ON IDENTIFYING THE CRITICAL NODES AND VULNERABLE EDGES FOR INCREASING NETWORK SECURITY

Mohamed Alshaer
University of the District of Columbia
School of Engineering and Applied Sciences
Department of Electrical and Computer Engineering

Paul Cotae
University of the District of Columbia
School of Engineering and Applied Sciences
Department of Electrical and Computer Engineering

Abstract
The recent increase in internet-related crimes and sophistication of hacker techniques has shown the importance of introducing new topics and researches on Cybersecurity which will help professionals reduce the probability of attacks. In this paper, we focus on the computer network security by finding the vulnerable components of the networks which are very critical for protecting infrastructure and network system performances. In a typical attack case, an attacker would first exploit the weak elements on a network, and then he/she only needs to target some critical edges or nodes. As a case of study, we choose which nodes or edges would be attacked on a social network by using four centrality measures such as degree, closeness, eigenvector, and betweenness centrality. We attempt to characterize and provide insights into the topology of the networks and collaborations within engineering education research. We balance all the above concepts by using similar ideas as in Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) which is a Multi-Criteria Decision Making (MCDM) technique focusing on node importance and asset criticality. This technique will help us to demonstrate and distinguish critical nodes which will help network and security professionals make the right decision.

Keywords: Centrality measures, vulnerability, social network, TOPSIS

Introduction

Due to the expansion of engineering education over the past decade, it led the field to a critical stage that demands new tools and methods to enable the community to expand and build on prior work. To expand the applicability of cybersecurity, the field needs to enlarge and improve its workforce. Currently, professionals in this area of study are either researchers or workers that accumulate several specialization courses, instead of a degree that focuses primarily on network security; thus, it reflects on the number of young students who have interest in academic support to pursue degrees in this path.

Social Network Analysis (SNA) has already proven to be an effective technique to analyze interaction in a network topology. Using SNA, the nodes (clusters) and ties (relationships) in networks can be visualized and analyzed using quantitative measures and graphical representations to examine the flow of interactions. The goal of this paper is to demonstrate how SNA can be used to study importance of nodes or groups. By implementing the centrality measure technique, we will have better understanding which nodes or groups are more important and vital than the other within a network.

The foundation of many complex networks, from biological and social network to the internet including the telecommunication and computer network, can be portrayed as a graph. The ability for analyzing networks to identify important nodes and discover hidden structure has led to important scientific and technological breakthroughs. Network analysis algorithms are used to discover communities of like-minded individuals, detect influential people and blogs, rank scientists and find important scientific papers. Centrality determines node's importance in a network. This measure is dependent on the network structure or network topology.

The increasing frequency, size and sophistication of cyber-attacks have produced a lot of research focusing on enhancing the network performance. Specifically, models and metrics that are designed to obtain an understanding of the vulnerability of enterprise networks have recently received a lot of attention. Depending on the definition of centrality, a node with high centrality is the one that can behave like a broker in the network, or that can reach the other individuals with the minimum number of "hops." In networks, centrality can be used to study the robustness of the network topology or to identify the nodes in the network that represent points of failure. This kind of analysis can be carried on off-line, once the network topology is known, as social scientists do on social networks and it is helpful to identify network nodes or links that need to be treated with special attention by the network manager.

Critical nodes in complex systems need to be identified for protection or removal. Removal of critical nodes decreases or minimizes a system's ability to diffuse entities such as information, goods, or diseases. Previous research suggested some vulnerability metrics, but there remains a lack of understanding how a metric change (e.g., upper bound and lower bound) and how it is related to the structure of a complex system [1].

Background

Previous works on vulnerability modeling of enterprise networks have focused their attention on two main concepts namely direct modeling of cyber graphs and modeling with attack graphs. Pure graph-based approaches are normally grounded in applying concepts and metrics such as reachability, shortest paths and other modes of centrality computations to analyze vulnerabilities in a given network [1].

A number of pure graph-based approaches suffer from high computational complexity issues because of the usage of a large perimeter in the form of the Internet connected hosts and servers [1]. Furthermore, pure graph-based approaches work based on topology without considering the underlying mechanics and the associated probabilities, that when included presents a more realistic picture.

The study of a network's vulnerabilities and its response to malfunction helps engineers design robust systems [2] and scientists understand complex phenomena such as neuro dysfunction [3], economic and financial risk [3], and disease spreading [3].

Methods for robustness analysis generally assume that the system is represented by a network composed of static links, focusing on the topological properties of a network that is vulnerable. Depending on the system and research question, a static representation may also incorporate weighted and directed edges, allowing richer dynamics to be modelled. However, static shortest paths miss the vital time order of links which result in the underestimation of the correct shortest path. Thus, the key contribution of this paper is the introduction of temporal centrality metrics for the identification of key nodes in temporal graphs based on temporal shortest paths. Often the topology of a network has distinctive features, such as

vertex order distribution, clustering, and characteristic path length, which can be explained in terms of its evolution and which in turn explain some aspects of its behavior.

Naturally, both these temporal extensions are associated to the identification of central nodes in the network with application to dynamic processes over a real network. In many systems, however, edges are not continuously active and the quantities their weights represent may vary with time. Furthermore, these time-varying systems may also be spatially embedded, and thus the capability for nodes to interact is controlled by the space in which they operate as well as their network connectivity.

Centrality measures aim at quantifying how important an element of the network is by relying only on the structural pattern of the network. The vertex centrality measures have been used by many works in different areas, including: strategic network formation [4], game theory [4], social behavior, transportation, influence and marketing [4], communication, scientific citation and collaboration, communities, group problem-solving. The goal of this paper is to enable and encourage researchers interested in cyber security and education research more generally, to perform analyses that use relational data and consider the importance of learning relationships to undergraduate education.

Method

The MCDM problems can be divided into two kinds. One is the classical MCDM set of when ratings and the weights of criteria are measured in single values. Another one is the multiple criteria decision-making set of problems where the ratings and the weights of criteria evaluated on missing or incomplete information, imprecision, subjective judgment and vagueness are usually expressed by interval numbers, numbers or fuzzy numbers.

In the classical TOPSIS method we assume that the ratings of alternatives and weights are represented by numerical data and the problem is solved by a single decision maker. Complexity arises when there are more than one decision makers because the preferred solution must be agreed on by interest groups who usually have different goals. The classical TOPSIS algorithm for a single decision maker and for group decision making is systematically described below.

1) We start with a $m \times n$ decision matrix $= d_{mxn}$, where the rows $i = 1, \ldots, m$ are the alternatives (e.g. nodes), and the columns $j = 1, \ldots, n$ are the criteria (e.g. centrality measures, node asset values). The alternatives are $\{A_i\}$, and the criteria are $\{K_j\}$. The set of criteria that is indexed by j, consists of both "benefit" criteria K^+ and "cost" criteria K^-. If there are no cost criteria, the TOPSIS algorithm simplifies.

2) We next form a normalized decision matrix $D = \delta_{ij}$ by normalizing every entry in D_j, the jth column of D, by its column norm $\| D_j \|$.

$$\delta_{ij} = \frac{d_{ij}}{\sqrt{\sum_{i=1}^{m} (d_{ij})^2}} , \quad j = 1,2,3 \ldots, m \; ; \quad j = 1,2,3 \ldots, n$$

3) We now assign a weight of $0 \leq w_j$ to each column and form the weighted normalized decision matrix:

$$V = v_{ij} \;\; m \times n \qquad v_{ij} = w_j \, . \, \delta_{ij} \quad i = 1, \ldots m \; ; \; j = 1, \ldots n$$

4) We next calculate the positive ideal solution $A^+$and the negative ideal solution A^-

$$A^+ = \{v^+_1, \ldots, v^+_n\} \; ; \; A^- = \{v^-_1, \ldots, v^-_n\}, \text{ where}$$

$$v_j^+ = max v_{ij} \; ; \quad v_j^- = min v_{ij}, \; j = 1 \ldots, n$$

Thus, A^+ corresponds to finding the largest value in every criteria (column), and A^- to the minimal.

5) Now we go through attribute by attribute (row) to see how the actual values compare to the positive ideal, and the negative ideal. We calculate the Positive (Negative) Separation S_i^+ (S_i^-) between each alternative and the positive ideal solution (negative ideal solution).

$$S_i^+ = \sqrt{\sum_{j=1}^{n}(v_{ij} - v_j^+)^2} \quad , \quad S_i^- = \sqrt{\sum_{j=1}^{n}(v_{ij} - v_j^-)^2} \quad \text{for } j = 1, \ldots n$$

6) We next calculate the relative closeness of alternate i to the ideal positive solution. We want to be close to the ideal positive solution and far from the ideal negative solution

$$C_i = \frac{S_i^-}{S_i^- + S_i^+}$$

When $Ci = 1$ ($Ci = 0$) alternative i is the best (worst) solution, that is it coincides with A^+ (A^-). Note that if we have a weighting wj and replace it with K wj the terms $S_i^+$and $S_i^-$are multiplied by K, but the Ci values are unchanged. In particular, if every column is equiweighted at 1, and we change the normalization to $wj = 1/n$ the Ci are unchanged.

7) Now we rank the alternatives, from highest to lowest, via the Ci value. The results are exemplified on the network structure illustrated in Fig 1 and summarized in Table 1.

Network Topology

Social Network Analysis Using R software teaches analysts how to visualize and analyze data from a social network like Twitter or Facebook with the text-based statistical language. It provides some useful R code examples on:

- directed and undirected graphs
- creating regular graphs, including full graphs, stars, rings, lattices and trees
- creating graphs from real-world data
- various random graphs
- importing and exporting graphs in various formats, such as edge list files and Pajek format
- vertex and edge sequences and their indexing; and network flows and minimum cuts.

SNA measures allow the analyst to gather more information and data and put more effort into the parts of a network that require further attention. Here's a brief summery on centrality measures that have been used on this paper:

Centrality is the measure which gives a rough indication of the social power of a node based on how well they "connect" the network. "Betweenness," "Closeness," "Degree," and "Eigenvector" are all measures of centrality.

Degree centrality assigns an importance score based purely on the number of links held by each node.

> **What it tells us:** How many direct, 'one hop' connections each node has to other nodes within the network.
>
> **When to use it:** For finding very connected individuals, popular individuals, individuals who are likely to hold most information or individuals who can quickly connect with the wider network.

Betweenness centrality measures the number of times a node lies on the shortest path between other nodes.

> **What it tells us:** This measure shows which nodes act as 'bridges' between nodes in a network. It does this by identifying all the shortest paths and then counting how many times each node falls on one.
>
> **When to use it:** For finding the individuals who influence the flow around a system.

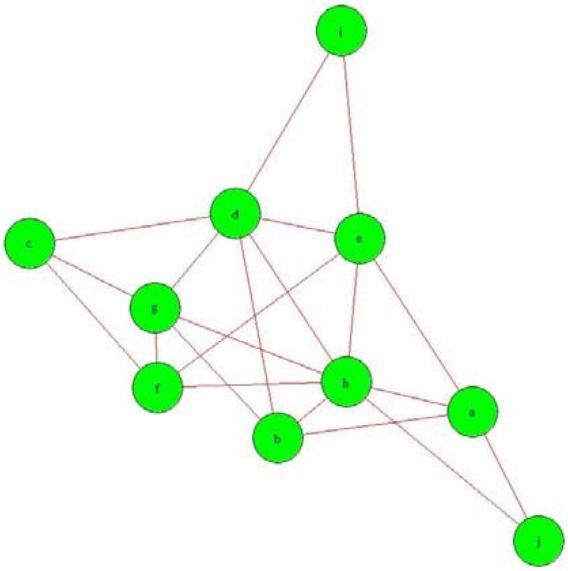

Fig. 1: Representing nodes and links.

Closeness this measure scores each node based on their "closeness" to all other nodes within the network.

> **What it tells us:** This measure calculates the shortest paths between all nodes, then assigns each node a score based on its sum of shortest paths.
>
> **When to use it:** For finding the individuals who are best placed to influence the entire network most quickly.

Eigen value Like degree centrality, Eigen Centrality measures a node's influence based on the number of links it has to other nodes within the network. Eigen Centrality then goes a step further by also taking into account how well connected a node is, and how many links their connections have, and so on through the network.

> **What it tells us:** By calculating the extended connections of a node, Eigen Centrality can identify nodes with influence over the whole network, not just those directly connected to it.
>
> **When to use it:** Eigen Centrality is a good 'all-round' SNA score, handy for understanding human social networks, but also for understanding networks like malware propagation.

We consider the example as shown in Fig. 1. This time we will use the normalizations that we discussed at the method section of the paper. we see the rank of each node (alternate) based on DC, CC, BC, or EC.

Table 1: Centrality Measures Table

#N	DC	CC	BC	EC
a=1	4.0000	0.6666	1.6666	0.5940
b=2	4.0000	0.6821	1.1190	0.6959
e=3	5.0000	0.6922	4.5357	0.7383
h=4	7.0000	0.8181	8.8452	1.0000
j=5	2.0000	0.5555	0.0000	0.3407
d=6	6.0000	0.7555	6.5952	0.8690
g=7	5.0000	0.6923	2.2023	0.7923
c=8	3.0000	0.5294	0.2551	0.4935
f=9	4.0000	0.6428	1.7857	0.6466
i=10	2.0000	0.5294	0.0000	0.3437

Table 2: Rank

Node	TOPSIS Rank
d	first
b	second
h	third
e	fourth
g	fifth
f	sixth
a	sixth
c	seventh
j	eighth
i	eighth

Results

Understanding how nodes relationship form in a network topology, as well as the impacts these relationships have on network performance and safety, can inform researchers in unique ways and improve educational reform. Social network analysis (SNA) provides the necessary tool kit for investigating questions involving relational data and the importance of a particular node. Recently, temporal networks played an important role in social network analysis due to network dynamics. Good visualization methods for time-series networks can provide better understanding on network evolution [5], thus becoming an important supplement to current social network analysis methods. For example, temporal email networks have been studied for analysis and visualization [6].

Criteria of the functions can be: benefit functions (more is better) or cost functions (less is better). When we use TOPSIS with equal weighting for each criterion which is = 0.25 for each criterion, we arrive at the above ranking. where w_j is the weight of the j-th criterion $\sum_{j=1}^{n} w_j = 1$, Which does not agree with any of the other four! Hence, TOPSIS can be successfully used when we have to balance different measures of centrality and node importance against each other.

Future Work

Measuring network centrality is an important problem for many applications. Most existing studies have focused on analyzing static networks, while in reality this assumption is not reliable since many networks are inherently dynamic; connections are added or removed over time. Our approach is in fact not very complex. Integrating the various mechanisms of vulnerability including the crucial human-user aspect would help build a less vulnerable community as well as encourage more people to look for and achieve expertise in the field. The result is a set of understandable metrics that can be analyzed that probabilistically quantify the risk either to a specific set of nodes or to the whole network.

Applying probability concept is very important in our approach to capture the mechanism of such a network in order to define our assets This includes targeted education of human users on enforcing the best practices of cyber-defense. Here we propose the topology for static structure without taking temporal effect in to consideration. We would also like to consider the temporal variation, its effect in the proposed topology and centrality evaluation; and consider a

multi-tier approach where each tier of the hierarchy will encompass a set of modules as virtual nodes and hence incorporate the dynamism [6].

Conclusions

After applying the method discussed and demonstrated above, we can find the critical values for our graph for any given topology. By setting a multi criteria matrix for decision making, we found out that the weight optimization is time sensitive, since decision making will most likely depend on temporal and dynamic networks that changes over the time. Our future work will be more focusing on weight optimization for spatiotemporal networks. More importantly, our goal is to provide members of the Engineering Education Research (EER) community with tools and infrastructure that allows them to understand the structure and networks. We have the ambitious goal of reaching a much broader range of knowledge such as simulation, data analysis and solutions.

Acknowledgments

This research was funded by the Army Research Office (ARO) - Department of Defense (DOD)– Award No. W911NF-15-1-0481.

References

[1] J. R. Johnson *et al.,* "A graph analytic metric for mitigating advanced persistent threat," *2013 IEEE International Conference on Intelligence and Security Informatics* (ISI 2013) Seattle, Washington, pp. 129– 133, June 2013.

[2] J. A. Danowski and N. Cepela, "Automatic mapping of social networks of actors from text corpora: Time series analysis," *Annals of Information Systems,* vol 12, pp.31-46, 2010, doi:10.1007/978-1-4419-6287-4_3.

[3] G. Hua, G. *et al.*, "Network analysis of US Air transportation network," *Data Mining for Social Network Data*, vol. 12, pp. 75- 89, 2010.

[4] D. Helbing, "Globally networked risks and how to respond," *Nature*, vol. 497, pp. 51–59 (doi:10.1038/nature 12047), May 2013

[5] E.H. Chi *et al.,* "Visualizing the evolution of web ecologies," Proceedings of SIGCHI Conference on Human Factors in Computing Systems*, ACM Press/Addison-Wesley,* pp. 400–407, New York,1998.

[6] S. Wen *et al.,* "Modeling propagation dynamics of social network worms," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 24, no. 8, pp. 1633– 1643, Aug 2013.

[7] C.L. Hwang *et al.,* "Multiple attribute decision making," *Methods and Applications,* Springer-Verlag, New York, 1981.

[8] P. Gloor, "Capturing team dynamics through temporal social surfaces," *Proceedings of 9th IEEE International Conference on Information Visualization* IV05, pp. 6–8, 2005.

[9] S. Opricovic and G. H. Tzeng, "Compromise solution by MCDM methods," *European Journal of Operational Research*, vol. 15, pp.445–455, 2004.

Authors

Mohamed Alshaer
Mr. Mohamed Alshaer is a Graduate Research Assistant, in the Department of Electrical and Computer Engineering, School of Engineering and Applied Sciences at the University of the District of Columbia, Washington D.C. He is interested in cybersecurity.

Paul Cotae
Dr. Paul Cotae is a Professor of Electrical and Computer Engineering in in the School of Engineering and applied Sciences (SEAS). He is the Director of SEAS Research Center at the University of the District of Columbia. His research is in Digital Communication, Information theory, Statistics and Applied Mathematics and Cybersecurity: Anomaly detection, Detection of Low Rate Denial of Service Attacks, Intrusion Detection, Information Visualization. He published more than 140 conference and journal papers, authored 2 books and coauthored three books in the area of digital communications systems. During the AY 2014-2015 he spent his sabbatical at the Center for High Assurance Computer Systems Code 5540, Naval Research Laboratory, Washington DC, 20375. Since 2009, he has been selected every summer as ONR (Office of the Naval Research) Senior Research Fellows for the ASEE Summer Faculty Research Program at NRL.