Numerical data are obtained from 100 realizations.

the load of $l_{ij}$. Thus the expected value of the load of $l_{ij}$ is

$$\begin{aligned} L(k_1,k_2) &= \sum_{d=1}^{D-2}\left((k_1-1)\frac{(z_1-1)^d-1}{z_1-2}+1\right)(k_2-1)(z_1-1)^{D-d-2} \\ &\quad + \left((k_1-1)\frac{(z_1-1)^{D-1}-1}{z_1-2}+1\right)+(k_2-1)(z_1-1)^{D-2} \\ &= \left(\frac{(D-2)(z_1-1)^{D-2}}{z_1-2}-\frac{(z_1-1)^{D-2}-1}{(z_1-2)^2}\right)((k_1-1)(k_2-1)) \\ &\quad + (k_1+k_2-2)\frac{(z_1-1)^{D-1}-1}{z_1-2}+1. \end{aligned} \tag{7}$$

When $k_1 \gg z_1, k_2 \gg z_1$, the above equation can be rewritten as

$$L(k_1,k_2)=\frac{((D-2)(z_1-2)-1)(z_1-1)^{D-2}+1}{(z_1-2)^2}(k_1-1)(k_2-1), \tag{8}$$

showing that the *load* is directly proportional to the product of $(k_1-1)$ and $(k_2-1)$ when $k_1$ and $k_2$ are large enough. For simplicity, we rewrite equation (8) as

$$L[(k_1-1)(k_2-1)] = c(k_1-1)(k_2-1), \tag{9}$$

where $c$ is the coefficient $(((D-2)(z_1-2)-1)(z_1-1)^{D-2}+1)/(z_1-2)^2$.

The above analytical results can be numerically verified in the following. We plot $R(k_1,k_2)$ in figure 1, and $L(k_1,k_2)$ in the inset of figure 1. From the inset of figure 1, it can be seen that the *load* is directly proportional to the product $(k_1-1)(k_2-1)$ when $(k_1-1)(k_2-1)$ is large enough.

Combining equations (6) and (9), we can see that the *load* and the *range* have a negative correlation, that is

$$R(L) \approx 1+\frac{\ln(cN\cdot z_1/L)/\ln(z_2/z_1)}{1-L/2Ncz_1}. \tag{10}$$

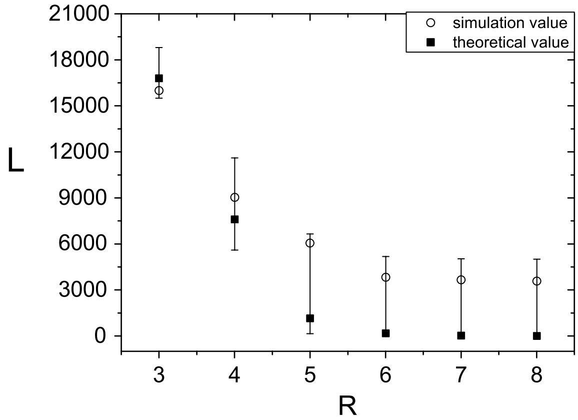


**Figure 2.** Average *load* as a function of *range*. Square: theoretical value; circle: averaged simulation value over 100 realizations. The error bar is also given. (All the parameters are the same as in figure 1). The values for $R = 2$ are much larger than for other $R$, and thus are not plotted in this figure.

This expression gives an estimation of the relation between $R$ and $L$, suggesting that short-range links are expected to be passed through a large number of shortest paths. From equations (8) to (9), the condition $k_1 \gg z_1, k_2 \gg z_1$ is used; thus equation (9) is valid when $(k_1 - 1)(k_2 - 1)$ and $L$ are large. As a result, equation (10) is valid when $R$ is small. Numerical verification is presented in figure 2. When $R = 2$ the numerical estimation is 1200 000, and the analytical value is 996 000; when $R = 3$ the simulation value is 16 000, and the analytical value is 16 800; when $R = 4$ the simulation value is 9031, and the analytical value is 7605. The simulation values in all three cases are well consistent with the corresponding analytical values. When $R = 5$, the simulation value is 6062, and the analytical value is 1151; when $R = 6$, the simulation value is 3832, and the analytical value is 176. The simulation values in the above two cases and in the cases of $R > 6$ have significant discrepancy with the corresponding analytical values. This is because, when $R$ is large and $L$ is small, the approximation $k_1 \gg z_1, k_2 \gg z_1$ does not hold any more, and our analysis is not valid either. From figure 2, we can see that equation (10) gives a good approximation of the relation between $R$ and $L$ for small values of $R$, and is not valid for large values of $R$.

When a link $l_{ij}$ is removed from the network, the decrement of the *efficiency* of the network is approximately

$$\Delta E \approx \frac{2}{N(N-1)} \sum_{(m,n)\in\Gamma} \frac{R(i,j)-1}{d_{mn}^2} \approx \frac{2(R(i,j)-1)L(i,j)}{D^2 N(N-1)}, \tag{11}$$

where the set $\Gamma$ is all the node-pairs whose shortest length should increase as a result of the removal of link $l_{ij}$, $R(i,j)$ and $L(i,j)$ are the *range* and *load* respectively of $l_{ij}$. Thus the product $(R-1)L$ is a natural quantity to characterize the impact of removing a link on the *efficiency*. For RSF networks,

$$\Delta E \approx hL(i,j) \frac{\ln(cN \cdot z_1/L(i,j))/\ln(z_2/z_1)}{1 - L/2Ncz_1}, \tag{12}$$

where $h$ is $2/D^2N(N-1)$. $\Delta E$ is an increasing function of $L$, and thus a decreasing function of $R$. It can be concluded that links with small range are more important for the efficiency of RSF networks.

In summary, by investigating the expected *range* and *load* of links in RSF networks, we obtain an approximate analytical relation between *range* and *load*, and then give an estimation of the impact of removal of links on the *efficiency*. Thus we prove analytically that attacks on short-range links are more destructive for RSF networks. An insufficiency in our work is that $R(L)$ has a significant discrepancy compared with numerical results for large value of $R$. However, the analytical results in this paper give a reasonable description for the trend of the true relation between $R$ and $L$ for RSF networks.

## Acknowledgments

The work is supported by the National Hi-Tech Research and Development Program with Grant No. 2006AA09A102-08 and the National Basic Research Program of China with Grant No. 2007CB209603. Lei Yang thanks the 100 Person Project of the Chinese Academy of Sciences, and China National Natural Science Foundation for their support with Grant No. 10775157.

## References

[1] Albert R, Jeong H and Barabási A-L, 2000 *Nature* **406** 378
Callaway D S, Newman M E J, Strogatz S H and Watts D J, 2000 *Phys. Rev. Lett.* **85** 5468
Solé R V and Montoya J M, 2001 *Proc. R. Soc.* B **268** 2039
Cohen R, Erez K, ben-Avraham D and Havlin S, 2001 *Phys. Rev. Lett.* **86** 3682
Jeong H, Mason S P, Barabási A-L and Oltvai Z N, 2001 *Nature* **411** 41
[2] Cohen R, Erez K, ben-Avraham D and Havlin S, 2000 *Phys. Rev. Lett.* **85** 4626
Callaway D S, Newman M E J, Strogatz S H and Watts D J, 2000 *Phys. Rev. Lett.* **85** 5468
[3] Girvan M and Newman M E J, 2002 *Proc. Nat. Acad. Sci.* **99** 8271
Holme P, Kim B J, Yoon C N and Han S K, 2002 *Phys. Rev.* E **65** 056109
[4] Motter A E, Nishikawa T and Lai Y-C, 2002 *Phys. Rev.* E **66** 065103(R)
[5] Watts D J, 1999 *Small World* (Princeton, NJ: Princeton University Press)
[6] Watts D J and Strogatz S H, 1998 *Nature* **393** 440
[7] Newman M E J, 2001 *Phys. Rev.* E **64** 016132
[8] Goh K-I, Kahng B and Kim D, 2001 *Phys. Rev. Lett.* **87** 278701
[9] Latora V and Marchiori M, 2001 *Phys. Rev. Lett.* **87** 198701
[10] Newman M E J, Strogatz S H and Watts D J, 2001 *Phys. Rev.* E **64** 026118

J. Stat. Mech. (2008) P02008