Spectral Radius as a Measure of Variation in Node Degree for Complex Network Graphs

Natarajan Meghanathan
Department of Computer Science
Jackson State University
Jackson, MS 39217, USA
e-mail: natarajan.meghanathan@jsums.edu

Abstract— **The spectral radius of a network graph is the largest eigenvalue of the adjacency matrix of the graph. We hypothesize the spectral radius to be a measure of the variation in the degrees of the nodes. In this pursuit, we define a metric called the spectral radius ratio for node degree as the ratio of the spectral radius to the average node degree. We validate our hypothesis by determining this metric on some of the commonly studied classical large real-world complex network graphs (undirected) for network analysis. Based on the results collected, we observe the spectral radius ratio for node degree to be positively correlated (correlation coefficient: 0.75) to the coefficient of variation in node degree (the ratio of the average node degree to the standard deviation in node degree), thus confirming our hypothesis.**

Keywords-Spectral radius; eigenvalue; network graphs; node degree; correlation

I. INTRODUCTION

Network analysis and visualization of large complex real-world networks (ranging anywhere from social networks, co-authorship networks, Internet, World wide web to biological networks and etc) is an actively researched area in recent years. The strength of network analysis is to abstract the complex relationships between the members of the system in the form of a graph with nodes (comprising of the constituent members) and edges (weighted or unit-weight as well as directed or undirected, depending on the nature of the interactions) and study the characteristics of the graph with respect to one or more metrics (like node degree, diameter, clustering coefficient and etc). The adjacency matrix A(G) of the network graph G essentially captures the presence of edges between any two vertices. For any two vertices i and j in graph G, the entry in the i^{th} row and j^{th} column of A(G) is 1 if there is an edge from vertex i to vertex j and 0 otherwise.

Spectral decomposition is a method of projecting the characteristics of a network graph in n-dimensions (mutually perpendicular directions) where n is the number of vertices in the graph. The projection in each direction is represented in the form of a scalar value (called the eigenvalue) and its corresponding vector with entries for each vertex (called the eigenvector). Though the number of dimensions in the spectrum is the number of vertices in the graph, most of the variations could be captured in the first few dimensions of the coordinate system represented by the eigenvalues and the eigenvectors. The largest eigenvalue of the projection is called the principal eigenvalue and the corresponding eigenvector is called the principal eigenvector. The principal eigenvalue (also called the spectral radius) and its corresponding eigenvector capture maximum amount of variability in the data (in the case of a network graph, the data are the edges connecting the vertices). In this paper, we make use of the spectral radius of the adjacency matrix of complex network graphs to analyze the variations in node degree and correlate with the coefficient of variation of node degree.

The rest of the paper is organized as follows: In Section 2, we introduce the power-iteration method to calculate the spectral radius of a network graph as well as introduce our hypothesis relating the spectral radius with variation in node degree based on examples of toy graphs (with fewer vertices). Section 3 validates our hypothesis on some of the commonly studied classical larger network graphs. Section 4 concludes the paper and lists future work.

II. POWER ITERATION METHOD TO CALCULATE SPECTRAL RADIUS AND HYPOTHESIS

The power iteration method [1] can be used to calculate the principal eigenvalue (i.e., spectral radius) and the corresponding principal eigenvector of a graph based on its adjacency matrix. The eigenvector X_{i+1} of a network graph at the end of the $(i+1)^{th}$ iteration is given by: $X_{i+1} = X_i \,/\, \|AX_i\|$, where $\|AX_i\|$ is the normalized value of the product of the graph adjacency matrix and the tentative eigenvector X_i at the end of iteration i. The initial value of X_i is [1, 1, ..., 1], a column vector of all 1s, where the number of 1s correspond to the number of vertices in the graph. We continue the iterations until the normalized value $\|AX_{i+1}\|$ converges to that of the normalized value $\|AX_i\|$. At this juncture, column vector X_i is the principal eigenvector of the network graph and the normalized value of convergence is the spectral radius. Figure 1 illustrates an example for computation of the spectral radius on a network graph. The converged normalized value of 2.21 is the spectral radius of the graph (denoted $\lambda_{sp}(G)$).

978-1-4799-7767-3/14 $31.00 © 2014 IEEE
DOI 10.1109/UNESST.2014.8

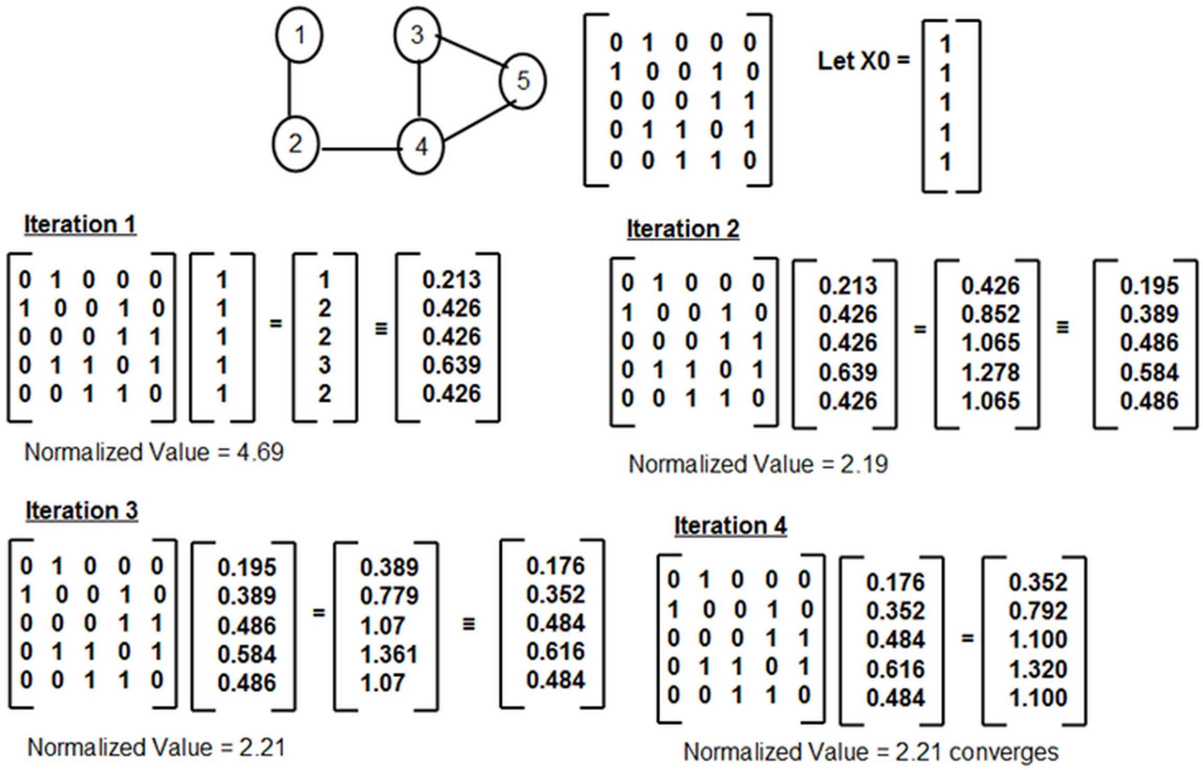
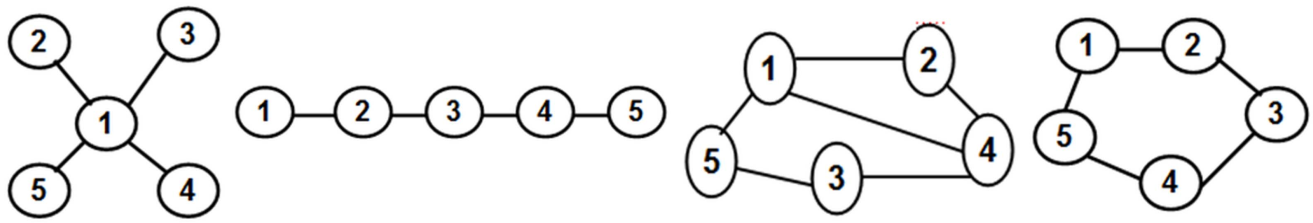


Figure 1. Example to compute spectral radius using the power iteration method.

If k_{min}, k_{avg} and k_{max} are the minimum, average and maximum node degrees of graph G, then $k_{min} \leq k_{avg} \leq \lambda_{sp}(G) \leq k_{max}$ [2]. We define the spectral radius ratio as the spectral radius divided by the average node degree. As per the above relationship, $\lambda_{sp}(G) / k_{avg} \geq 1$. The coefficient of variation for node degree is the standard deviation of the node degree (k_{SD}) divided by the average node degree. In the toy examples shown in Figure 2, we notice that the spectral radius gets closer to the average node degree (i.e., the spectral radius ratio approaches 1) as the variation in the node degrees in the graphs reduces. This forms the basis of our hypothesis that the spectral radius ratio and coefficient of variation of node degree are positively correlated. The reasoning behind our hypothesis is that the spectral radius represents the length of longest projection of the variation of the system. As $k_{min} \leq k_{avg} \leq \lambda_{sp}(G) \leq k_{max}$, we hypothesize that as the variation in the node degree reduces and k_{min} and k_{max} approach k_{avg}, the value of the ratio $\lambda_{sp}(G) / k_{avg}$ should also approach 1.

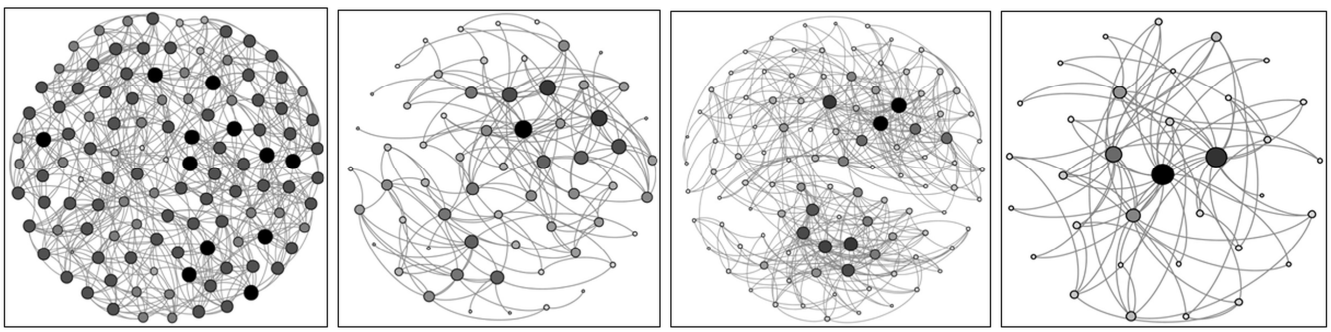


k_{avg}, $\lambda_{sp}(G)$, k_{SD} = {1.6, 2, 1.2} k_{avg}, $\lambda_{sp}(G)$, k_{SD} = {1.6, 1.73, 0.49} k_{avg}, $\lambda_{sp}(G)$, k_{SD} = {2.4, 2.48, 0.49} k_{avg}, $\lambda_{sp}(G)$, k_{SD} = {2.0, 2.0, 2.0}

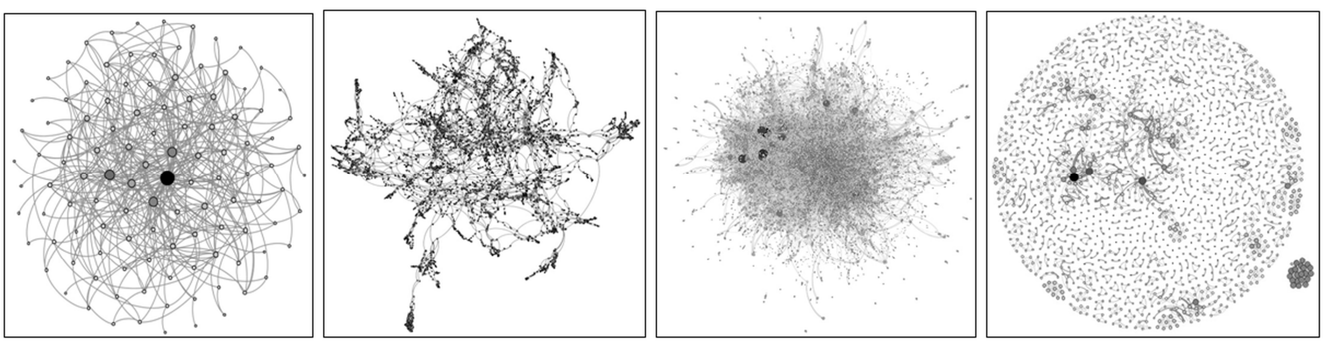
Figure 2. Toy examples to motivate hypothesis: Spectral radius-average degree ratio is positively correlated to the coefficient of variation of node degree.

III. Validation of the Hypothesis: Analysis of Complex Real-world Networks

We validate our hypothesis by analyzing the spectral radius and distribution of node degree on large complex network graphs obtained from real-world scenarios. We analyze a total of 8 classical network graphs that have been used in Network Science studies (available as .gml files at: http://www-personal.umich.edu/~mejn/netdata/). The networks analyzed (listed in the increasing order of the Spectral Radius Ratio for Node Degree) are: (i) American College Football Network [3] of the Fall 2000 season: nodes (115 nodes) are college teams and there is an edge (613 edges) between two nodes if and only if the corresponding teams have competed against each other earlier; (ii) Dolphin Social Network [4]: An undirected social network of frequent associations (159 edges) between 62 dolphins in a community living off Doubtful Sound, New Zealand; (iii) US Politics Books Network [5]: Nodes represent a total of 105 books about US politics sold by the online bookseller Amazon.com. A total of 441 edges represent frequent co-purchasing of books by the same buyers, as indicated by the "customers who bought this book also bought these other books" feature on Amazon; (iv) Zachary's Karate Club [6]: Social network of friendships (78 edges) between 34 members of a karate club at a US university in the 1970s; (v) Word Adjacencies Network [7]: This is a word co-appearance network representing adjacencies of common adjective and noun in the novel "David Copperfield" by Charles Dickens. A total of 112 nodes represent the most commonly occurring adjectives and nouns in the book. A total of 425 edges connect any pair of words that occur in adjacent position in the text of the book; (vi) Power Grid Network [8]: It is an undirected network (4941 nodes and 6594 edges) representing the topology of the power grid covering the western states of the US; (vii) Condensed Matter Collaborations [9]: It is a weighted network of co-authorships between scientists (16,726 nodes and 47,595 edges) posting pre-prints on the Condensed Matter e-print archive between Jan. 1, 1995 and Dec. 31, 1999; (viii) Network Science Co-authorships [7]: A network of authors (1589 nodes) who co-author and collaborate on any publication; two nodes are connected if the authors are co-authors for any publication (2742 edges). All networks are modeled as undirected graphs.



(i) American College Football Network (ii) Dolphins' Social Network (iii) US Politics Books Network (iv) Zachary's Karate Club Network



(v) Word Adjacencies Network (vi) Power Grid Network (vii) Condensed Matter Collaboration Network (viii) Network Science Co-authorship Network

Figure 3. Degree distribution of real-world networks (Increasing order of spectral radius - average node degree ratio).

TABLE I. NODE DEGREE AND SPECTRAL RADIUS FOR THE REAL-WORLD NETWORKS

Network #	# nodes	# edges	k_{min}	k_{max}	k_{avg}	k_{SD}	$\lambda_{sp}(G)$	$\lambda_{sp}(G)/ k_{avg}$	k_{avg} / k_{SD}
(i)	115	613	7	12	10.66	0.88	10.77	1.01	0.08
(ii)	62	159	1	12	5.13	2.93	7.18	1.40	0.57
(iii)	105	441	2	25	8.4	5.45	11.87	1.41	0.65
(iv)	34	78	1	17	4.59	3.82	6.72	1.46	0.83
(v)	112	425	1	49	7.59	6.85	13.15	1.73	0.90
(vi)	4,941	6,594	1	19	2.67	1.79	7.48	2.80	0.67
(vii)	16,726	47,595	0	107	5.69	6.42	24.97	4.39	1.13
(viii)	1,589	2,742	0	34	3.45	3.47	19.02	5.51	1.01

We developed our own Java program to compute the spectral radius, average node degree and the standard deviation of node degree for the above 9 network graphs. Figure 2 illustrates the distribution of the node degrees seen for each of the nine networks listed above (the larger the node size and darker the node color, the larger is the node degree).

Table 1 lists the values for minimum, maximum, average, standard deviation of the node degree, spectral radius, spectral radius ratio for node degree and the coefficient of variation for node degree (denoted respectively as k_{min}, k_{max}, k_{avg}, k_{SD}, λ_{sp}(G), λ_{sp}(G)/ k_{avg}, k_{SD} / k_{avg}). As it can be seen, the two ratios are positively correlated and have a high correlation coefficient of 0.75 (based on the well-known Pearson's Correlation Coefficient method [10]).

IV. CONCLUSIONS AND FUTURE WORK

The high-level contribution of this research is the identification of a positive correlation between the ratio of the spectral radius to the average node degree to that of the coefficient of variation of node degree. The positive correlation has been observed for networks of different sizes, ranging from networks of 5 nodes, to a few hundred nodes, all the way to networks of thousands of nodes. With this observation, we can now confidently say that the closer the value of the spectral radius ratio for node degree to 1.0, the smaller the variation among the degrees of the nodes in the network. The variation in the node degrees for any two networks can also be simply compared based on the spectral radius ratio for node degrees observed for the two networks, rather than requiring to compute the standard deviation of the node degrees for the two networks. As part of future work, we wish to extend this analysis for directed graphs based on the distribution of the in-degree and out-degree of the vertices as well as the spectral radius computed for both these variants of node degree.

REFERENCES

[1] G. Straing, Linear Algebra and its Applications, Cengage Learning, Stamford, CT, July 2005.

[2] B. G. Horne, "Lower Bounds for the Spectral Radius of a Matrix," Linear Algebra and its Applications, September 1997, vol. 263, pp. 261-273.

[3] 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, April 2002, vol. 19, no. 12, pp. 7821-7826.

[4] D. Lusseau, K. Schneider, O. J. Boisseau, P. Haase, E. Slooten and S. M. Dawson, "The Bottlenose Dolphin Community of Doubtful Sound Features a Large Proportion of Long-lasting Associations," Behavioral Ecology and Sociobiology, 2003, vol. 54, pp. 396-405.

[5] V. Krebs, http://www.orgnet.com/divided.html, last accessed: October 27, 2014.

[6] W. W. Zachary, "An Information Flow Model for Conflict and Fission in Small Groups," Journal of Anthropological Research, 1997, vol. 33, no. 4, pp. 452-473.

[7] M. E. J. Newman, "Finding Community Structure in Networks using the Eigenvectors of Matrices," Physics Review, September 2006, E 74, 036104.

[8] D. J. Watts, and S. H. Strogatz, "Collective Dynamics of Small-World Networks," Nature, June 1998, vol. 393, pp. 440-442.

[9] M. E. J. Newman, "The Structure of Scientific Collaboration Networks," Proceedings of the National Academy of Sciences of the United States of America, January 2001, vol. 98, no. 2, pp. 404-409.

[10] J. Benesty, J. Chen, Y. Huang, and I. Cohen, " Pearson Correlation Coefficient," Springer Topics in Signal Processing: Noise Reduction in Speech Processing, March 2009, vol. 2, pp. 1-4.