Journal of Information Hiding and Multimedia Signal Processing ©2017 ISSN 2073-4212
Ubiquitous International Volume **8**, Number **5**, September **2017**

Mining Critical Links in Complex Networks based on Entropy Weights and Grey Relational Analysis

Jin Si

Department of Computer and Information Technology
Zhejiang Police College
Hangzhou, 310053, P.R. China
sijin@zjjcxy.cn

Zhen Wu and Zhe-Ming Lu*

School of Aeronautics and Astronautics
Zhejiang University
Hangzhou, 310027, P. R. China
*Corresponding author: zheminglu@zju.edu.cn

Wen-Hua Ren

Department of Computer and Information Technology
Zhejiang Police College
Hangzhou, 310053, P.R. China
renwenhua999@163.com

Received January 2017; revised March 2017

Abstract. *Numerous and various complex networks have emerged with the development of science and technology. Finding the critical links in complex networks is very important for attackers and defenders. In order to evaluate the importance of links complex network links, it is crucial to choose the criteria comprehensively, objectively and independently. This paper presents an importance evaluation scheme to mine the critical links based on entropy weights and grey relational analysis. Experimental results demonstrate the superiority of our scheme in compassion with typical methods.*
Keywords: Complex networks, Network Security, Entropy weights, Grey relational analysis, Critical links mining.

1. **Introduction.** A complex network usually has a fragility that a relatively localized damage in one system may lead to a failure in another [1]. Understanding the fragility properly is one of the major challenges in the design of resilient infrastructures [2]. Small-world [3] and scale-free [4] are two important properties existing in most natural networks. Many previous works on the security of scale-free networks preferred attacks on nodes rather than on links [5, 6]. However, some other studies showed that attacks on links are as important as those on nodes [7]. Nevertheless, attacks on a few links may cause a sharp decline of network performance [8].

To solve this problem, some evaluation criteria have been proposed. One is the decline rate of spanning trees [9, 10]. The principle is that when an edge is deleted from the graph, the less the number of spanning trees is, the more important the edge is. Another criterion is the increasing rate of the average shortest path length[11, 12], although the edge-betweenness is also vital [5]. Actually, each criterion has its insufficiencies. First, most of the criteria use a single index, and hence it cannot be used for evaluation directly. Second, if multi indices are used, the correlation information between the indicators tends to be ignored. Thus, it is necessary to find an effective way to integrate various indicators. Recently, Ren and Lu[13] considered the multi-criteria based comprehensive ranking method for evaluating Link importance for communication networks using the gray relational analysis method. They characterized link importance based on three criteria.

To further improve the performance, this paper proposes a method to integrate the criteria effectively based on entropy weights and grey relational analysis. In Section 2, the grey relational analysis is introduced for criteria selection. In Section 3, we use entropy weights to solve the problem of subjectivity. In Section 4, experimental results and comparisons are performed based on two classical networks. A brief conclusion is given in Section 5.

2. **Criteria Selection Based on Grey Relational Analysis.** The grey relational analysis has been proved that it is an effective method to find the numerical relationship between various factors based on the trend of the differences or similarities. For mining critical links, the procedures are given below.

Step 1. Suppose there are n links in the network, and each edge has m factors. The following equation denotes the importance measure vector of the i-th edge.

$$X_i = (x_{i1}, x_{i2}, ..., x_{im}), 1 \leq i \leq n \tag{1}$$

where x_{ik} means the k-th measure factor of the i-th edge.

Step 2. Based on a comprehensive comparison of all links, the reference importance measure vector is given as below.

$$Y = (y_1, y_2, ..., y_m) \tag{2}$$

where y_k is the best value among the k-th measure factors of all links.

Step 3. To deal with the different magnitude of the factors, we use the average method to make indices dimensionless. The dimensionless importance measure vector is given as follows.

$$X_i^* = (\frac{x_{i1}}{ave_1}, \frac{x_{i2}}{ave_2}, ..., \frac{x_{im}}{ave_m}) \tag{3}$$

And the dimensionless reference vector is given as follows

$$Y^* = (\frac{y_1}{ave_1}, \frac{y_2}{ave_2}, ..., \frac{y_m}{ave_m}) \tag{4}$$

where ave_k is the average value over the k-th factors of all links, i.e.,

$$ave_k = \sum_{j=1}^{n} \frac{x_{jk}}{n} \tag{5}$$

Step 4. Compute the difference matrix as follows.

$$D = \begin{bmatrix} |y_1^* - x_{11}^*| & |y_2^* - x_{12}^*| & |y_3^* - x_{13}^*| & ... & |y_m^* - x_{1m}^*| \\ |y_1^* - x_{21}^*| & |y_2^* - x_{22}^*| & |y_3^* - x_{23}^*| & ... & |y_m^* - x_{2m}^*| \\ ... & ... & ... & ... & ... \\ |y_1^* - x_{n1}^*| & |y_2^* - x_{n2}^*| & |y_3^* - x_{n3}^*| & ... & |y_m^* - x_{nm}^*| \end{bmatrix} \tag{6}$$

Step 5. Calculate the correlation coefficient between the i-th edge and the k-th factor as follows.

$$r_{ik} = \frac{D_{\min} + rD_{\max}}{|y_k^* - x_{ik}^*| + rD_{\max}} \tag{7}$$

where r_{ik} means the correlation coefficient, and r is the resolution coefficient falling in the range of [0, 1].

Step 6. Calculate the final correlation degree as follows.

$$R_i = \frac{1}{m}\sum_{k=1}^{m} r_{ik} \tag{8}$$

where R_i means the correlation degree of the i-th edge and i=1, 2, 3,..., n.

Step7. Compare the correlation degrees of all links to output a sorted list of links according to their correlation degrees, and a bigger correlation degree value corresponds to a more important edge.

3. **The Improved Scheme.** Entropy is a physical quantity to indicate the degree of the state may appear. In information theory, entropy is a measure of the uncertainty. The smaller the information entropy of a target is, the greater the degree of variation of the index is. Thus, the greater the amount of information is provided, and the greater the weight is. On the contrary, if the information entropy of a target is larger, then the information provided is smaller and the weight should be correspondingly smaller. Following steps are a brief introduction to the entropy weighting method.

Step1. For a problem with n evaluation objects and m measure indices, its initial matrix is as follows.

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1m} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & x_{n3} & \dots & x_{nm} \end{bmatrix} \tag{9}$$

Step2. After normalizing the initial matrix, the normalized matrix is obtained as follows.

$$X^* = \begin{bmatrix} x^*_{11} & x^*_{12} & x^*_{13} & \dots & x^*_{1m} \\ x^*_{21} & x^*_{22} & x^*_{23} & \dots & x^*_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ x^*_{n1} & x^*_{n2} & x^*_{n3} & \dots & x^*_{nm} \end{bmatrix} \tag{10}$$

Where the normalized equation is given as

$$x^*_{ij} = \frac{x_{ij} - x_{\min}(j)}{x_{\max}(j) - x_{\min}(j)} \tag{11}$$

Where $x_{\mathtt{max}}(j)$ and $x_{\mathtt{min}}(j)$ represent the optimal value and the worst value of the j-th index over all objects.

Step3. Determine the entropy of the j-th index as follows

$$S_j = -\frac{1}{\ln n}\sum_{i=1}^{n} P_{ij}\ln P_{ij} \tag{12}$$

Where P_{ij} is defined as follows.

$$P_{ij} = \frac{1 + x^*_{ij}}{\sum_{j=1}^{m}(1 + x^*_{ij})} \tag{13}$$

Step4. Determine the entropy weight of the j-th index.

$$\omega_j = \frac{1 - S_j}{m - \sum_{i=1}^{m} S_i} \tag{14}$$

Based on the above grey relational analysis, entropy weights for all indices can be obtained. Using these entropy weights, we can modify Eq. (8) to obtain the final overall index for the i-th object as follows:

$$R_i = \sum_{k=1}^{m} \omega_k r_{ik} \tag{15}$$

where R_i means the overall index for the i-th edge, ω_k means the k-th entropy weight, r_{ik} is the correlation coefficient between the i-th edge and the k-th index. The bigger the R_i value is, the more important the edge is. In our method, we choose three classical evaluation criteria as our measure indices, i.e., the decreasing rate of the number of spanning trees, the increasing rate of the average shortest path length, and the edge-betweenness. For a network G=(V, E), where V denotes the set of nodes and E denotes the set of links, their definitions can be given as follows.

3.1. **Edge Betweenness.** Edge betweenness is a measure to quantify the ability of an edge in controlling the communication between nodes in a complex network. Let p_{kl} denote the number of shortest paths between Node $\mathtt{v}_k$ and Node $\mathtt{v}_l$, $p_{kl}(\mathtt{e}_{ij})$ be the number of shortest paths between Node $\mathtt{v}_k$ and Node $\mathtt{v}_l$ which must pass through the edge $\mathtt{e}_{ij}$, and $B_i j$ denote the edge betweenness of $\mathtt{e}_{ij}$. Then we have

$$B_{ij} = \sum_{k \neq l} \frac{p_{kl}(\mathtt{e}_{ij})}{p_{kl}} \qquad \mathtt{v}_k, \mathtt{v}_l \in \mathtt{V}, \mathtt{e}_{ij} \in \mathtt{E} \tag{16}$$

The larger the value B_{ij} is, the more important the edge $\mathtt{e}_{ij}$ is.

3.2. **The Decreasing Rate of the Number of Spanning Trees.** First we calculate the number of spanning trees of a network, according to the Matrix-Tree theorem, i.e., for an undirected network, let b_{ij} denote the associated number between Node $\mathtt{v}_i$ and Edge $\mathtt{e}_j$, if Node $\mathtt{v}_i$ and Edge $\mathtt{e}_j$ are connected, $b_{ij} = 1$; otherwise, $b_{ij} = 0$. Thus, the matrix **B** denotes the incidence matrix, and the number of spanning trees can be calculated as follows:

$$\tau(G) = |\mathtt{det}(C_r)| \tag{17}$$

Here, $\tau(G)$ is the number of spanning trees, C_r is the (n-1)th order principal minor of Kirchhoff matrix, and Kirchhoff matrix can be calculated as $\mathtt{BB}^T$. When we delete Edge $\mathtt{e}_i$, recalculate the number of spanning trees of the new network $\tau_i(G)$. Then we use $T(\mathtt{e}_i)$ to denote the decreasing rate of the number of spanning trees when Edge $\mathtt{e}_i$ is deleted, and $T(\mathtt{e}_i)$ can be calculated as follows

$$T(\mathtt{e}_i) = 1 - \frac{\tau_i(\mathtt{G})}{\tau(\mathtt{G})} \tag{18}$$

The larger the decreasing rate is, the more important the edge is.

3.3. **The Increasing Rate of the Average Shortest Path Length.** To obtain the increasing rate of the average shortest path length, we first calculate the average shortest path length of the original network, denoted as $L(\mathtt{G})$. For an undirected graph, the average shortest path length is defined as the average value over all the distances between every two nodes. When an edge $\mathtt{e}_i$ is deleted, we recalculate the average distance of the new network, denoted as $L_i(\mathtt{G})$. Thus, the increase rate of the average distance $D(\mathtt{e}_i)$ can be calculated as follows:

$$D(\mathtt{e}_i) = \frac{L_i(\mathtt{G}) - L(\mathtt{G})}{L(\mathtt{G})} \tag{19}$$

$$L(\mathtt{G}) = \frac{1}{N(N-1)} \sum_{j \neq i} d_{ij} \tag{20}$$

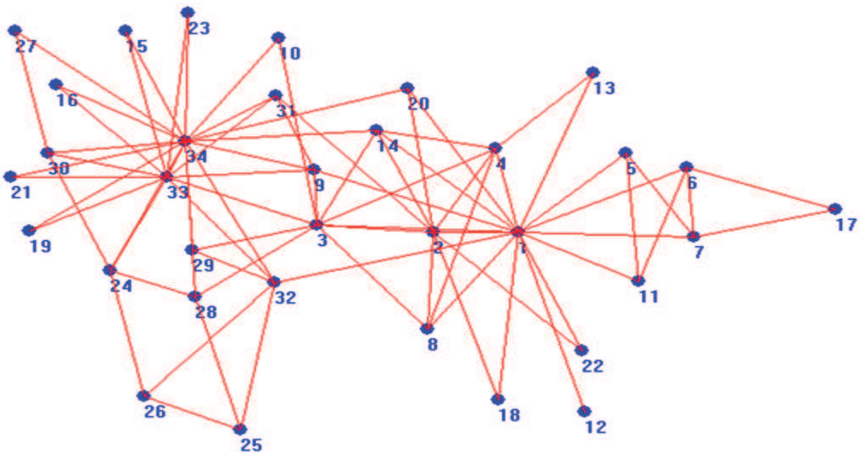

Figure 1. The karate club network.

Table 1. Critical links ranking of the Karate club network.

Critical links	**Method 1**	**Method 2**	**Method 3**	**Our**
1	1-12	1-12	1-32	2-20
2	7-17	1-32	1-7	1-32
3	6-17	1-11	14-34	3-33
4	27-30	1-5	1-6	14-34
5	27-34	1-6	1-9	1-9

Here, d_{ij} denotes the length of the shortest paths between Node $\mathtt{v}_i$ and Node $\mathtt{v}_j$,and N is the number of nodes. The larger the increase rate is, the more important the node is.

4. **Simulations.** Two classical networks are chosen to test our methods, i.e., the karate club network and the sex relationship network. Specifically, there are 34 nodes and 78 links in the karate club network as shown in Fig. 1. We chose each single evaluation criteria to determine the critical links. The results are listed in Table 1, where Method 1, Method 2 and Method 3 are based on the decreasing rate of the number of spanning trees, the increasing rate of the average shortest path length, and the edge-betweenness, respectively. In contrast, our method is based on combining entropy weights with grey relational analysis. From Table 1, it is easy to find that using each single method, the results may be correct in the local network, but are not comprehensive. Instead, using our scheme, we can get more reasonable results.

As shown in Fig. 2, the sex relationship network contains 40 nodes and 41 links. The same procedure is used to compare different methods, and the results are shown in Table 2. According to Method 1, almost 32 links have the same results, i.e., the decline rate of the number of spanning trees is 1.0. Using Method 2, for 32 links, the increasing rate of the average shortest path length is infinity. Thus, for this classical network, Method 1 and Method 2 do not work. According to the comparison between Method 3 and out method, we can see that the edge 11-16 is more important than the edge 16-22, because the edge 11-16 can cut the network into two independent networks while the edge 16-22 cannot. According to above experimental results, we can see that our method outperforms the previous works in mining the critical links.

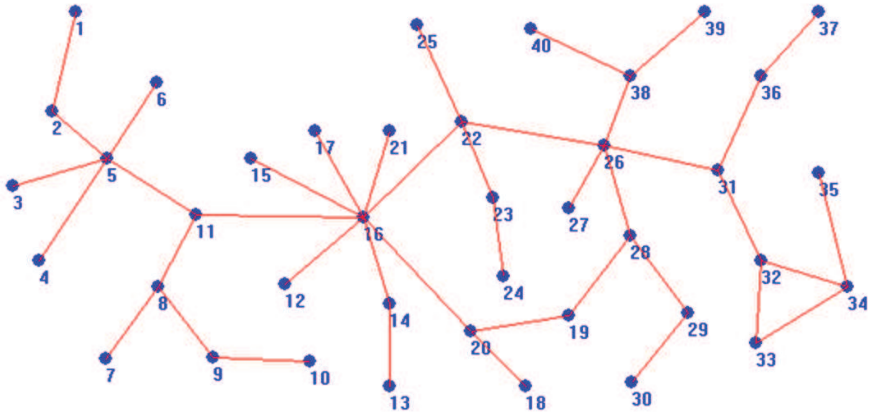

FIGURE 2. The karate club network.

TABLE 2. Critical links ranking of the sex relationship network.

Critical links	**Method 1**	**Method 2**	**Method 3**	**Our**
1	1-2	1-2	16-22	11-16
2	2-5	2-5	11-16	26-31
3	3-5	3-5	22-26	5-11
4	4-5	4-5	26-31	16-22
5	5-6	5-6	5-11	8-11

5. **Conclusions.** In this paper, we propose a new method to mine the critical links in complex networks. That is, the entropy weights and grey relational analysis are combined to integrate three criteria into one index. Two classical networks are used to test our method. According to the experimental results, we can conclude that our improved method in mining the critical links has two following advantages. First, our method can be used in different networks which may be not appropriate to other single available method. Second, the results of our method are comprehensive which cannot be achieved in each single method.

Acknowledgment. This work was partly supported by the Zhejiang Provincial Natural Science Foundation of China (Grant No. LY17F030008, No. LY15F010003 and No. LY14F020024), the Public Good Research Project of Science and Technology Program of Zhejiang Province, China. (NO.2016C31097) and the Science and Technology Project of QuZhou, Zhejiang, China.(NO.2015Y005).

REFERENCES

[1] A.Vespignani, Complex networks: The fragility of inter-dependency, *Nature*, vol. 464, no. 7291, pp. 984–985, 2010.
[2] S. M. Rinaldi, J. P. Peerenboom, and T. K. Kelly, Identifying, understanding, and analyzing critical infrastructure interdependencies, *IEEE Control System Magazine*, vol. 21, no. 6, pp. 11–25, 2001.
[3] D. J. Watts, and S. H. Strogatz, Collective dynamics of 'small-world' networks, *Nature (London)*, vol. 393, no. 6684, pp. 440–442, 1998.
[4] A. L. Barabasi, and R. Albert, Emergence of scaling in random networks, *Science*, vol. 286, no. 5439, pp. 509–512, 1999.

[5] M. Girvan, and M. E. J. Newman, Community structure in social and biological networks, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 21, pp. 8271, 2002.

[6] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han, Attack vulnerability of complex networks, *Physical Review E*, vol. 65, pp. 056109, 2002.

[7] Y. C.Lai, A. E. Motter, and T.Nishikawa, Attacks and cascades in complex networks, *Lecture Notes in Physics*, vol. 650, pp. 299–310, 2004.

[8] A. E. Motter, T. Nishikawa, and Y.-C. Lai, Range-based attack on edges in scale-free networks: Are long-range edges responsible for the small-world phenomenon?, *Physical Review E*, vol. 66, pp. 065103, 2002.

[9] F. P. Tsen, T. Y. Sung, and M. Y. Lin, Finding the most vital edges with respect to the number of spanning trees, *IEEE Transactions on Reliability*, vol. 43, no. 4, pp. 600–603, 1994.

[10] V. V. B Rao, Most-vital edge of a graph with respect to spanning trees, *IEEE Transactions on Reliability*, vol. 47, no. 1, pp. 6–7, 1998.

[11] E. Nardelli, G. Proietti, and P. Widmayer, Finding the detour-critical edge of a shortest path between two nodes, *Information Processing Letters*, vol. 67, no.1, pp. 51–54, 1998.

[12] E. Nardelli, G. Proietti, and P. Widmayer, A faster computation of the most vital edge of a shortest path, *Information Processing Letters*, vol. 79, no.2, pp. 81–85, 2001.

[13] W. H. Ren, and Z. M. Lu, Link importance evaluation based on gray relational analysis for communication networks, *Journal of Network Intelligence*, vol. 1, no. 1, pp. 38–45, 2016.