An Introduction to Centrality Measures with a Transportation Network Defense Exercise

James Kupetz, Jr.

Karl Levy, Ph.D.

1. Introduction

Greetings. This module is intended for use in an introductory discrete math course. It should take between 1 and 2 hours to complete (perhaps with some of the questions finished up as part of a homework assignment).

1.1. Audience

This module was written with first- and second-year college students in mind. It would fit quite well into an introduction to discrete math course that includes some amount of graph theory.

1.2. Required knowledge

- Precalculus
- A basic understanding of matrices
- A basic understanding of sigma notation

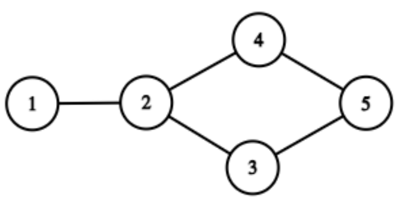
1.3. Objectives

- Introduce graphs
- Introduce measures of centrality
- Introduce an application of measures of centrality to real world networks, specifically with soft target defense and natural disaster response in mind.

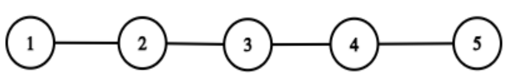
2. Introduction to Graphs

A **graph** (G) is a set of **vertices** (V) which are connected by a set of **edges** (E). In mathematical terms, G = (V, E), but don't let the math discourage you from delving further into this, it's really quite intuitive! We can visualize a graph as a set of points (called vertices) with lines (called edges) connecting related vertices. A simple graph modeling a subway network would include stations as vertices and the lines connecting them as edges. An edge is present between two vertices if and only if there is a direct subway connection between the two stations.

We will be examining graphs that are **connected.** A **path** is a sequence of vertices with consecutive pairs of vertices connected by an edge. Two vertices x and y on a graph are connected if there is a path from x to y. A graph is connected if each vertex on the graph is connected to every other vertex. In a simple graph modeling a subway network, the number of edges on a path between two vertices is its length. A shortest path between to vertices is called a **geodesic.**



Graph 2



Graph 1

(Exercise 1) What are the shortest paths, i.e, the geodesics, from vertex 1 to vertex 5 in Graph 1 and in Graph 2?

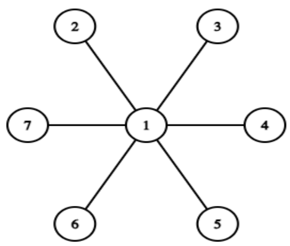
Solution: In Graph 1 the shortest path has length 4 and is (v_1, v_2, v_3, v_4, v_5). In Graph 2 there are two geodesics of length 3, which are (v_1, v_2, v_4, v_5) and (v_1, v_2, v_3, v_5).

3. Centrality Measures

When studying a graph that models a network, be it a subway system, a power grid, a communication network, etc., we often want to know which vertices are the most important, in one sense or another. We call measures of a vertex's importance **measures of centrality**. This module explores 3 basic measures of centrality but, rest assured, many more do exist.[1]

3.1. Degree Centrality

The simplest measure of centrality is **degree**. The degree of a vertex, deg(v), is equal to the number of edges connecting it to other vertices.



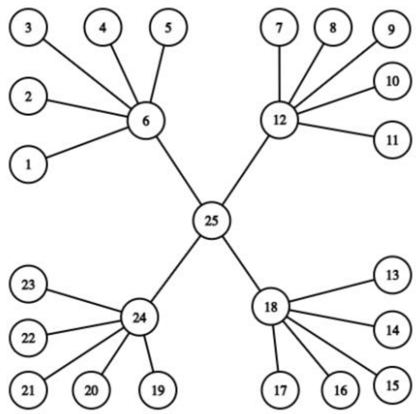
Graph 3

(Exercise 2) What are the degrees of the vertices of Graph 3.

Solution: deg(v_1) = 6 and deg(v_2) = deg(v_3) = deg(v_4) = deg(v_5) = deg(v_6) = deg(v_7) = 1

Although degree centrality can tell you a lot about a vertex it is not always the best measure of a vertex's importance.

1 We suggest that the curious reader takes a look at this very cool periodic table of centrality measures: http://schochastics.net/sna/periodic.html Can you come up with your own measure of centrality?



Graph 4

In Graph 4 imagine that each vertex indicates a person and an edge indicates a friendship between two people. A person is asked to get an important message to all of their friends ("first degree friends") and also to all of their friends' friends ("second degree friends").

(Exercise 3) In Graph 4, would the person represented by v_6 or vertex v_{25} get the message out to more people?

Solution: Even though $deg(v_{25}) < deg(v_6)$, v_{25} get the message out to more people; $deg(v_{25}) = 4$ "first degree friends" + 20 "second degree friends" means v_{25} would get the message out to 24 people; but $deg(6) = 6$ "first degree friends" + 2 "second degree friends" means v_6 would only get the message out to 8 people.

Perhaps there is some other measure that captures this kind of importance, that is, a vertex's closeness to not just its neighbors but also its neighbors neighbors and so forth.

3.2. Closeness Centrality

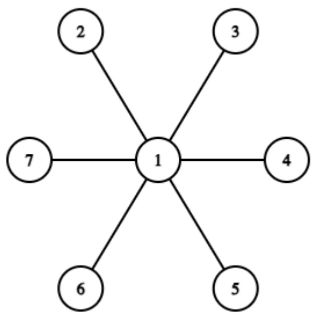
Closeness centrality is *a* measure of the average length of the shortest paths between a given vertex and all other vertices. To see how this works, we reproduce Graph 3 from above and form its distance matrix.

Two quick reminders before we start on closeness:

Sigma notation is a fast way of writing a long sum. For example, the sum x = 1+2+3+4+5 can be written as $x = \sum_{n=1}^{5} n$. We start with n=1 and do the summation operation as we increase the number until it reaches 5.

When working with matrices, we identify elements of the matrix as d_{ij} - the element in the i^{th} row and j^{th} column. For example, $d_{2,3}$ in the following matrix is 8.

$$\begin{matrix} 4 & 2 & 3 \\ 7 & 1 & 8 \\ 9 & 0 & 6 \end{matrix}$$



Graph 3 (again)

The Distance Matrix for the Graph

$$3D = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 2 & 2 & 2 & 2 & 2 \\ 1 & 2 & 0 & 2 & 2 & 2 & 2 \\ 1 & 2 & 2 & 0 & 2 & 2 & 2 \\ 1 & 2 & 2 & 2 & 0 & 2 & 2 \\ 1 & 2 & 2 & 2 & 2 & 0 & 2 \\ 1 & 2 & 2 & 2 & 2 & 2 & 0 \end{bmatrix}$$

Each matrix entry d_{ij} is the distance, which is the number of edges in a shortest path between vertex v_i and vertex v_j. For Graph 3's central vertex v_1 we see, in both column 1 and row 1, that the distance between it and all other vertices is 1. But since the distance is smaller the closer a vertex is, we define a vertex's **closeness centrality** as the inverse of the average distance to all other vertices

$C_i = \frac{(n-1)}{\sum_{j \neq i} d_{ij}}$ where n is the number of vertices in our graph. This way the closer a vertex is to all others, the higher its closeness centrality is.

Let's look at part of what is happening in the previous calculation in a little more detail. Instead of just averaging the distances between a vertex and all of the other vertices as a measure of closeness, we then find its inverse. The advantage of this is a measure that places all of the possible results between 0 and 1. If we look at three different vertices A, B and C where the average of the distances from the vertex to all other vertices is 2 for A, 10 for B and 100 for C, the inverse function calculates closeness measures of $C_A = 0.5$, $C_B = 0.1$ and $C_C = 0.01$ putting them in the correct numerical order of closeness, with vertex A having the highest closeness value, and vertex C the lowest. In this way, a vertex with a closeness value of 1 would be adjacent to all of the other vertices in the graph, and as closeness values get closer to 0, the vertices would be further from their neighbors, that is, less close.

Graph 5

(Exercise 4) In Graph 5, guess which vertex has the highest closeness centrality and then calculate C_1, C_2 and C_3. Were you correct?

Solution: $C_2 = (3-1)/(1+1) = 2/2 = 1$. $C_1 = C_3 = (3-1)/(1+2) = 2/3$.

(Exercise 5) Calculate C_1 and C_7 using the distance matrix for Graph 3.

Solution: $C_1 = (7-1)/(1+1+1+1+1+1) = 6/6 = 1$. $C_7 = (7-1)/(1+2+2+2+2+2) = 6/11$.

(Exercise 6) Write down the 25×25 distance matrix for the Graph 4. Which vertex would you guess has the highest closeness centrality? Which vertices would you guess have the lowest closeness centrality? Calculate C_i for all the vertices (hint: there are only 3 different values). Were your guesses about highest and lowest correct?

Solution:

0	2	2	2	2	1	4	4	4	4	4	3	4	4	4	4	4	3	4	4	4	4	4	3	2
2	0	2	2	2	1	4	4	4	4	4	3	4	4	4	4	4	3	4	4	4	4	4	3	2
2	2	0	2	2	1	4	4	4	4	4	3	4	4	4	4	4	3	4	4	4	4	4	3	2
2	2	2	0	2	1	4	4	4	4	4	3	4	4	4	4	4	3	4	4	4	4	4	3	2
2	2	2	2	0	1	4	4	4	4	4	3	4	4	4	4	4	3	4	4	4	4	4	3	2
1	1	1	1	1	0	3	3	3	3	3	2	3	3	3	3	3	2	3	3	3	3	3	2	1
4	4	4	4	4	3	0	2	2	2	2	1	4	4	4	4	4	3	4	4	4	4	4	3	2
4	4	4	4	4	3	2	0	2	2	2	1	4	4	4	4	4	3	4	4	4	4	4	3	2
4	4	4	4	4	3	2	2	0	2	2	1	4	4	4	4	4	3	4	4	4	4	4	3	2
4	4	4	4	4	3	2	2	2	0	2	1	4	4	4	4	4	3	4	4	4	4	4	3	2
4	4	4	4	4	3	2	2	2	2	0	1	4	4	4	4	4	3	4	4	4	4	4	3	2
3	3	3	3	3	2	1	1	1	1	1	0	3	3	3	3	3	2	3	3	3	3	3	2	1
4	4	4	4	4	3	4	4	4	4	4	3	0	2	2	2	2	1	4	4	4	4	4	3	2
4	4	4	4	4	3	4	4	4	4	4	3	2	0	2	2	2	1	4	4	4	4	4	3	2
4	4	4	4	4	3	4	4	4	4	4	3	2	2	0	2	2	1	4	4	4	4	4	3	2
4	4	4	4	4	3	4	4	4	4	4	3	2	2	2	0	2	1	4	4	4	4	4	3	2
4	4	4	4	4	3	4	4	4	4	4	3	2	2	2	2	0	1	4	4	4	4	4	3	2
3	3	3	3	3	2	3	3	3	3	3	2	1	1	1	1	1	0	3	3	3	3	3	2	1
4	4	4	4	4	3	4	4	4	4	4	3	4	4	4	4	4	3	0	2	2	2	2	1	2
4	4	4	4	4	3	4	4	4	4	4	3	4	4	4	4	4	3	2	0	2	2	2	1	2
4	4	4	4	4	3	4	4	4	4	4	3	4	4	4	4	4	3	2	2	0	2	2	1	2
4	4	4	4	4	3	4	4	4	4	4	3	4	4	4	4	4	3	2	2	2	0	2	1	2
4	4	4	4	4	3	4	4	4	4	4	3	4	4	4	4	4	3	2	2	2	2	0	1	2
3	3	3	3	3	2	3	3	3	3	3	2	3	3	3	3	3	2	1	1	1	1	1	0	1
2	2	2	2	2	1	2	2	2	2	2	1	2	2	2	2	2	1	2	2	2	2	2	1	0

$$C_1 = \frac{25-1}{100} = .24 = C_2 = C_3 = C_4 = C_5 = C_7 = C_8 = C_9 = C_{10} = C_{11} = C_{13} = C_{14} = C_{15} = C_{16}$$

$$= C_{17} = C_{19} = C_{20} = C_{21} = C_{22} = C_{23}$$

$$C_6 = \frac{25-1}{57} = .421 = C_{12} = C_{18} = C_{24}$$

$$C_{25} = \frac{25-1}{44} = .545$$

These calculations should reinforce our intuition that vertex 25 is the most central vertex in graph 4, followed by vertices 6, 12, 18 and 24.

3.3. Betweenness Centrality

Closeness centrality might be a good measure of importance for graphs that model networks of mostly static relationships like friendships, family bonds, physical locations, etc. But what about graphs that model networks of flows like the flow of electricity, commuters, information, etc.?

In these latter cases, the number of shortest paths that go through a certain vertex may matter more than just how close that vertex is to all the others. We here define **betweenness centrality** of a vertex v_i as

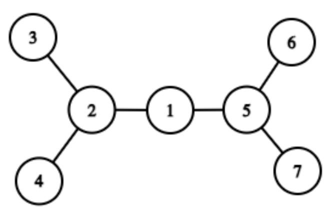
$$x_i = \frac{1}{2}\sum \frac{n_{jk}^{i}}{g_{jk}}$$

where n_{jk}^{i} is the number of shortest paths between vertex v_j and v_k that pass through vertex v_i.[2] g_{jk} is the total number of shortest paths between vertex v_j and v_k. We multiple by 1/2 because n_{jk}^{i} and n_{kj}^{i} count the same path(s) twice. But enough with the definition! In the simple example we explore below, all of the n_{jk}^{i} equal either 0 or 1 and all of the g_{jk} equal 1.

Calculating the betweenness centrality takes a bit more work than calculating the closeness centrality. We have to construct a "betweenness matrix"

$$B_i = \left[\frac{n_{jk}^{i}}{g_{jk}}\right]$$

for each vertex rather than the one distance matrix for all of them. We then add up all of a matrix's entries and multiple by 1/2 to get our betweenness centrality x_i.



Graph 6

For Graph 6, let's calculate x_i. First we construct the betweenness matrix

2 If the path starts or ends on v_i or if it starts or ends on the same vertex—equivalently, if j=i, k=i or j=k—then $n_{jk}^{i} = 0$.

$$B_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

Then we add up all the entries of B_1 and divide by 2 (or only add up the entries above or below the diagonal of zeros, noticing the symmetry) and get $x_1 = 9$.

(Exercise 7) For Graph 6, construct the betweenness matrices B_2 and B_3 and then calculate the betweenness centralities x_2 and x_3.

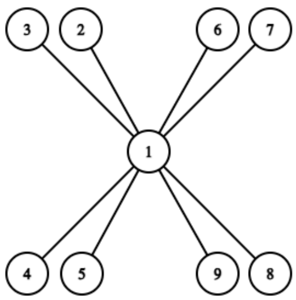
Solution:

$$B_2 = \begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \text{ therefore, } x_2 = 9$$

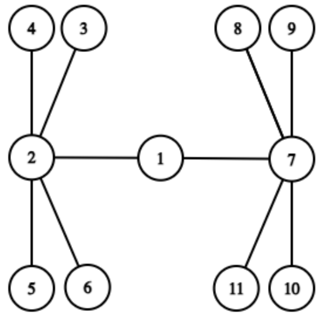
$$B_3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \text{ therefore, } x_3 = 0$$

(Exercise 8) Are more calculations necessary to find the remaining x_i?

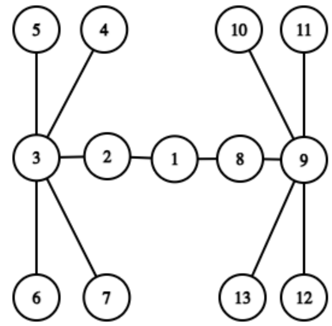
Solution: No. Examining the symmetries in Graph 6 should make it clear that $x_3 = x_4 = x_6 = x_7$ and that $x_2 = x_5$.



Graph 7



Graph 8



Graph 9

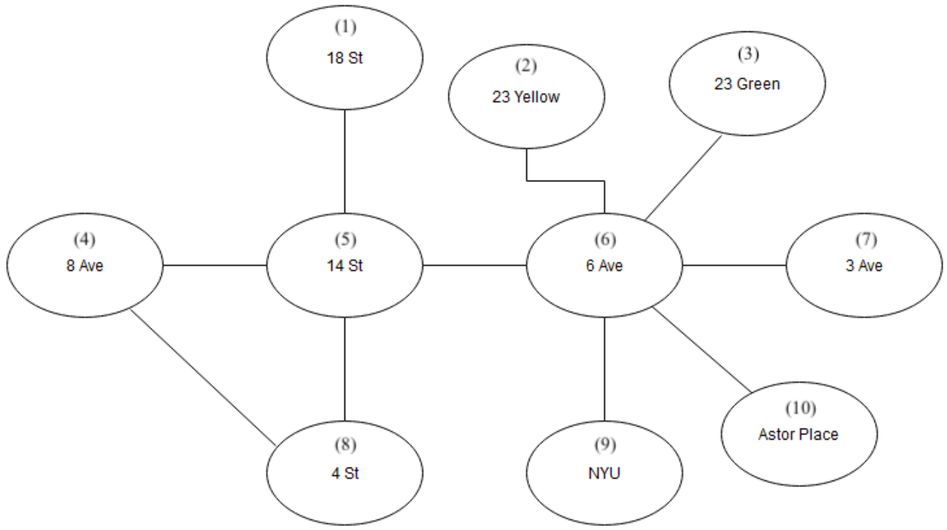
(Exercise 9) How do the degree centrality, the closeness centrality and the betweenness centrality change for v_1 for Graph 7, Graph 8 and Graph 9?

4. A Transportation Network Exercise

In the past, transportation networks have been attacked with chemical and biological weapons. One way to prevent such attacks is by the use of detectors that can sense small amounts of chemical and biological substances. These detectors are often very expensive and so, with limited resources, the choice of where to place the detectors is critical.

Similarly, natural disasters like huge storms have damaged transportation networks. One way to prevent such damage is to fortify the network's stations against flooding. But again, with limited resources, the choice of which stations to fortify is critical.

Let's examine a portion of the New York City transit system and use some of the concepts we have developed so far.



Graph 10 (A Portion of the NYC Subway System)

(Exercise 9) In your opinion, which stations in Graph 10 are the most important to place detectors in or to fortify against flooding? Explain your answers.

Responses will vary. Explanation for intuitive responses will most likely include degree of the vertices as a reason.

(Exercise 10) Calculate the degree centrality, closeness centrality and betweenness centrality for all of the stations. How do these results either reinforce or go against your answer to Exercise 9?

Solution:

Degree centrality:

vertex	1	2	3	4	5	6	7	8	9	10
degree	1	1	1	2	4	6	1	2	1	1

Distance matrix

	1	2	3	4	5	6	7	8	9	10
1	0	3	3	2	1	2	3	2	3	3
2	3	0	2	3	2	1	2	3	2	2
3	3	2	0	3	2	1	2	3	2	2
4	2	3	3	0	1	2	3	2	3	3
5	1	2	2	1	0	1	2	1	2	2
6	2	1	1	2	1	0	1	2	1	1
7	3	2	2	3	2	1	0	3	2	2
8	2	3	3	2	1	2	3	0	3	3
9	3	2	2	3	2	1	2	3	0	2
10	3	2	2	3	2	1	2	3	2	0

Closeness centrality:

vertex	1	2	3	4	5	6	7	8	9	10
formula	$\frac{9}{22}$	$\frac{9}{20}$	$\frac{9}{20}$	$\frac{9}{22}$	$\frac{9}{13}$	$\frac{9}{12}$	$\frac{9}{17}$	$\frac{9}{22}$	$\frac{9}{20}$	$\frac{9}{20}$
closeness	. 409	.450	.450	.409	.692	.750	.529	.409	.450	.450

Betweenness centrality:

The betweenness matrices for vertices 1, 2, 3, 4, 7, 8, 9 and 10 have all 0 entries, therefore a betweenness measure of 0, for each of those vertices. We only need to calculate the betweenness of vertices 5 and 6, so we start with the matrix of each.

Betweenness matrix for vertex 5

	1	2	3	4	5	6	7	8	9	10
1	0	1	1	1	0	1	1	1	1	1
2	1	0	0	1	0	0	0	1	0	0
3	1	0	0	1	0	0	0	1	0	0
4	1	1	1	0	0	1	1	0	1	1
5	0	0	0	0	0	0	0	0	0	0
6	1	0	0	1	0	0	0	1	0	0
7	1	0	0	1	0	0	0	1	0	0
8	1	1	1	0	0	1	1	0	1	1
9	1	0	0	1	0	0	0	1	0	0
10	1	0	0	1	0	0	0	1	0	0

Therefore $x_5 = 20$

Betweenness matrix for vertex 6

	1	2	3	4	5	6	7	8	9	10
1	0	1	1	0	0	0	1	0	1	1
2	1	0	1	1	1	0	1	1	1	1
3	1	1	0	1	1	0	1	1	1	1
4	1	1	1	0	0	0	1	0	1	1
5	0	1	1	0	0	0	1	0	1	1
6	0	0	0	0	0	0	0	0	0	0

7	0	1	1	1	1	0	0	1	1	1
8	1	1	1	0	0	0	1	0	1	1
9	0	1	1	1	1	0	1	1	0	1
10	0	1	1	1	1	0	1	1	1	0

Therefore $x_6 = 29$

Performance Task

Find a transportation network near you or, if none are near you, one of interest to you. The New York City Subway is a good one to use. Choose a minimum of 20 nodes on the network and create a graph showing the connections between the nodes. From your graph, calculate the degree centrality, closeness centrality and betweenness centrality for each node. As the graph becomes more complicated, does the calculation of centrality differ from your intuition of which nodes are central to the network?

Bibliography and Acknowledgements

Chapter 7 of The Newman's *Networks* offers an introduction to the subject of measures of centrality with amble references and bibliography for a deeper dive:

Newman, Mark. *Networks*. Second edition., OUP Oxford, 2018, https://doi.org/10.1093/oso/9780198805090.001.0001.

The following site was used to draw some of our graphs:

https://csacademy.com/app/graph_editor/