The Connections of Connections: New Results on the Centrality and Communicability of Networks

Alexander Farrugia
alex.farrugia@um.edu.mt

Abstract: In network science, the centrality, or importance, of a node in a network is of crucial importance, for instance to rank webpages or to measure the rate of the spread of news on a social network. Also of importance is measuring the communicability between two nodes in a network, which assesses how well the nodes can communicate between them. Two methods of calculating the centrality and communicability of nodes in a network, arising from the subgraph centrality approach utilizing walk of graphs are surveyed, with a focus on the more recently introduced of these two methods. A connection between the centrality and communicability scores produced by this scheme is presented. The question of how the centrality and communicability of nodes vary by the introduction of new links in the network is put forward. To answer this question, formulae that derive these centrality and communicability differences in terms of existing centralities of nodes within the network are presented.

Keywords: network, centrality, communicability, walks of graphs, resolvent matrix

During the past decade, the study of the interactions existing in networks associated with the natural, social, and technological sciences gave rise to the field of network science. In this area of study, a network can represent websites linked together on the World Wide Web, the interconnections of neurons in a brain, people connected on a social network, individuals interacting with each other, the structure of a molecule or of a protein, the connections of power lines supplying various buildings, and so on.[1] Network science thus unifies the theories established in these different areas of study, and others, to be able to provide collective answers to questions related to the connectivity of such networks.

Mathematics allows the unification of such networks arising from different fields of study by representing a network as a graph. A *graph* is a collection of nodes and edges linking several of these nodes together. The area of graph theory in mathematics is thus essential to network science, in which the centrality and communicability of nodes in networks play an important role. Figure 1 shows an example of a graph, or network, having six nodes and eight edges. Throughout this paper, the number of nodes in the network shall be denoted by n. For instance, in the example of Figure 1, $n = 6$.

In network science, the *centrality* of a node in a network quantifies how well that node is connected to all others. On the other hand, the *communicability* between two distinct nodes in the network assigns a value according to how well those two nodes can communicate with each other.[2] A node with high centrality is considered more important, or has more influence, over nodes having lower centrality. For example, in the World Wide Web, which can be considered as being a giant network linking websites together, a node with high centrality signifies a website that is more influential, or has overall more traffic, than others having a lower centrality score. On the other hand, in an epidemiology network representing the interactions between persons coming in contact with each other, two persons (nodes) having a low communicability score signifies that these persons have a low probability of transmitting potential contagious diseases to each other.

These two measures – centrality and communicability – are then combined to provide the *betweenness centrality* of a node, which is a measure of how the overall communicability of the network changes when that node is removed.[3] The betweenness centrality of a node can also be understood as measuring how much information passes through that particular node in order to reach others.[4]

1 Ernesto Estrada and Desmond J. Higham, 'Network Properties Revealed Through Matrix Functions', *SIAM Review*, 52 (4) (2010), 696–714.

2 M.E.J. Newman, 'A Measure of Betweenness Centrality Based on Random Walks', *Social Networks*, 27 (1) (2005), 39–54.

3 Estrada and Higham, 698.

4 Newman, 40.

The various ways by which these values may be produced is where mathematics comes to the fore.

Subgraph centrality

There are a multitude of methods used to calculate the centrality of nodes of networks. Benzi and Klymko provide an overview of these different methods.[5] One of these methods, the eigenvector centrality, was made famous by the PageRank algorithm, on which the search algorithm used by Google is based.[6] In this paper, the focus is on one method of calculating the centrality of nodes called the *subgraph centrality*, first put forward by Ernesto Estrada[7] in 2000 and later refined by Estrada and Rodríguez-Velázquez.[8]

Pertinent to this method is the concept of a *walk* on a graph (network). A walk on a network that starts from node A and ends at node B is a sequence of nodes, having first node A and last node B, such that any two consecutive nodes in the sequence are linked together by an edge in the network. The *length* of the walk is one less than the number of nodes that its sequence of nodes possesses. The sequence may contain repeated nodes, or it may not; indeed, the starting and ending nodes might be the same, in which case the walk is *closed*. In Figure 1, the walk $1 \rightarrow 2 \rightarrow 4 \rightarrow 6 \rightarrow 4 \rightarrow 5 \rightarrow 4$ is a walk of length six starting at node 1 and ending at node 4. An example of a closed walk on the same network that starts and ends at node 2 is the walk $2 \rightarrow 1 \rightarrow 2 \rightarrow 4 \rightarrow 3 \rightarrow 2$, having length five.

5 Michele Benzi and Christine Klymko, 'On the Limiting Behavior of Parameter-Dependent Network Centrality Measures', *SIAM J. Matrix Anal. Appl.*, 36 (2), 686–706.

6 Sergey Brin and Lawrence Page, 'The Anatomy of a Large-Scale Hypertextual Web Search Engine', *Computer Networks and ISDN Systems*, 30 (1) (1998), 107–17.

7 Ernesto Estrada, 'Characterization of 3D Molecular Structure', *Chem. Phys. Lett.*, 319 (2000), 713–8.

8 Ernesto Estrada and Juan A. Rodríguez-Velázquez, 'Subgraph Centrality in Complex Networks', *Physical Review E*, 71 (056103) (2005).

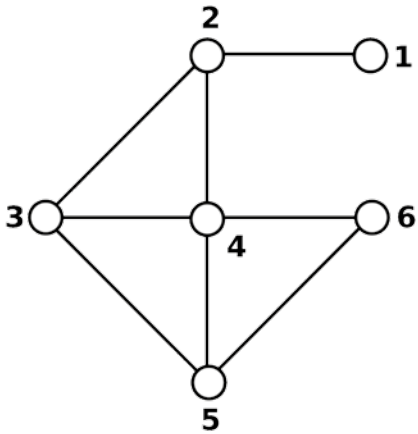

Figure 1 A simple network having six nodes labelled 1, 2, 3, 4, 5 and 6

A node B is *incident* to another node A if they are linked by an edge. One simple way to describe the centrality of a node is to simply count the number of closed walks of length two starting and ending at this node. This will count the number of nodes incident to that node; this number is called the *degree* of the node. This simple measure, called the *degree centrality*, is also one of the earliest employed – indeed, it was introduced in 1954 by Shaw.[9] However, a usually better measure of centrality of a node is obtained by counting all possible closed walks that start and end at that same node, and then weighting these walk counts according to their length. Applications usually dictate that the shorter a walk is, the more important it is deemed to be. This is how subgraph centrality works.

Of course, choosing different weights for the walk counts gives rise to different subgraph centrality measures. In Estrada's original paper that introduced the concept of the subgraph centrality,[10] he proposed to weigh the walk lengths as follows: a walk of length one is twice as important as a walk of length two, a walk of length two is three times as important as a walk of length three, a walk of length three is four

9 Linton C. Freeman, 'Centrality in Social Networks Conceptual Clarification', *Social Networks*, 1 (1978/79), 215–39.

10 Estrada., 715.

times as important as a walk of length four, and so on. This is called *exponential weighting*. Walks of length zero, which are only possible in closed walks, are given a score of one, while longer walks are counted as fractions of zero-length walks. This weighting of subgraph centrality spurred a large amount of interest from several authors coming from a wide range of fields such as biochemistry (particularly in the study of protein folding),[11] statistical thermodynamics,[12] quantum chemistry,[13] network theory,[14] and information theory.[15] Indeed, nowadays, the sum of the subgraph centralities of all the nodes in the network, using this weighting, is known as the *Estrada index*.[16] The communicability between two distinct nodes A and B in the network is also determined in the same way: walks starting at A and ending at B of various lengths are counted, then these counts are exponentially weighted as described above.

A second weighting, proposed by Estrada and Higham in 2010 is the following:[17] assuming the network has n nodes, any walk of length k is $(n - 1)$ times as important as a walk of length $(k + 1)$. For example, the walks of a network with nine nodes would be weighted as follows: a walk of length one is eight times as important as a walk of length two, a walk of length two is eight times as important as a walk of length three, a walk of length three is eight times as important as a walk of length four, and so on. As before, walks of length zero are given a score of one. We shall call this the *resolvent weighting*, as it is related to the *resolvent matrix* in mathematics.[18] It is important to note that the exponential weighting mentioned in the previous paragraph is related to a matrix in mathematics called the *matrix exponential* – hence its name.[19] The centrality and communicability measures arising from the

11 Ibid.,717.

12 Ernesto Estrada and Naomichi Hatano, 'Statistical-mechanical Approach to Subgraph Centrality in Complex Networks', *Chem. Phys. Lett.*, 439 (2007), 247–51.

13 Ernesto Estrada, Juan A. Rodríguez-Velázquez and Milan Randić, 'Atomic Branching in Molecules', *Int. J. Quantum Chem.*, 106 (2006), 823–32.

14 Estrada and Rodríguez-Velázquez.

15 Ramon Carbó-Dorca, 'Smooth Function Topological Structure Descriptors Based on Graph Spectra', *J. Math. Chem.*, 44 (2008), 373–8.

16 José Antonio de la Peña, Ivan Gutman and Juan Rada, 'Estimating the Estrada Index', *Lin. Algebra Appl.*, 427 (2007), 70–6.

17 Estrada and Higham.,702.

18 Carl D. Meyer, *Matrix Analysis and Applied Linear Algebra* (Philadelphia, PA, USA, 2000).

19 Roger A. Horn and Charles R. Johnson, *Matrix Analysis*, 2nd edn. (Cambridge, 2013).

resolvent weighting are called the resolvent centrality and resolvent communicability respectively.[20]

Matrices

Matrices were briefly mentioned in the previous paragraph. In mathematics, a *matrix* is a two-dimensional array of numbers. We can neatly package all the centrality and communicability measures of a network in a matrix, such that the entries (numbers) on the main diagonal (the one starting from the top left corner and ending at the bottom right corner) of the matrix are the centralities of each node, while the entries off this diagonal are the communicability measures of the network. For example, the third diagonal entry of such a matrix would be the centrality measure of node 3 of the network, while the entry in the second row and fourth column of the matrix would be the communicability measure between node 2 and node 4 of the network.

The network depicted in Figure 1 has the following two matrices associated with it, both containing the centrality and communicability measures of the network. The first matrix *E* uses exponential weighting, while the second matrix *R* uses resolvent weighting.

Exponential Weighting:
$$E = \begin{bmatrix} 1.673 & 1.789 & 1.118 & 1.194 & 0.680 & 0.444 \\ 1.789 & 3.985 & 3.664 & 4.032 & 2.756 & 1.874 \\ 1.118 & 3.664 & 4.742 & 4.914 & 4.032 & 2.624 \\ 1.194 & 4.032 & 4.914 & 6.172 & 4.858 & 3.719 \\ 0.680 & 2.756 & 4.032 & 4.858 & 4.666 & 3.351 \\ 0.444 & 1.874 & 2.624 & 3.719 & 3.351 & 3.180 \end{bmatrix}.$$

Resolvent Weighting:
$$R = \begin{bmatrix} 1.048 & 0.238 & 0.068 & 0.072 & 0.032 & 0.021 \\ 0.238 & 1.188 & 0.341 & 0.359 & 0.161 & 0.104 \\ 0.068 & 0.341 & 1.223 & 0.416 & 0.359 & 0.155 \\ 0.072 & 0.359 & 0.416 & 1.306 & 0.413 & 0.344 \\ 0.032 & 0.161 & 0.359 & 0.413 & 1.220 & 0.327 \\ 0.021 & 0.104 & 0.155 & 0.344 & 0.327 & 1.134 \end{bmatrix}.$$

The following briefly describes how the numbers in the sixth row and sixth column of the above two matrices, namely 3.180 and 1.134, were

20 Estrada and Higham, 702.

produced. The other numbers were determined in a similar manner. The closed walks starting and ending at node 6 are counted. We notice that there is one walk of length zero, no walks of length one, two walks of length two, two walks of length three, eleven walks of length four, and so on. These can all be easily verified by inspection, except perhaps the last claim that there are eleven closed walks of length four starting and ending at node 6. These are listed below for confirmation:

$6 \rightarrow 4 \rightarrow 6 \rightarrow 4 \rightarrow 6$
$6 \rightarrow 4 \rightarrow 6 \rightarrow 5 \rightarrow 6$
$6 \rightarrow 4 \rightarrow 5 \rightarrow 4 \rightarrow 6$
$6 \rightarrow 4 \rightarrow 2 \rightarrow 4 \rightarrow 6$
$6 \rightarrow 4 \rightarrow 3 \rightarrow 4 \rightarrow 6$
$6 \rightarrow 4 \rightarrow 3 \rightarrow 5 \rightarrow 6$
$6 \rightarrow 5 \rightarrow 6 \rightarrow 5 \rightarrow 6$
$6 \rightarrow 5 \rightarrow 6 \rightarrow 4 \rightarrow 6$
$6 \rightarrow 5 \rightarrow 4 \rightarrow 5 \rightarrow 6$
$6 \rightarrow 5 \rightarrow 3 \rightarrow 5 \rightarrow 6$
$6 \rightarrow 5 \rightarrow 3 \rightarrow 4 \rightarrow 6$

Using the exponential weighting, the centrality of node 6 is thus

$$1 + \left(\frac{1}{1}\right)0 + \left(\frac{1}{1}\right)\left(\frac{1}{2}\right)2 + \left(\frac{1}{1}\right)\left(\frac{1}{2}\right)\left(\frac{1}{3}\right)(2) + \left(\frac{1}{1}\right)\left(\frac{1}{2}\right)\left(\frac{1}{3}\right)\left(\frac{1}{4}\right)(11) + \cdots$$

Using the resolvent weighting, the centrality of node 6 is

$$1 + \left(\frac{1}{5}\right)0 + \left(\frac{1}{5}\right)\left(\frac{1}{5}\right)2 + \left(\frac{1}{5}\right)\left(\frac{1}{5}\right)\left(\frac{1}{5}\right)(2) + \left(\frac{1}{5}\right)\left(\frac{1}{5}\right)\left(\frac{1}{5}\right)\left(\frac{1}{5}\right)(11) + \cdots$$

Thus, infinitely many numbers must be summed up in both cases. However, both summations can be proved to always *converge* to some particular values, and do not become infinitely large as more numbers are added. Indeed, this can be proved to be true for any network, for the centrality of any node, and for the communicability of any distinct pairs of nodes.[21] Thus, the summation can be continued until the required degree of accuracy is achieved. In this case, the above summations

21 Ibid.,700.

converge to 3.180… and to 1.134… respectively. This explains why the matrix E has the value 3.180 at its sixth row and sixth column, while matrix R has the value 1.134 at the same position. Naturally, producing these numbers is the perfect job for a computer, and this is indeed the way that these numbers are usually produced.

Note also that if the degree centrality was used instead of these subgraph centralities, then nodes 2, 3 and 5 would have been given an equal score of 3, since these nodes all have three nodes incident to them. Using exponential weighting and resolvent weighting, however, these three nodes are given different scores, so that node 3 is deemed to be slightly more well-connected than node 5, which is, in turn, slightly more well-connected than node 2.

Relation between resolvent centrality and resolvent communicability of networks

In this paper, the focus is exclusively on the resolvent centrality and resolvent communicability, that is, on the numbers forming matrices akin to matrix R above. A method to determine the resolvent centrality of any node in a network in terms of each resolvent communicability measure between that node and any node incident to it is described. The question of how the resolvent centralities of nodes and the resolvent communicabilities of distinct pairs of nodes change by the introduction of a new link in the network is then posed. The answer to this question is shown to be surprisingly complicated.

The following results are presented:

Result 1: *The resolvent centrality of node A in a network having n nodes is the sum of the resolvent communicabilities between A and each node incident to A, divided by* $(n-1)$*, plus one.*

Result 2: *If node A is not incident to node B, then the resolvent communicability between nodes A and B in a network having n nodes is the sum of the resolvent communicabilities between A and each node incident to B, divided by* $(n-1)$*. (Swapping A and B in this result is permissible.)*

Both of the above results are proved together, by first introducing the *adjacency matrix* of the network. The adjacency matrix has a '1' at row A and column B if nodes A and B are linked by an edge; otherwise, it has a '0'. For example, the adjacency matrix of the network in Figure

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}.$$

It turns out that the matrix R containing the resolvent centralities and resolvent communicabilities of the network may be written in terms of the adjacency matrix A as the matrix $(n-1)\,((n-1)\,I-A)^{-1}$.[22] Here, I is the identity matrix, which is the matrix whose entries (numbers) on its main diagonal are all ones and whose off-diagonal entries are all zeros.

By definition of the matrix inverse, we have

$$((n-1)I-A)\,((n-1)\,I-A)^{-1} = I$$

Expanding,

$$(n-1)((n-1)\,I-A)^{-1} - A(n-1)I-A)^{-1} = I$$

Rearranging,

$$(n-1)((n-1)\,I-A)^{-1} = I + A((n-1)\,I-A)^{-1}$$

But since $R = (n-1)((n-1)\,I-A)^{-1}$
the above relation may be written as follows:

$$R = I + \frac{1}{\mathrm{n}-1} AR.$$

22 Alexander Farrugia, 'The Increase in the Resolvent Energy of a Graph Due to the Addition of a New Edge', *Applied Mathematics and Computation*, 321, (2018) 25–36.

Results 1 and 2 are then proved by equating each entry of matrix *R* on the left hand side of the above relationship with its corresponding entry on the right hand side.

Result 1 is illustrated using the network in Figure 1. The resolvent centrality of node 6 is 1.134, according to matrix *R*. Since node 6 is incident to node 4 and node 5, this number should be equal to the sum of the communicability between nodes 4 and 6 and the communicability between nodes 5 and 6, divided by 5 (one less than the number of nodes in the network), plus one. Indeed, $\frac{0.344+0.327}{5}+1$ is equal to 1.134.

Moving on to Result 2, according to the same matrix *R*, the resolvent communicability between nodes 3 and 1 is 0.068. Node 3 is incident to nodes 2, 4 and 5, so by Result 2, 0.068 should be one fifth of the sum of the communicabilities between nodes 1 and 2, nodes 1 and 4 and nodes 1 and 5. We confirm that this is the case, since $\frac{0.238+0.072+0.032}{5}=0.068$ The same result can also be obtained by noting that node 1 is only incident to node 2, so by swapping nodes 3 and 1 and reapplying Result 2, 0.068 should also be equal to the communicability between nodes 3 and 2, divided by five. Indeed, $\frac{0.341}{5}=0.068$ too.

The change in the resolvent centrality and resolvent communicability caused by the introduction of a new link to the network

The centrality of each node and the communicability between any two nodes in the network must increase after any two nodes are joined by an edge. The reason for this is that this new link will increase the number of walks of various lengths in the network. This will, in turn, directly affect all centrality and communicability scores in the network, each ending up increasing slightly.

The problem, then, is to quantify this increase, because the centrality score of each node in the network will possibly be increased by different amounts. For the resolvent weightings of graphs, the change in the centrality of a node and the communicability of pairs of nodes have been quantified in the recent paper by Farrugia.[23] Unfortunately, the equations that provide these changes are rather complicated.

Before proceeding, we denote the resolvent centrality at node A

23 Ibid., 29.

by C_A or by $C_{A,A}$. Moreover, the resolvent communicability between the distinct nodes A and B is denoted by $C_{A,B}$ Furthermore, we assume that nodes A and B were not linked together by an edge prior to the introduction of the new link in the network.

Result 3: *The resolvent centrality at node N after nodes A and B are linked together increases by*

$$\frac{2(n-1-C_{A,B})C_{N,A}C_{N,B}+C_A\left(C_{N,B}\right)^2+C_B\left(C_{N,A}\right)^2}{\left(n-1-C_{A,B}\right)^2-C_AC_B}$$

Result 4: The resolvent communicability between nodes M and N after nodes A and B are linked together increases by

$$\frac{(n-1-C_{A,B})(C_{M,A}C_{N,B}+C_{M,B}C_{N,A})+C_AC_{M,B}C_{N,B}+C_BC_{M,A}C_{N,A}}{\left(n-1-C_{A,B}\right)^2-C_AC_B}.$$

Recall that if one (or both) of M or N is/are the same as one (or both) of A or B, then the notation $C_{A,A}$ may be simplified to C_A. For example, the resolvent communicability increase between nodes A and B themselves after they are linked together is the slightly simpler quantity

$$\frac{(n-1)\left(C_AC_B+\left(C_{A,B}\right)^2\right)+C_{A,B}\left(C_AC_B-\left(C_{A,B}\right)^2\right)}{\left(n-1-C_{A,B}\right)^2-C_AC_B}.$$

Again, we illustrate these results using the example network of Figure 1. Suppose nodes 1 and 6 are linked together by an edge. By Result 3, the increase in resolvent centrality of node 5 owing to the presence of this new link in the network amounts to

$$\frac{2(5-0.021)(0.032)(0.327)+(1.048)(0.327)^2+(1.134)(0.032)^2}{(5-0.021)^2-(1.048)(1.134)}$$

$$=\frac{0.104+0.112+0.001}{24.790-1.188}=\frac{0.217}{23.602}=0.009.$$

This means that the effect on the resolvent centrality of node 5 after the new link between nodes 1 and 6 is introduced is an increase from 1.220 to 1.220 + 0.009, or 1.229.

The increase in resolvent communicability between nodes 3 and 4 caused by the introduction of the new link between nodes 1 and 6 is now investigated. By Result 4, this amounts to

$$\frac{2(5-0.021)(0.032)(0.327)+(1.048)(0.327)^2+(1.134)(0.032)^2}{(5-0.021)^2-(1.048)(1.134)}$$
$$=\frac{0.104+0.112+0.001}{24.790-1.188}=\frac{0.217}{23.602}=0.009.$$

Hence, connecting nodes 1 and 6 together increases the resolvent communicability between nodes 3 and 4 in the network from 0.416 to 0.416 + 0.010 = 0.426.

These values may be confirmed by calculating them directly using the method described at the end of the 'Matrices' section of this paper.

Conclusion

This paper presented expressions for the resolvent centrality in terms of certain resolvent communicability scores in the network. Moreover, formulae for the difference in resolvent centrality and resolvent communicability of nodes because of the introduction of a new link in the network were revealed.

Similar expressions for the exponential centrality and communicability of nodes, rather than those with resolvent weightings as discussed in this paper, are much more difficult to derive. The main difficulty to overcome in such an endeavour is the noncommutativity of matrix multiplication. The rule $e^x e^y = e^{x+y}$ for any numbers x and y is well-known, even by schoolchildren. Unfortunately, the corresponding law for matrix exponentials, that is $\exp(A)\exp(B) = \exp(A+B)$, only holds when $AB = BA$. In fact, this is a necessary and sufficient condition, in the sense that if $AB \neq BA$, then $\exp(A)\exp(B)$ and $\exp(A+B)$ are guaranteed to be different matrices.[24]

24 Cleve Moler and Charles Van Loan, 'Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later', *SIAM Review*, 45 (1) (2003), 3–49.

Further research on this area is suggested. Indeed, it would be interesting to attempt to derive similar results to those presented in this paper for the exponential subgraph centrality of networks. After all, as mentioned earlier in this paper, the exponential subgraph centrality, on which the Estrada index is based, is already being utilized in plenty of important applications.

Alexander Farrugia is mathematics lecturer at the University of Malta Junior College and a part-time lecturer at the University of Malta. His area of expertise includes algebraic and spectral graph theory, with applications in molecular chemistry and graph energies. He has several publications in renowned journals.