

A Metric to Compare Vulnerability of the Graphs of Different Sizes

Kushal Kanwar^{1,2}

*University Institute of Engineering and Technology
Panjab University
Chandigarh, India 160014*

Harish Kumar, Sakshi Kaushal³

*University Institute of Engineering and Technology
Panjab University
Chandigarh, India 160014*

Abstract

Node immunization and estimation of vulnerability are essential for a number of applications from safe guarding computer networks against virus attacks to protecting human population from highly contagious biological viruses. The concepts of node immunization can be used in the fast diffusion of information on a network and viral marketing. In this article, normalized eigenvalue based measure to compare vulnerability of Graphs of different sizes is proposed. The measure is based on largest eigenvalue of the adjacency matrix of the graph. Erdos-Renyi Random Graphs of different size and connectivity been used for experimental verification using susceptible infected (SI) model. It has been found that the proposed metric is good to compare vulnerability of graphs when graphs are moderately or highly connected, or when contagion strength is not too weak.

Keywords: Eigenvalue, Graph Vulnerability, Vulnerability Metric, Comparative Vulnerability.

1 Introduction

Node immunization is essential to safeguard computer networks against virus attack, to contain biological epidemic and effective viral marketing [4] [21]. All these high-end applications requires the quantification of importance of individual node or group of nodes in terms of their contribution towards vulnerability. A number of measures for individual nodes are available. Researchers have proposed some metrics to measure overall graph vulnerability using concepts of immunology and graph theory [3], [13], [20]. The largest (first) eigenvalue, λ is a good metric to judge vulnerability of a given graph $G = G(V, E)$, where V is set of vertices, and E is set of edges [20]. The λ is a single real number. Higher the value of λ more is the vulnerability of the graph. This metric can be used to compare the vulnerability of the graphs of the same size i.e. those having equal number of vertices. However, λ cannot be used to compare vulnerability of two graphs with different sizes. Hence, A metric is required which enables the comparison of vulnerability of the graphs of different size.

In this paper, a new Normalized Vulnerability Metric (*NVM*) is proposed which is based on highest eigenvalue of the adjacency matrix of the graph. Experimentally, it is shown that *NVM* can be used to estimate vulnerability of a graph under moderate conditions.

Rest of the article is organized as follows: section 2 gives related work, section 3 formally defines problem, section 4 presents experimental results and discussion, and the last section 5 concludes the paper.

2 Related Work

In this section, related work is reviewed which can be categorized in three parts: measuring the importance of single node on graph, measuring importance of a group of nodes and eigenvalue based approach for epidemic threshold.

¹ This work is supported by the Visvesvaraya PhD Scheme for Electronics & IT, MEITY, GOI.

² Email: kushalneo@gmail.com

³ Email: harisk@pu.ac.in (Harish Kumar), sakshi@pu.ac.in (Sakshi Kaushal)

Measuring importance of single node on graph: A number of measures are available which are based on (a) shortest path (b) random walks (c) shortest path and random walk (d) PageRank (e) HITS (f) coreness (g) abnormality score (h) articulation points (i) k -vertex cut etc. These measures use topological properties of the graphs [9], [10], [11], [13], [14], [15].

Bonachi et al, have surveyed different definitions of centrality measures and algorithms to compute them. Exact algorithms, approximation algorithms, sampling-based algorithm and highly-scalable MapReduce algorithms for large graphs are discussed [1]. Moskowitz et al, have suggested that if network topology is not fixed then centrality measures should be viewed probabilistically. It is shown that study of transmission behaviour based only on graphs topological and degree based properties are insufficient to model network propagation [12]. Satio et al have proposed a new centrality measure, *super-mediators*, which is *data-driven* and compared it to other measures such as betweenness, degree, closeness, etc. [17].

Measuring importance of a group of nodes: Valente and Fujimoto proposed a measure of brokerage in networks based on the strength of weak ties. It identifies the need for finding node-based measures of brokerage considering entire network structure, not just a node's local environment [19]. Everett and Valente extended work of Valente and Fujimoto using betweenness centrality, their new approach have advantages of less computational cost and more robustness to normalize the measure to control network size [8].

Chen et al proposed a *Shield-value* score to measure the importance of a group of k nodes. They used graphs highest eigenvalue as a measure of overall graph vulnerability. It uses carefully approximated effect of removal of group of nodes by using matrix perturbation theory and greedy heuristic. Chen et al also proved that *Shield-value* score is sub-modular and hence the algorithm developed by them is near optimal to find a set of nodes with highest *Shield-value* score.

Eigenvalue based approach: Finding the essence of graph with fewest parameters is a challenge, it is more difficult because these parameters are intrinsic to graph itself. Wang et al proposed the use of largest eigenvalue as a close measure for epidemic threshold, and proved that for scale free graphs epidemic threshold is close to zero [20]. It is a very important measure because with a single real number prediction about graphs can be made. Chakrabarti et al extended work of Wang et al and proved some important theorems relating epidemic threshold and largest eigenvalue of graph [3].

3 Problem Definition

The problem of finding a comparable vulnerability measure of a graph which (a) gives a value which measures the vulnerability of a graph (b) and this value can be used to compare vulnerability across graphs irrespective of their size. The measure should be consistent with vulnerability measure $V(G) = \lambda$, used by [20] to compare vulnerability of graphs of same size. The problem is formally defined as follows:

Problem: To define a metric which can give the vulnerability measure of a graph and this metric should be comparable for different graphs irrespective of their size. It is named as Normalized Vulnerability Measure (*NVM*).

Given: A set of un-directed un-weighted connected graphs $G_1, G_2, G_3, \dots, G_n$ with corresponding adjacency matrices $A_1, A_2, A_3, \dots, A_n$ respectively.

Proposed: A single number $NVM(G_i)$ for each Graph G_i , reflecting the vulnerability of a graph and it can be used to compare vulnerability with other graphs. Mathematically,

$$NVM(G) = V(G)/(N - 1) = \lambda/(N - 1) \quad (1)$$

Where λ is the largest eigenvalue of the adjacency matrix A of the graph and $|V(G)| = N$ [18]. The intuition behind division of λ by $N - 1$ is that the most vulnerably graph of any size is a Complete Graph. The inspection of eigen-spectrum of Complete Graph of N nodes yields that the highest eigenvalue will be $N - 1$ [2]. Hence by dividing λ by $N - 1$ will restrict *NVM* in the range $[0, 1]$; *NVM* is 0 for isolated graph, and 1 for Complete Graph. To check the claim of comparability of *NVM* across graphs of different sizes, experiments of infection spread are performed using Susceptible Infected (SI) Model on E-R Random graphs with various infection rates [7], [6].

4 Experiments and Results

For the problem formulated in the previous section an experiment is designed. Erdos-Renyi (ER) random graphs of different sizes (20, 30, 40 and 50) are generated using igraph library in R [5]. The λ and *NVM* is calculated for these graphs and values are shown in Table 1. Afterwards, these graphs are grouped together which are having same or close (in the range of ± 0.02) *NVM* value. On each group SI simulation of contagion spread is performed and the time taken by contagion to infect 80% of the nodes is recorded. It was found that graphs have same or close *NVM* score are equally vulnerable irrespective of their size, in terms of time required to infect them.

Table 1

For various combinations of n and p , $G(n, p)$ is shown with corresponding largest eigenvalue λ and NVM value.

$G(n, p)$ $n \rightarrow$ $p \downarrow$	20		30		40		50	
	λ	<i>NVM</i>	λ	<i>NVM</i>	λ	<i>NVM</i>	λ	<i>NVM</i>
0.1	2.94	0.15	4.17	0.14	5.63	0.14	5.68	0.12
0.2	4.27	0.22	7.22	0.25	8.33	0.21	10.4	0.21
0.3	5.94	0.31	9.9	0.34	11.48	0.29	15.63	0.32
0.4	7.27	0.38	11.85	0.41	15.88	0.41	20.35	0.42
0.5	9.85	0.52	13.52	0.47	20.38	0.52	25.99	0.53
0.6	12	0.63	17.74	0.61	23.55	0.6	29.36	0.6
0.7	13.78	0.73	20.43	0.7	27.62	0.71	34.65	0.71
0.8	16.36	0.86	23.35	0.81	31.41	0.81	39.58	0.81
0.9	17.77	0.94	26.23	0.9	35.49	0.91	44.35	0.91
1	19	1	29	1	39	1	49	1

Given a set of graph(s) and rate of infection α , a vulnerability measure should be able to predict the level of spread i.e. whether contagion will die or there will be an epidemic or endemic equilibrium. Various kinds of contagion spread models like Susceptible-Infected (SI), Susceptible-Infected-Susceptible (SIS), Susceptible-Infected-Recovered (SIR) etc. are available to understand the dynamics of a spread process on a network [16]. In this work, the SI model was programmed in R to develop the simulation using various infection rates. It should be noted that the simulation time is the number of steps to spread the contagion to desired number of nodes, not the actual physical time.

From Table 1, five sets of graphs which have same or very close NVM score are chosen as shown in Table 2. Sets are named A through E for sake of ease. Each set contains three graphs of different size but same level of connectivity (equal p value).

On each set A through E, a SI infection simulation was run on them with different infection rates. The average time to infect 80% nodes was recorded

Table 2
Set of graphs with similar NVM score

Sr. No	Set of Graphs	NVM	Set Name
1	G(20, 0.3), G(40, 0.3), G(50, 0.3)	(0.31,0.29,0.32)	A
2	G(30, 0.4), G(40, 0.4), G(50, 0.4)	(0.41,0.41,0.42)	B
3	G(30, 0.6), G(40, 0.6), G(50, 0.6)	(0.61,0.6,.06)	C
4	G(30, 0.9), G(40, 0.9), G(50, 0.9)	(0.9,0.91,0.91)	D
5	G(30, 1.0), G(40, 1.0), G(50, 1.0)	(1,1,1)	E

as shown in Table 3. For each set A-E simulation run was performed with infection rates α mentioned in the Table 3. In Table 2, each set A-E contains three graphs, hence the each entry in Table 3 under *Average Time of Infection* refers to the infection time for each element in respective order.

The results of Table 3 can be summarized as follows:

- (a) In Table 2, set A and B contain less dense graphs ($p = 0.3, 0.4$) as compared to set C, D and E ($p \geq 0.6$). So for low rate of infection $\alpha = 0.02$ and $\alpha = 0.05$, the time required to infect 80% of the nodes is more and have high variance as compared to graphs with high connectivity. But for $\alpha \geq 0.08$, the values shows less variance and magnitude.
- (b) For sets C and D, all values for time of infection is small and coherent, except for $\alpha = 0.02$. It is because graphs are well connected.
- (c) For set E (complete graphs), all the time to infection shows small and nearly same values except for $\alpha = 0.02$.

From above mentioned observation, it can be concluded, that if graphs have same or comparable *NVM* score then for certain threshold infection rate (here in our case $\alpha = 0.02$) the time to infect most of the graph (in our case 80% nodes) is nearly same.

Table 3
Infection rates for sets A to E for different values of α

Set Name	Average Time of Infection				
	$\alpha = 0.02$	$\alpha = .05$	$\alpha = .08$	$\alpha = .1$	$\alpha = .2$
A	(31, 24, 20)	(16, 9, 8)	(11, 9, 7)	(9, 7, 6)	(6, 5, 5)
B	(20, 18, 14)	(9, 8, 7)	(7, 7, 6)	(6, 6, 5)	(5, 4, 4)
C	(14, 12, 10)	(7, 6, 5)	(6, 5, 5)	(5, 5, 4)	(4, 4, 4)
D	(10, 9, 8)	(5, 5, 4)	(5, 4, 4)	(4, 4, 4)	(3, 3, 3)
E	(9, 8, 7)	(5, 5, 4)	(4, 4, 4)	(4, 4, 4)	(3, 3, 3)

These experiments and observations validate our claim that *NVM* can be used as a metric for vulnerability of a graph and same value can be used to compare vulnerability of the graphs of different sizes.

5 Conclusion and Future Work

In this paper, a metric to compare vulnerability of graphs of different sizes is proposed, named as *NVM*. Based on experiments, it has been found that *NVM* is good measure to compare vulnerability of graphs when graph is moderately connected, or when contagion strength is not too weak.

In future this work will be further extended in following directions (a) A mathematical quantifier will be developed to quantify when two different *NVM* values can be taken as same and with how much significance (b) The experiments will be performed for large graphs and variety of degree distribution (c) The *NVM* scores will be tested with different epidemic models like SIS and SIR etc.

References

- [1] Bonchi, F., G. De Francisci Morales and M. Riondato, *Centrality measures on big graphs: Exact, approximated, and distributed algorithms*, in: *Proceedings of the 25th International Conference Companion on World Wide Web, WWW '16 Companion* (2016), pp. 1017–1020.
- [2] Brouwer, A. E. and W. H. Haemers, “Spectra of graphs,” Springer Science & Business Media, 2011.
- [3] Chakrabarti, D., Y. Wang, C. Wang, J. Leskovec and C. Faloutsos, *Epidemic thresholds in real networks*, ACM Transactions on Information and System Security (TISSEC) **10** (2008), p. 1.
- [4] Chen, C., H. Tong, B. A. Prakash, C. E. Tsourakakis, T. Eliassi-Rad, C. Faloutsos and D. H. Chau, *Node immunization on large graphs: Theory and algorithms*, IEEE Transactions on Knowledge and Data Engineering **28** (2016), pp. 113–126.
- [5] Csardi, G. and T. Nepusz, *The igraph software package for complex network research*, InterJournal, Complex Systems **1695** (2006), pp. 1–9.
- [6] Erdős, P. and A. Rényi, *On random graphs, i*, Publicationes Mathematicae (Debrecen) **6** (1959), pp. 290–297.
- [7] Erdős, P. and A. Rényi, *On the evolution of random graphs*, Publ. Math. Inst. Hungar. Acad. Sci **5** (1960), pp. 17–61.
- [8] Everett, M. G. and T. W. Valente, *Bridging, brokerage and betweenness*, Social Networks **44** (2016), pp. 202–208.
- [9] Freeman, L. C., *A set of measures of centrality based on betweenness*, Sociometry (1977), pp. 35–41.
- [10] Kleinberg, J. M., *Authoritative sources in a hyperlinked environment*, Journal of the ACM (JACM) **46** (1999), pp. 604–632.
- [11] Moody, J. and D. R. White, *Structural cohesion and embeddedness: A hierarchical concept of social groups*, American Sociological Review (2003), pp. 103–127.

- [12] Moskowitz, I. S., P. Hyden and S. Russell, *Network topology and mean infection times*, Social Network Analysis and Mining **6** (2016), pp. 1–14.
- [13] Nagamochi, H. and T. Ibaraki, **123**, Cambridge University Press New York, 2008.
- [14] Newman, M. E., *A measure of betweenness centrality based on random walks*, Social networks **27** (2005), pp. 39–54.
- [15] Page, L., S. Brin, R. Motwani and T. Winograd, *The pagerank citation ranking: bringing order to the web* (1999).
- [16] Prakash, B. A., D. Chakrabarti, N. C. Valler, M. Faloutsos and C. Faloutsos, *Threshold conditions for arbitrary cascade models on arbitrary networks*, Knowledge and information systems **33** (2012), pp. 549–575.
- [17] Saito, K., M. Kimura, K. Ohara and H. Motoda, *Super mediator—a new centrality measure of node importance for information diffusion over social network*, Information Sciences **329** (2016), pp. 985–1000.
- [18] Strang, G., “Introduction to linear algebra,” Wellesley Cambridge Press, 2011.
- [19] Valente, T. W. and K. Fujimoto, *Bridging: locating critical connectors in a network*, Social Networks **32** (2010), pp. 212–220.
- [20] Wang, Y., D. Chakrabarti, C. Wang and C. Faloutsos, *Epidemic spreading in real networks: An eigenvalue viewpoint*, in: *Reliable Distributed Systems, 2003. Proceedings. 22nd International Symposium on*, IEEE, 2003, pp. 25–34.
- [21] Zhao, D., L. Wang, S. Li, Z. Wang, L. Wang and B. Gao, *Immunization of epidemics in multiplex networks*, PloS one **9** (2014), p. e112018.