7th International Symposium on Reliability Engineering and Risk Management (ISRERM2020)
Jun.5-8, 2020, Beijing, China
Lu, Z.H., Zhao, Y.G. & Ji, X.W. (editors)

Centrality-based selective recursive decomposition algorithm for efficient network reliability analysis

Dongkyu Lee[1] and Junho Song[2]

[1] *Graduate Student, Dept. of Civil and Environmental Engineering, Seoul National University, Seoul, Republic of Korea.*
E-mail: idongkr@snu.ac.kr
[2] *Professor, Dept. of Civil and Environmental Engineering, Seoul National University, Seoul, Republic of Korea.*
E-mail: junhosong@snu.ac.kr

It is imperative to evaluate the reliability of infrastructure networks such as transportation networks and water supply systems, which play a central role in maintaining urban functions. System reliability analysis of real lifeline networks generally faces various technical challenges, which are caused by a large number of components, complex network topology, and interdependency between component failures. An existing non-simulation-based algorithm, the selective recursive decomposition algorithm (S-RDA) has improved the original RDA, but still has limitations in terms of computational time and memory required for large-scale complex networks. In this paper, a new approach utilizing a network centrality index, termed a centrality-based selective recursive decomposition algorithm (CS-RDA), is developed. By decomposing subgraphs containing critical components identified based on centrality preferentially, CS-RDA can further reduce the number of the subgraphs required to achieve the target bound width and enables us to handle more complex networks without increasing the computational costs significantly. It is also noted that one can further improve the efficiency of the algorithm by assigning more weights to the origin and destination nodes when they have low values of betweenness centrality in a network. The efficiency of the proposed approach is demonstrated by numerical examples of a hypothetical network and a real transportation network.

Keywords: system reliability, infrastructure network, non-simulation-based approach, probability bounds, network centrality.

1 Introduction

As cities become denser and more complicated, modern society is highly dependent on the reliability of urban lifeline networks such as transportation, electricity, and gas networks. When major disasters such as earthquakes and typhoons occur, the performance of the lifeline network becomes far more critical; the damaged lifeline networks can disrupt prompt response, and lead to socio-economic losses caused by a delayed recovery. Therefore, it is essential to rapidly evaluate the system failure probability, e.g., the probability of disconnection to identify appropriate follow-up measures.

However, system reliability analysis of infrastructure networks is very intricate because of an excessive number of components, complex network topology, and complicated interdependence between each component. To overcome these difficulties and analyze system reliability promptly but accurately, many simulation-based approaches that have straightforward applicability and high flexibility have been proposed. These sampling methods however have an intrinsic limitation: a tremendous amount of computational time is required to achieve a statistically significant level of convergence for rare events. In addition, these approaches make it nearly impossible to perform various probabilistic inferences or to quantify the contribution of each component to system failure events using the analysis result.

To address these limitations of simulation-based approaches, various non-simulation-based approaches have been developed and utilized. For example, Li and He (2002) proposed an algorithm to decompose a network recursively. The recursive decomposition algorithm (RDA) identifies cut sets (i.e., sets of network components whose joint failures cause the disconnection of the given origin-destination (O/D) pair) and link sets (i.e., sets of network components whose joint survivals ensure the connectivity of the given O/D pair) systematically. RDA identifies these link and cut sets such that they are mutually exclusive to each other, which results in a great advantage in calculating system reliability.

Lim and Song (2012) proposed a method termed as a 'selective recursive decomposition algorithm' (S-RDA), which improved the RDA by prioritizing *critical* sets that have the most likelihood of connection or disconnection of the given network. Since the probabilities of the identified link and cut sets using S-RDA are much higher than those identified by the original RDA, the convergence speed of the upper and lower bounds is significantly enhanced and the computational cost with respect to time and memory is saved. This improved efficiency facilitates analyzing large-scale lifeline networks that cannot be handled with existing algorithms. Furthermore, Lim and Song (2012) showed how inter- and intra-event uncertainties in spatially correlated ground motions can be incorporated into network reliability analysis using S-RDA. It was found that the assumption of statistical independence that disregards spatial correlations can overestimate or underestimate system reliability significantly.

Although the performance of S-RDA is much more efficient than RDA, there still exist some computational limitations for direct application to real urban lifeline networks. In addition, the calculation considering spatial correlations has a trade-off of the accuracy against the computational complexity. Consequently, the reduction in total calculation time is negligible. For meaningful calculation time and memory savings and fast convergence of bounds of network reliability, utilizing network centrality of components can be used as a key factor in identifying critical sets and components. In some past research efforts regarding the relationship between network robustness and centrality, critical components in terms of network connectivity were detected based on *betweenness centrality* (Carvalho et al. 2009, Lordan et al. 2014). However, these results were based on the assumption that the failure probabilities of all components are the same, so a measure considering the reliability of individual components as well as betweenness centrality is necessary to gain general applicability of such an algorithm.

In this paper, a new network reliability analysis method termed a 'centrality-based selective recursive decomposition algorithm' (CS-RDA) is proposed. To identify critical components that contribute to the connection between the origin and destination nodes, we consider some network centrality concepts including betweenness centrality. Moreover, several alternatives are presented to compensate for the blind spot of the methodology based on network centrality. The performance of the proposed CS-RDA is demonstrated by numerical examples and compared with that of S-RDA.

2 Review of Existing Recursive Decomposition Algorithms

Using graph theory, let us consider a lifeline network as a graph $G = (N, A)$ where N and A respectively denote the sets of the nodes and arcs in the graph. In general, graphs can be classified into three categories in terms of weight assignment: (1) node weight graph, i.e., only nodes are assigned weights; (2) arc weight graph, i.e., only arcs are assigned weights; and (3) general weight graph, i.e., both nodes and arcs are assigned weights (Li and He 2002). A lifeline network, meanwhile, consists of both line-type (pipelines and electricity wires) and node-type components (stations and power plants). At first glance, it seems that only the general weight graph can handle the failure probabilities of both types of components. However, adopting an 'arc-weighted' approach in which every component is considered as a node regardless of the component types and the reliability of each component is reflected in the weight of neighboring arcs, we can transform any type of graphs into arc weight graphs. Thus, the node set N consists of the nodes representing node-type elements and the 'link nodes' that are introduced to indicate the states of line-type elements (Lim and Song 2012). The arc set A is made up of arcs whose weights are equal to the reliability of the arrival nodes. Reflecting the reliability of nodes in this way does not consider the reliability of the origin node, but it does not affect the result because the origin node is contained in all paths.

Suppose after a natural or man-made disaster occurs, each component can have binary states: operative or failed. Likewise, an O/D pair can be disconnected from each other or remain connected. To express the connectivity of the O/D pair and the states of individual components, a 'structure function' $\Psi(G)$ and 'node functions' a_i, $i = 1, \ldots, N$, i.e., Bernoulli random variables representing the connectivity of a graph G and the state of i^{th} component respectively, are introduced as follows:

$$\Psi(G) = \begin{cases} 1, & \text{if O/D pair is connected} \\ 0, & \text{if O/D pair is disconnected} \end{cases} \tag{1}$$

$$a_i = \begin{cases} 1, & \text{if } i^{\text{th}} \text{ component operates} \\ 0, & \text{if } i^{\text{th}} \text{ component fails} \end{cases} \tag{2}$$

Consider the shortest path from origin to destination node that is represented as $A_1 = a_1 a_2 \ldots a_n$. Then according to the Boolean laws, the structure function $\Psi(G)$ is expressed as a linear function of node and structure functions of subgraphs in ascending order of component numbering, i.e.,

$$\begin{aligned}\Psi(G) = a_1 a_2 \ldots a_n + \bar{a}_1 \Psi(G_1) + a_1 \bar{a}_2 \Psi(G_2) + \cdots + a_1 a_2 \cdots a_{i-1} \bar{a}_i \Psi(G_i) + \cdots \\ + a_1 a_2 \cdots a_{n-1} \bar{a}_n \Psi(G_n)\end{aligned} \tag{3}$$

where G_i represents the subgraph of G, obtained by removing i^{th} node from G. If G_i still has the shortest path from origin to destination node, $\Psi(G_i)$ is recursively expanded in the same way as mentioned in Equation (3). In S-RDA, to achieve faster convergence of the bounds on network failure probability, critical disjoint link sets are found by finding the most reliable paths instead of the shortest paths. Furthermore, when choosing a subgraph to be expanded, those with the highest likelihood of being cut sets are identified preferentially, not following node numbering. These differences help S-RDA converge faster than the original RDA.

Using the structure function mentioned above, system reliability can be expressed as a probability about the structure function. If all disjoint link sets between the origin and the destination node, $L_i, i = 1, \ldots, N_L$, are identified by an algorithm, the system reliability R, can be calculated by summing up their probabilities, that is,

$$R = P[\Psi(G) = 1] = P\left(\bigcup_{i=1}^{N_L} L_i\right) = \sum_{i=1}^{N_L} P(L_i) \tag{4}$$

The system failure probability P_f, can be obtained in the same way using all disjoint cut sets between the origin and the destination node, C_i, $i = 1, \dots, N_C$, that is,

$$P_f = P[\Psi(G) = 0] = P\left(\bigcup_{i=1}^{N_C} C_i\right) = \sum_{i=1}^{N_C} P(C_i) \quad (5)$$

However, it is impossible to identify all disjoint link and cut sets for a large-scale network. In these cases, one can alternatively calculate the upper and lower bound of the system failure probability. In this case, both identified link and cut sets are utilized to obtain the information about the bounds of system reliability (Figure 1(a)) or system failure probability (Figure 1(b)), that is,

$$\begin{gathered} \sum_{i=1}^{n_L} P(L_i) \leq R \leq 1 - \sum_{i=1}^{n_C} P(C_i) \\ \sum_{i=1}^{n_C} P(C_i) \leq P_f \leq 1 - \sum_{i=1}^{n_L} P(L_i) \end{gathered} \quad (6)$$

where n_L and n_c are the numbers of identified disjoint link and cut sets.

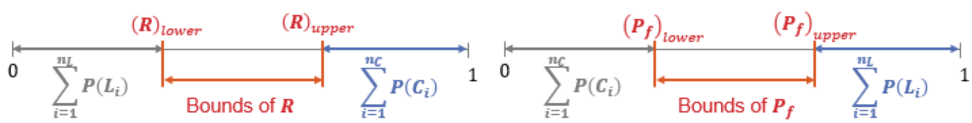


Figure 1. The upper and lower bounds of (a) system reliability R, and (b) system failure probability P_f

3 Importance Measures in terms of Network Topology

For selective decomposition of networks, which would improve the efficiency of network reliability analysis, we prioritize the nodes which are most likely to be included in critical cut sets of the network. Critical nodes are appraised in terms of network topology to achieve this goal. Various network centralities have been developed to quantify importance of nodes in a network: influential people in social networks, key nodes in the Internet, and superspreaders of diseases (Özgür et al. 2008). In this paper, we will focus on the three most common indices: degree centrality, closeness centrality, and betweenness centrality.

3.1 *Degree centrality*

The simplest index of the network centrality is degree centrality, which is defined as the degree (i.e., the number of neighboring nodes) of each node. This is based on the assumption that the more nodes a given node is connected to, the greater the impact on the whole network. Using the adjacency matrix A, which takes 1 if there is an arc from node i to node j, and 0 otherwise, the degree centrality of i^{th} node, $C_D(i)$, is calculated as

$$C_D(i) = \sum_{j \neq i} A_{ij} \quad (7)$$

3.2 *Closeness centrality*

The closeness of a node is based on the average distance of the shortest path between the node and the others in the network. On this account, the distance of all arcs in the network should be defined to use closeness centrality. Unlike degree centrality introduced above, the direction of arcs can produce completely different results, e.g., a famous person has a high closeness centrality from incoming arcs, but a low closeness centrality from outgoing arcs. The closeness centrality of i^{th} node, $C_C(i)$, is defined as

$$C_C(i) = \frac{n-1}{\sum_{j \neq i} d_{ij}} \tag{8}$$

where d_{ij} represents the shortest distance from node i to node j, and n represents the number of nodes in the given network.

3.3 *Betweenness centrality*

Betweeness centrality, another centrality measure defined based on the shortest path, measures the centrality of a node by how many of the shortest paths pass through the node. The calculation process is quite similar to that of closeness centrality, but the difference is that while closeness centrality concentrates on the length of the shortest path, betweenness centrality focuses on the number of the shortest paths via the given node. In detail, the betweenness centrality of i^{th} node, $C_B(i)$, is given as

$$C_B(i) = \frac{1}{(n-1)(n-2)} \sum_{t \neq s,i} \sum_{s \neq i} \frac{v_i(s,t)}{v(s,t)} \tag{9}$$

where $v(s,t)$ is the number of shortest paths from node s to node t, and $v_i(s,t)$ is the number of those passing through node i.

4 Centrality-based Selective Recursive Decomposition Algorithm

In the aforementioned S-RDA, identifying the subgraph with the highest probability makes the convergence of the bounds much faster than the original RDA. In addition to identifying critical disjoint sets that contributes significantly to system reliability, it is also important to identify critical components that are likely to belong to the minimum cut set. Albert et al. (2000) showed that the robustness of a network decreases dramatically when key components with high centrality (e.g., degree centrality, closeness centrality, or betweenness centrality) malfunction. It is known that betweenness centrality is most closely related to network robustness among various centrality indices (Iyer et al. 2013).

In this paper, a new algorithm utilizing betweenness centrality, termed a centrality-based selective recursive decomposition algorithm (CS-RDA), is proposed. However, there is a fatal drawback to using betweenness centrality in a straightforward manner. Because the centrality depends on network topology, decompositions based only on betweenness centrality can lead to inefficient results in terms of system reliability analysis. The problem worsens especially for nodes consisting of routes have high centrality but low reliability. The likelihood of these identified disjoint sets is so small that these sets rarely contribute to the convergence of bounds. In the proposed algorithm, centrality reflects both the reliability of components and network topology to prevent such incompetent cases. An arc weight graph $G = (N, A)$ representing the target lifeline network is composed of the node set N including node-type elements and link-type elements and the arc set A containing the information about the network topology and the failure probability of the arrival nodes. Since the arc weights are for comparting the relative reliability of each path, the omission of the failure probability of the origin node does not affect the result. In the graph constructed in this way, the most reliable paths are identified first as S-RDA. When arranging nodes in the identified paths, they are sorted in descending order of centrality, not following component numbering choice. This simple ordering change makes the nodes with a high likelihood belong to cut sets to be removed first, resulting in huge computational savings as fewer subgraphs are created.

However, if the centralities of the origin and destination nodes are too low, the performance of CS-RDA is almost no different from that of S-RDA. In these cases, one can achieve the same goal through fewer decompositions by assigning more weights to the origin and destination nodes, named *weighted* CS-RDA or wCS-RDA. In this process, the property of betweenness centrality, i.e., each shortest path is taken into account in measuring the centrality of all nodes, play an important role because we do not need to weight all nodes.

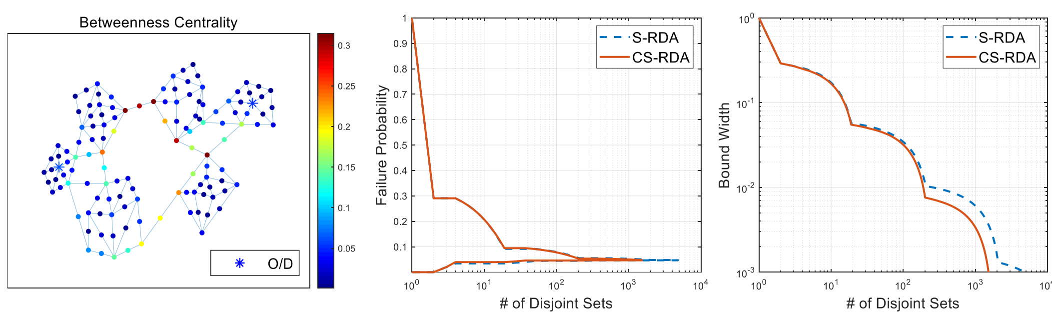
5 Numerical Examples

In this section, two numerical examples are used to demonstrate the improvement by CS-RDA. To compare the efficiency of each algorithm, the number of disjoint sets and computational times required to achieve the target bound width of 0.1% and 0.001% respectively are presented. The first example deals with a hypothetical network in Figure 2(a) with 42 node-type elements and 85 link-type elements. The failure event of the i^{th} component and its probability is formulated as

$$E_i = \{z_i \leq 0\} = \{R_i - S_i \leq 0\} \tag{10}$$

$$P_{f_i} = \Phi(-E[z_i]/\sigma[z_i]) \tag{11}$$

where R_i represents the capacity of the i^{th} component which follows the normal distribution with mean 1.5 and standard deviation 0.4, S_i represents the demand of the i^{th} component which follows the normal distribution with mean 0.48 and standard deviation 0.3, and z_i is the difference between R_i and S_i. Capacities and demands are assumed to be statistically independent, respectively. The second example deals with a real highway network in eastern Massachusetts in Figure 3(a) with 74 node-type elements only, whose failure probabilities are computed in the same manner as the previous example. In the second example, we compare the results of the O-D pair with and without weights to check the effect of weights.



(a) (b) (c)

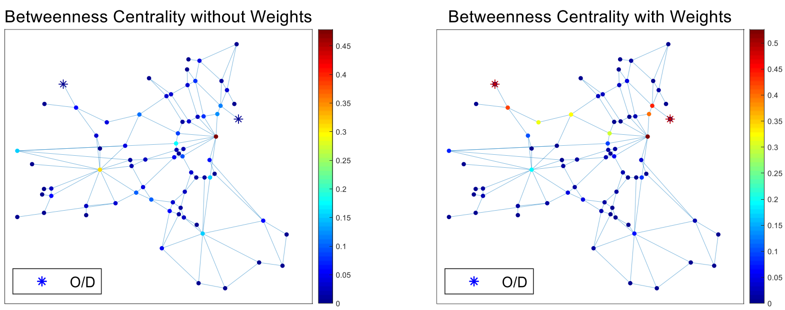
Figure 2. (a) Betweenness centrality of a hypothetical network, (b) the bounds on the system failure probability, and (c) the bound widths in terms of the number of disjoint sets (logarithmic scale)

5.1 *Hypothetical network*

Figures 2(b) and 2(c) show the bounds on the system failure probability and bound widths obtained by S-RDA and CS-RDA. As the number of disjoint sets increases, the difference between bound widths obtained by S-RDA and CS-RDA increases. In other words, a much smaller number of disjoint sets are needed to achieve a target bound width of 0.1% when using CS-RDA. Table 1 shows the bounds on system failure probability, the number of disjoint sets and computational time for each algorithm. The number of disjoint sets and computational time needed for CS-RDA to achieve the target bound width is about 1/3 of that needed for S-RDA. We can simply estimate the system failure probability using Monte Carlo simulations, but its computational time is quite long in comparison with other non-simulation methods.

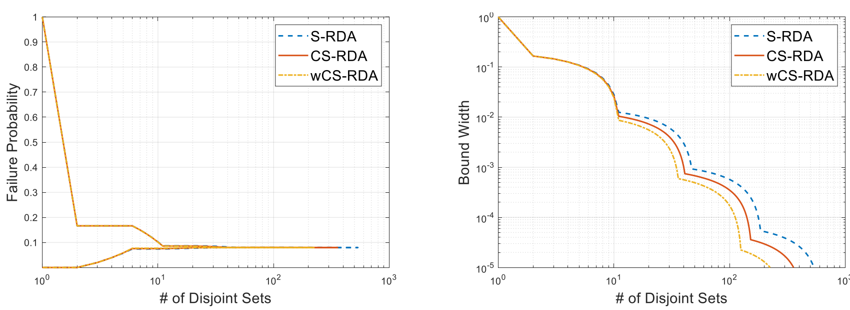
Table 1. The results of system reliability analysis (SRA) for a hypothetical network

SRA Algorithm	Bounds on the system failure probability (P_f)	# of disjoint sets	Time (s)
M.C.S. ($\delta = 0.1\%$)	0.0469	-	2032
S-RDA	0.0468~0.0478	4,828	6.276
CS-RDA	0.0467~0.0477	1,520	2.128



(a) (b)

Figure 3. (a) Betweenness centrality of Eastern Massachusetss highway network without weights, and (b) with weights to the origin and destination nodes



(a) (b)

Figure 4. (a) The bounds on the system failure probability, and (b) the bound widths in terms of the number of disjoint sets (logarithmic scale) in Eastern Massachusetss highway network

5.2 *Eastern Massachusetts highway network*

In this network example, the betweenness centralities of the nodes located on the most reliable path between the given O-D pair is relatively low as seen in Figure 3(a). In this case, wCS-RDA works more effectively as mentioned above. Figure 3(b) shows that betweenness centralities assigning more weights to the origin and destination nodes for wCS-RDA. Table 2 shows the bounds on system failure probability, the number of disjoint sets and computational time for each algorithm. Figures 4(a) and 4(b) visualize these results. Although the target bound width is much smaller than that of the hypothetical network, fewer disjoint sets are needed because the network size is small compared to the previous network. However, it is clear that wCS-RDA can improve CS-RDA even though CS-RDA still shows higher efficiency than S-RDA. Monte Carlo simulations require excessively large calculation time in this example as well.

Table 2. The results of SRA for the Eastern Massachusetts highway network

SRA Algorithm	Bounds on the system failure probability (P_f)	# of disjoint sets	Time (s)
M.C.S. ($\delta = 0.1\%$)	0.07963	-	1156
S-RDA	0.07961~0.07962	537	0.6198
CS-RDA	0.07961~0.07962	358	0.4634
wCS-RDA	0.07961~0.07962	224	0.3704

6 Conclusions

In this paper, a centrality-based selective recursive decomposition algorithm (CS-RDA) is developed for efficient reliability analysis of large-scale networks in real life, which cannot be handled by existing non-simulation-based algorithms. For this purpose, the number of components that make up disjoint sets is minimized. As the lengths of disjoint sets are minimized, not only the expected value of the number of subgraphs decomposed reduces, but also the probability of each disjoint set increases. To eliminate the redundancy of these sets, nodes on the most reliable paths are sorted in descending order of betweenness centrality. Moreover, the efficiency can be further improved by assigning more weights to the origin and destination nodes especially when these nodes have low centralities. The computational cost by CS-RDA allows for more detailed network reliability analysis and expand the range of large-scale networks that can be addressed by non-simulation-based approaches. The numerical examples successfully demonstrate the efficiency of the proposed methodology.

Acknowledgments

This research was supported by a grant (20SCIP-B146946-03) from Smart Civil Infrastructure Research Program funded by Ministry of Land, Infrastructure and Transport of Korean government.

References

Albert, R., Jeong, H., and Barabási, A. L., Error and attack tolerance of complex networks, *nature*, 406(6794), 378, 2000.

Carvalho, R., Buzna, L., Bono, F., Gutiérrez, E., Just, W. and Arrowsmith, D., Robustness of trans-European gas networks, *Physical review E*, 80(1), 016106, 2009.

Iyer, S., Killingback, T., Sundaram, B., and Wang, Z., Attack robustness and centrality of complex networks, *PloS one*, 8(4), e59613, 2013.

Li, J. and He, J., A recursive decomposition algorithm for network seismic reliability evaluation, *Earthquake engineering & structural dynamics*, 31(8), 1525-1539, 2002.

Lim, H. W. and Song, J., Efficient risk assessment of lifeline networks under spatially correlated ground motions using selective recursive decomposition algorithm, *Earthquake Engineering & Structural Dynamics*, 41(13), 1861-1882, 2012.

Lordan, O., Sallan, J. M., Simo, P. and Gonzalez-Prieto, D., Robustness of the air transport network, *Transportation Research Part E: Logistics and Transportation Review*, 68, 155-163, 2014.

Özgür, A., Vu, T., Erkan, G., and Radev, D. R., Identifying gene-disease associations using centrality on a literature mined gene-interaction network, *Bioinformatics*, 24(13), i277-i285, 2008.