IV: Section (vi) Network Centrality: Degree centrality, Closeness Centrality, Eigenvector Centrality & Katz Centrality.

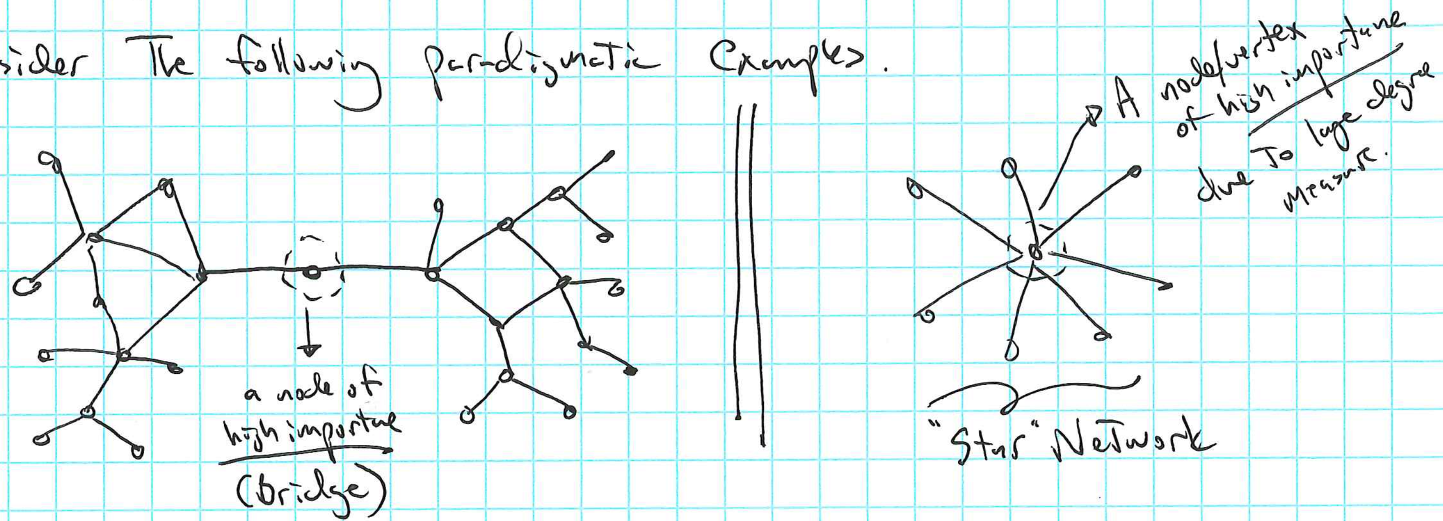
In the study of graphs/networks & their applications, it is natural to analyze the significance/importance of individual vertices. For networks, the notion of the importance of a particular vertex is generally expressed as $\boxed{\text{vertex centrality}}$.

We consequently desire a coherent definition of centrality that captures, in some sense, the degree to which a vertex is "influential" or important. But what characterizes the influence of a vertex in a graph? Should we privilege global or local aspects of "influence"? Because of the multi-faceted nature of such centrality considerations, a variety of useful metrics for vertex centrality exist. In total we explore (4) of the most common definitions of centrality in this lecture. Before a careful investigation of these different measures, however, we first offer some motivating insights for network centrality considerations.

First, it is helpful to relate vertex importance to the extent to which a vertex partakes in or facilitates the flow of information/traffic in a given network.

Thinking in this light, we might then concern ourselves with (2) essential centrality criteria: (1) How much of the network's resources flow through this vertex & (2) How crucial is this particular node to the (global) network flow - is it fungible, inessential?

Consider the following paradigmatic examples.



With the network example on the left, the node in the middle seemingly has a high degree of importance because any information flowing from the left-subgraph to the right-subgraph (or vice versa) must pass through this node; this vertex is sometimes called a bridge (note that it also serves as a cut-vertex of the graph); furthermore, despite the small degree measure of this vertex it is nevertheless important to the network from a global perspective.

For the graph on the right, the middle vertex is significant in both a global & local sense. Locally, $deg(v) = \Delta(G)$, in other words it is a vertex of maximum degree, and in addition it serves as something of an informational "hub", since its deletion generates a disconnected graph.

Given these considerations, we offer (3) general criteria for assessing vertex centrality (certainly there exist other cogent criteria).

1 Connectivity Measure: How well-connected is a given vertex to other vertices in the graph in a local sense? (see: Degree Centrality)

2 Closeness Measure: How "close" is a given vertex to all other vertices in the graph (a global measure)? (see: Closeness Centrality)

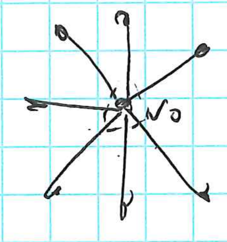
3 "Prestige" Measure: Is the vertex connected to many other vertices of high importance? (see: Eigenvector Centrality, Katz Centrality)

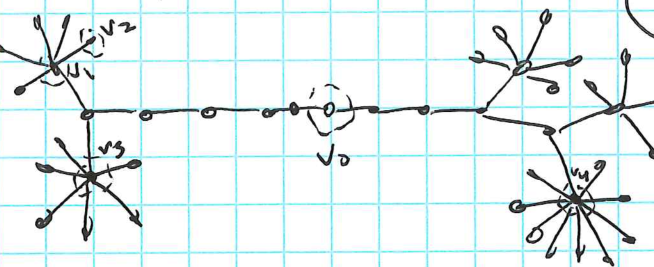
I Degree Centrality

Quite obviously, a vertex with a high degree in a graph (one may think of a person with many friends in a social network, for example), has, in some intrinsic sense, a large degree of importance in that graph. Degree centrality accordingly assigns a centrality measure equal to the degree of the given vertex, for all vertices in the graph.

$\boxed{C_D(v) = \deg(v)}$ for all $v \in V(G)$.

Ex.





Consider, for instance, the networks shown above. For the network on the left (a <u>tree</u>), $C_D(v_0) = 7$, while $C_D(v_i) = 1$ where v_i is any <u>leaf</u> in the graph. Without much controversy, we can say that the maximum centrality measure corresponds with our intuitive sense of the most important node in the graph.

Now consider the network on the right. Clearly, $C_D(v_1) = 6$, $C_D(v_2) = 1$, $C_D(v_3) = 7$, $C_D(v_4) = 10$, $C_D(v_0) = 2$ and so forth. Using this centrality metric, then, we would identify vertex $\boxed{v_4}$ as the most important in the network – and indeed this may be the case, depending on additional details of the network. However, at least in one sense we may credibly argue that vertex $\boxed{v_0}$ is, contrarily, <u>more important</u> than v_4, since it is, on average, "closer" to all other vertices in the graph. This discrepancy will shortly motivate our definition of <u>closeness centrality</u>.

Note that degree centrality is often "normalized" and alternatively defined as: $C_D(v) = \frac{\deg(v)}{\sum \deg(v_i)}$. Furthermore, in the case of a <u>digraph</u>, oftentimes <u>two</u> degree centrality measures are invoked – one for "indegree" & one for "outdegree".

If, for instance, your network represents, say, a network of academic papers and their associated citations, "indegree" (viz. how many papers reference your paper) might be a measure of crucial significance, whereas "outdegree" would be of negligible importance.

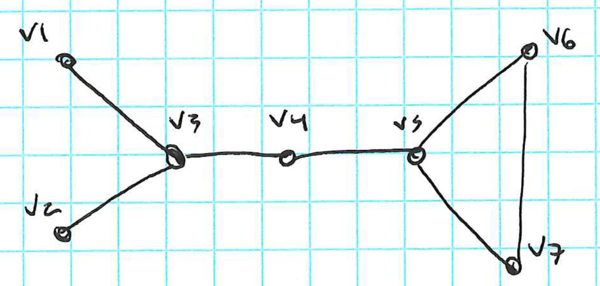
II Closeness Centrality

In connected graphs a natural measure of centrality is defined as the literal "closeness" of a vertex to all other vertices; here we take the reciprocal of the sum of the distance of the given vertex to all other vertices.

$$\boxed{C(x) = \frac{1}{\sum_y d(y,x)}} \quad \text{for each } x \in V(G)$$

·) where: $d(y,x) :=$ the shortest distance (measured by the number of edges) of any path connecting x & y in the graph. Note that this often called the geodesic distance between the two vertices. If G is disconnected & so such path exists, we say: $d(y,x) = \infty$, whereupon $\frac{1}{d(x,y)} = \frac{1}{\infty} = 0$.

Ex. We calculate the closeness centrality for each vertex in the following network.



Note That: $d(v_1,v_2)=2$, $d(v_1,v_3)=1$, $d(v_1,v_4)=2$,

$d(v_1,v_5)=3$, $d(v_1,v_6)=4$, $d(v_1,v_7)=4$

Thus, $$C(v_1) = \frac{1}{\sum_y (v_1,y)} = \frac{1}{2+1+2+3+4+4} = \boxed{\frac{1}{16}}$$

Continuing in This fashion, we get:

$C(v_1) = \frac{1}{16}$ $\quad C(v_3) = \frac{1}{11}$ $\quad C(v_5) = \frac{1}{11}$ $\quad C(v_7) = \frac{1}{15}$

$C(v_2) = \frac{1}{16}$ $\quad \boxed{C(v_4) = \frac{1}{10}}$ $\quad C(v_6) = \frac{1}{15}$

This shows That vertex $\boxed{v_4}$ has The highest closeness centrality of all vertices in The graph, confirming our intuition.

III Eigenvector Centrality Bonacich (1987)

In many contexts, we can imagine That The importance of a vertex in a network should not only be determined by The "local connectedness" (see degree centrality) as a quantitative measure, but also as a qualitative measure. To This end, consider a social network in which "Bob" is very gregarious & hence friends with a great many people (however, suppose none of Them are Fields Medal recipients, say), and "Jane" is a bit of a recluse (i.e. small friend cadre) but she nevertheless happens To hobnob on a regular basis with several distinguished Fields Medal winners. Now, maybe you'd prefer not to get beers with Jane on the weekend - but still, whom would you say carries with them more (Mathematical) "prestige" in the network? This notion of awarding a centrality measure to a vertex in proportion to the sum of the scores of its neighbors is the key idea behind $\boxed{\text{eigenvector centrality}}$, as detailed below.

Sketch of the idea:

Begin by initializing all centrality values to one, $C(x_i)=1$ for all $x_i \in V(G)$.

Now, define an updated centrality, $C'(x_i)$, as the sum of the scores for x_i's neighbors, thus: $\boxed{C'(x_i) = \sum_j A_{ij} \cdot C(x_j)}$, where

A_{ij} is an element of the adjacency matrix. Equivalently, we may write this expression in matrix form as: $\vec{C}'(\vec{x}) = A\vec{C}(\vec{x})$, where $\vec{C}(\vec{x})$ is the vector of node scores.

If we repeat this process iteratively to achive better estimates of centralities, after T steps we have:

$$\boxed{\vec{C}(\vec{x}(t)) = A^T \vec{C}(\vec{x}(0))}$$

Now, since A is symmetric, it is guaranteed to be diagonizable.

Whence, we can write the initial score vector: $\vec{C}(\vec{x}(0))$ as a linear combination of the eigenvectors of A: $\vec{C}(\vec{x}(0)) = \sum_i c_i \vec{v}_i$ ($\vec{v}_i$: eigenvector).

This result yields the following formula:

$$\underbrace{\boxed{\vec{C}(\vec{x}(t))}}_{\text{Centrality of vert. @ time } t} = A^t \vec{C}(\vec{x}(0)) = A^t \sum_i c_i \underbrace{v_i}_{\text{eigenvectors}} = \underbrace{\sum_i c_i \lambda_i^t \vec{v}_i}_{\text{Recall! } A\vec{v}_i = \lambda_i \vec{v}_i}$$

$$= \boxed{\lambda_1^t \sum_i c_i \left(\frac{\lambda_i}{\lambda_1}\right)^t \vec{v}_i}$$, where λ_i's are the eigenvalues of A, and $\underline{\underline{\lambda_1}}$ is the maximum such eigenvalue.

But notice then, if we consider the <u>long-term behavior</u> of this iterative process of computing centrality measures, we get:

$$\underbrace{\lim_{t \to \infty}}_{\substack{\text{limit as time} \\ \text{"marches on"}}} \vec{C}(\vec{x}(t)) = \lim_{t\to\infty} \lambda_1^t \sum_i c_i \underbrace{\left(\frac{\lambda_i}{\lambda_1}\right)^t}_{\substack{(*) \text{ Tends to zero} \\ \text{for } \lambda_i \neq \lambda_1}} \vec{v}_i = \boxed{c_1 \lambda_1^t \vec{v}_1}$$

In matrix form then, the limit of the iterative updates of centrality yields the vector $\vec{x}$, satisfying:

$$\boxed{A\vec{x} = \lambda_1 \vec{x}}$$

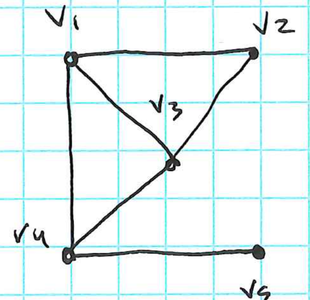
Key Formula!

Solving for vertex x_i gives us: $\boxed{C(x_i) = \frac{1}{\lambda_1} \sum_j A_{ij} \cdot C(x_j)}$.

As was promised, eigenvector centrality is therefore <u>proportional</u> <u>to the sum of the centralities of a vertex's neighbors.</u>

Quick sketch of why This worked:

$\boxed{\text{Ex.}}$ Consider The graph given below & associated adjacency matrix.



G

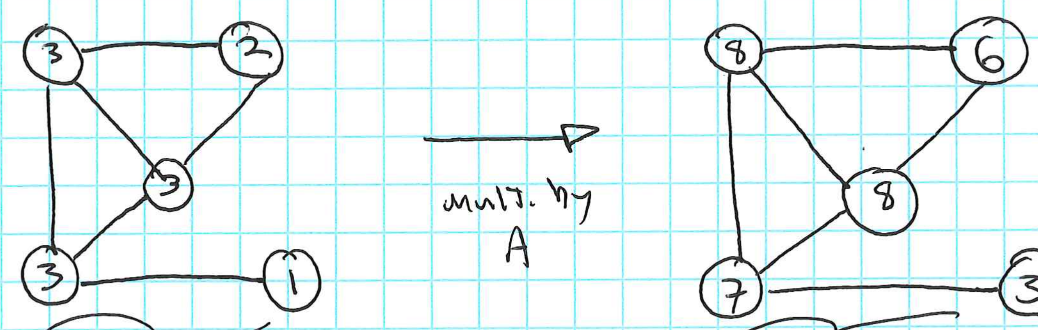
$$\longrightarrow \quad A(G) = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

Suppose we have an initial centrality vector, ordered as: $\begin{bmatrix} 3 \\ 2 \\ 3 \\ 3 \\ 1 \end{bmatrix} = \vec{x}$

Consider The multiplication:

$$A\vec{x} = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ 3 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 0\cdot3 + 1\cdot2 + 1\cdot3 + 1\cdot3 + 0\cdot1 \\ 1\cdot3 + 0\cdot2 + 1\cdot3 + 0\cdot3 + 0\cdot1 \\ 1\cdot3 + 1\cdot2 + 0\cdot3 + 1\cdot3 + 0\cdot1 \\ 1\cdot3 + 0\cdot2 + 1\cdot3 + 0\cdot3 + 1\cdot1 \\ 0\cdot3 + 0\cdot2 + 0\cdot3 + 1\cdot3 + 0\cdot1 \end{bmatrix} = \begin{bmatrix} 8 \\ 6 \\ 8 \\ 7 \\ 3 \end{bmatrix}$$

If we consider This calculation closely, we can see That The "effect" of This multiplication is To propagate centrality measures from a vertex's neighbors To itself.

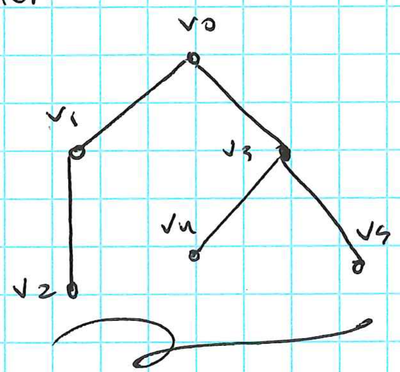


G with "$\vec{x}(0)$" label

G with "$\vec{x}(1)$" label

As we demonstrated earlier, this process converges (to a scalar multiple) of $\vec{v}_1$ (i.e. $A\vec{x} = \lambda_1 \vec{x}$).

$\boxed{\text{Ex.}}$ Given G below, we compute the eigenvector centrality of each vertex.



G

$$A(G) = \begin{array}{c} \begin{array}{cccccc} v_0 & v_1 & v_2 & v_3 & v_4 & v_5 \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \end{array}$$

For this particular matrix A, we have: $\lambda_1 \approx 1.902$, with $\vec{v}_1 = \begin{bmatrix} 1.618 \\ 1.176 \\ .618 \\ 1.902 \\ 1 \\ 1 \end{bmatrix}$

Thus, eigenvector centrality ranking for the graph G is as follows: $\boxed{v_3, v_0, v_1, v_4, v_5, v_2}$ (Ranked greatest to least).

IV Katz Centrality

In some instances, such as measures of centrality for digraphs, eigenvector centrality is problematic. For example, if a particular vertex has no incoming edges then its centrality is zero, and this zero centrality can propagate to neighboring vertices with incoming edges emanating from this vertex, leading to a great many erroneous zero centrality assignments.

To remedy this problem, we consider an addendum to eigenvector centrality known as **Katz Centrality.** We iteratively define the centrality measure of a vertex v_i as:

$$\boxed{x_i = \alpha \sum_j A_{ij} x_j + \beta}$$

Here the sum: $\sum_j A_{ij} x_j$ is identical the eigenvector centrality, with an added "attenuation" factor: $\boxed{\alpha}$, where α is chosen so that $\alpha < \frac{1}{\lambda_1}$, so that convergence is guaranteed.

In the formula, $\beta > 0$ is a shifting scalar so that all vertices are assigned non-zero values throughout the computation.

If $\beta = 1$, we can write the Katz centrality formula in terms of matrices as: $\vec{x} = \alpha A \vec{x} + \beta \vec{1}$, and solving for $\vec{x}$ we have: $\boxed{\vec{x} = (I_n - \alpha A)^{-1} \vec{1}}$.

We note that this formula yields an exact (as opposed to an iterative) solution, in approximately $O(n^3)$ time, due to the required computation of an inverse matrix.