Centrality Measures
Or, What's Important in This Network?

Dave Darmon

March 6, 2014

Centrality Measures

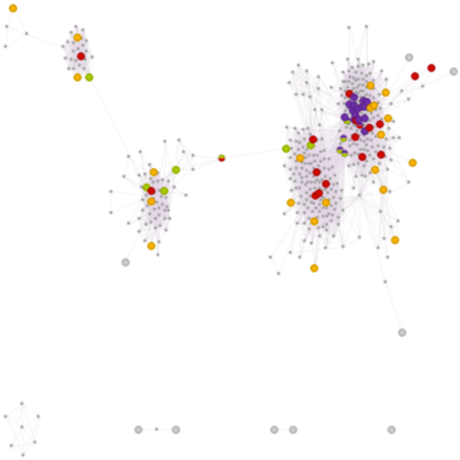
The Basic Idea, in Words

- Suppose we have a network, specified by its adjacency matrix $\mathbf{A}$ where

$$a_{ij} = \begin{cases} 1 & : \text{node } j \text{ points to node } i \\ 0 & : \text{otherwise} \end{cases} .$$

- Based on the network structure, we would like to rank the vertices $i = 1, 2, \ldots, n$ in terms of their importance or **centrality** to the network.
- 'Centrality' can mean anything we want it to, though its origins are in social network analysis.

A Social Network



A Gene Co-Expression Network

Centrality Measures

The Basic Idea, in Math

- Assign to each vertex ($i = 1, \ldots, n$) in the network a 'centrality' score x_i.
  - The centrality x_i is meant to capture how 'important' vertex i is.
  - Typically, 'importance' is determined by placement in the network.

Centrality Measures

The Basic Idea, in Math

- Assign to each vertex ($i = 1, \ldots, n$) in the network a 'centrality' score x_i.
  - The centrality x_i is meant to capture how 'important' vertex i is.
  - Typically, 'importance' is determined by placement in the network.
- **Notation:** Package all of the centralities in the vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Centrality Measures

A Simple Example – Degree Centrality

- Let k_i^{in} be the in-degree of vertex i.
  - The number of edges **pointing to** vertex i.
- We might then assume that a node is 'important' if it has high in-degree, in which case we define

$$x_i = k_i^{\text{in}}, \quad i = 1, \ldots, n.$$

Centrality Measures

A Simple Example – Degree Centrality

- Let k_i^{in} be the in-degree of vertex i.
  - The number of edges **pointing to** vertex i.
- We might then assume that a node is 'important' if it has high in-degree, in which case we define

$$x_i = k_i^{\text{in}}, \quad i = 1, \ldots, n.$$

- This assignment of importance, $\mathbf{x}^{\text{dc}} = \mathbf{A1}$, where $\mathbf{1}$ is the ones vector

$$\mathbf{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix},$$

corresponds to **degree centrality**.

Centrality Measures

Undirected vs. Directed Networks

- What happens if we multiply on the left by $\mathbf{1}$ instead of on the right? This gives us a (possibly different) degree centrality:

$$\mathbf{x}^{\text{dc}} = \mathbf{1}^T\mathbf{A}.$$

Centrality Measures

Undirected vs. Directed Networks

- What happens if we multiply on the left by $\mathbf{1}$ instead of on the right? This gives us a (possibly different) degree centrality:

$$\mathbf{x}^{\mathrm{dc}} = \mathbf{1}^T \mathbf{A}.$$

- We're summing across *columns* of $\mathbf{A}$, so this degree centrality is equivalent to the **out-degree** of each vertex,

$$x_j^{\mathrm{dc}} = \sum_{i=1}^{n} a_{ij}.$$

Centrality Measures

Undirected vs. Directed Networks

- What happens if we multiply on the left by $\mathbf{1}$ instead of on the right? This gives us a (possibly different) degree centrality:

$$\mathbf{x}^{\text{dc}} = \mathbf{1}^T \mathbf{A}.$$

- We're summing across *columns* of $\mathbf{A}$, so this degree centrality is equivalent to the **out-degree** of each vertex,

$$x_j^{\text{dc}} = \sum_{i=1}^{n} a_{ij}.$$

- This is a recurring theme: looking at left- and right-vectors gives *different* centralities.

Centrality Measures

Problems with Degree Centrality

- Degree centrality ranks a vertex i based on the number of edges pointing to it.
- But what if the vertices pointing to vertex i have a lot of vertices pointing to them?
- This motivates considering 'neighbors of neighbors' when computing the centrality of a vertex.

Centrality Measures

Fixing Degree Centrality: An Iterative Approach

- Start by assigning arbitrary (positive) centralities to each vertex, say $\mathbf{x}(0) = \mathbf{1}$, and then update the centralities using

$$
\begin{aligned}
x_i(n) &= \sum_{j=1}^{n} a_{ij} x_j(n-1), \quad n = 1, 2, \ldots \\
&= \underset{\sim}{\mathbf{a}_i} \mathbf{x}(n-1), \quad n = 1, 2, \ldots,
\end{aligned}
$$

Centrality Measures

Fixing Degree Centrality: An Iterative Approach

- Start by assigning arbitrary (positive) centralities to each vertex, say $\mathbf{x}(0) = \mathbf{1}$, and then update the centralities using

$$\begin{aligned} x_i(n) &= \sum_{j=1}^{n} a_{ij} x_j(n-1), \quad n = 1, 2, \ldots \\ &= \underset{\sim}{\mathbf{a}_i} \mathbf{x}(n-1), \quad n = 1, 2, \ldots, \end{aligned}$$

- **In words:** Make the centrality x_i of each vertex i the sum of the centralities x_j of the vertices j **pointing to** i.

Centrality Measures

Fixing Degree Centrality: An Iterative Approach

- Start by assigning arbitrary (positive) centralities to each vertex, say $\mathbf{x}(0) = \mathbf{1}$, and then update the centralities using

$$\begin{aligned} x_i(n) &= \sum_{j=1}^{n} a_{ij} x_j(n-1), \quad n = 1, 2, \dots \\ &= \underset{\sim}{\mathbf{a}_i} \mathbf{x}(n-1), \quad n = 1, 2, \dots, \end{aligned}$$

- **In words:** Make the centrality x_i of each vertex i the sum of the centralities x_j of the vertices j **pointing to** i.
- An 'exercise': How does this relate to degree centrality?

Centrality Measures

Fixing Degree Centrality: An Iterative Approach

- In vector notation,

$$x_i(n) = \sum_{j=1}^{n} a_{ij} x_j(n-1), \quad n = 1, 2, \ldots$$

becomes

$$\mathbf{x}(n) = \mathbf{A}\mathbf{x}(n-1), \quad n = 1, 2, \ldots$$

Centrality Measures

Fixing Degree Centrality: An Iterative Approach

- In vector notation,

$$x_i(n) = \sum_{j=1}^{n} a_{ij} x_j(n-1), \quad n = 1, 2, \ldots$$

  becomes

$$\mathbf{x}(n) = \mathbf{A}\mathbf{x}(n-1), \quad n = 1, 2, \ldots$$

- This is a linear system of difference equations.
  - A rich theory exists to linear systems of **difference** equations, most of it analogous to the theory for linear systems of **differential** equations.

Centrality Measures

Fixing Degree Centrality: An Iterative Approach

- Fortunately, we only care (why?) about the asymptotic behavior, that is what happens to $\mathbf{x}(n)$ as $n \to \infty$?

Centrality Measures

Fixing Degree Centrality: An Iterative Approach

- Fortunately, we only care (why?) about the asymptotic behavior, that is what happens to $\mathbf{x}(n)$ as $n \to \infty$?
- More importantly, we only care about the *direction* that $\mathbf{x}(n)$ points in, since we only care about the relative centrality of each vertex.

Centrality Measures

Fixing Degree Centrality: An Iterative Approach

- Fortunately, we only care (why?) about the asymptotic behavior, that is what happens to $\mathbf{x}(n)$ as $n \to \infty$?
- More importantly, we only care about the *direction* that $\mathbf{x}(n)$ points in, since we only care about the relative centrality of each vertex.
- Asymptotically,

$$\mathbf{x}(n) \approx c_1 \lambda_1^n \mathbf{v}_1$$

where c_1 is a constant depending on $\mathbf{x}(0)$, λ_1 is the largest positive eigenvalue of $\mathbf{A}$, and $\mathbf{v}_1$ is its associated eigenvector[1].

---

[1] For a strongly connected component, we are guaranteed such an eigenvalue-eigenvector pair exists by the Perron-Frobenius theorem for non-negative, irreducible matrices.

Centrality Measures

Eigenvector Centrality

- Asymptotically,

$$\mathbf{x}(n) \approx c_1 \lambda_1^n \mathbf{v}_1$$

  where c_1 is a constant depending on $\mathbf{x}(0)$, λ_1 is the largest positive eigenvalue of $\mathbf{A}$, and $\mathbf{v}_1$ is its associated eigenvector.
- Thus, the centrality $\mathbf{x}(n)$ points in the direction of the eigenvector associated with the dominant eigenvalue.
- Hence the name **eigenvector centrality**,

$$\mathbf{x}^{\text{ec}} = \mathbf{v}_1.$$

Centrality Measures

Eigenvector Centrality

- What does eigenvector centrality *mean*?
- Using the identity $\mathbf{A}\mathbf{x}^{\text{ec}} = \lambda_1 \mathbf{x}^{\text{ec}}$, we see that $\mathbf{x}^{\text{ec}} = \frac{1}{\lambda_1}\mathbf{A}\mathbf{x}^{\text{ec}}$, giving us

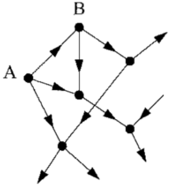
$$\begin{aligned} x_i^{\text{ec}} &= \frac{1}{\lambda_1}\sum_j a_{ij} x_j^{\text{ec}} \\ &\propto \sum_j a_{ij} x_j^{\text{ec}} \end{aligned}$$

- Thus, the eigenvector centrality of vertex i is proportional to the sum of the eigenvector centralities of the vertices incident on i.

Centrality Measures

Problems with Eigenvector Centrality

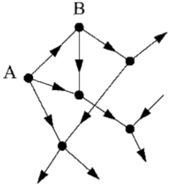
- What happens if a vertex i has **no** vertices pointing to it?



Centrality Measures

Problems with Eigenvector Centrality

- What happens if a vertex i has **no** vertices pointing to it?



- Then

$$x_i^{\mathrm{ec}} \propto \sum_j a_{ij} x_j^{\mathrm{ec}} = 0.$$

- The same is true for the eigenvector centrality of any vertex that is only pointed to by such an 'island' vertex.

Centrality Measures

Fixing Eigenvector Centrality

- We start again by assigning arbitrary (positive) centralities to each vertex, say $\mathbf{x}(0) = \mathbf{1}$, but now we update the centralities using

$$x_i(n) = \alpha \sum_{j=1}^{n} a_{ij} x_j(n-1) + \beta, \quad n = 1, 2, \ldots$$

Centrality Measures

Fixing Eigenvector Centrality

- We start again by assigning arbitrary (positive) centralities to each vertex, say $\mathbf{x}(0) = \mathbf{1}$, but now we update the centralities using

$$x_i(n) = \alpha \sum_{j=1}^{n} a_{ij} x_j(n-1) + \beta, \quad n = 1, 2, \ldots$$

- That is, we attribute an intrinsic centrality β to each vertex, independent of its placement in the network.
- **NOTE:** We have introduced *two* tuning parameters, α and β, that balance between pure eigenvector centrality and pure 'uniform' centrality.

Centrality Measures

Fixing Eigenvector Centrality

- In matrix-vector notation, we have

$$\mathbf{x}(n) = \alpha \mathbf{A}\mathbf{x}(n-1) + \beta \mathbf{1},$$

where $\mathbf{1}$ is the n-vector of all ones.

Centrality Measures

Fixing Eigenvector Centrality

- In matrix-vector notation, we have

$$\mathbf{x}(n) = \alpha \mathbf{A}\mathbf{x}(n-1) + \beta \mathbf{1},$$

  where $\mathbf{1}$ is the n-vector of all ones.
- Again, we only care about the asymptotic behavior of this system of difference equations.

Centrality Measures

Fixing Eigenvector Centrality

- Newman skips a lot of steps, and claims we want to solve

$$\mathbf{x} = \alpha \mathbf{A} \mathbf{x} + \beta \mathbf{1}$$

to find the limiting behavior of the update equation

$$\mathbf{x}(n) = \alpha \mathbf{A} \mathbf{x}(n-1) + \beta \mathbf{1}. \tag{1}$$

Centrality Measures

Fixing Eigenvector Centrality

- Newman skips a lot of steps, and claims we want to solve

  $$\mathbf{x} = \alpha \mathbf{A} \mathbf{x} + \beta \mathbf{1}$$

  to find the limiting behavior of the update equation

  $$\mathbf{x}(n) = \alpha \mathbf{A} \mathbf{x}(n-1) + \beta \mathbf{1}. \tag{1}$$

- This only makes sense if (1) has a limit point, which only occurs if all of the eigenvalues of $\mathbf{B} = \alpha \mathbf{A}$ are inside the unit circle.
- This becomes important in choosing α.

Centrality Measures

Fixing Eigenvector Centrality

- Solving

$$\mathbf{x} = \alpha \mathbf{A} \mathbf{x} + \beta \mathbf{1}$$

we get

$$\begin{aligned} \mathbf{x} &= (\mathbf{I} - \alpha \mathbf{A})^{-1} (\beta \mathbf{1}) = \beta (\mathbf{I} - \alpha \mathbf{A})^{-1} \mathbf{1} \\ &\propto (\mathbf{I} - \alpha \mathbf{A})^{-1} \mathbf{1}. \end{aligned}$$

Centrality Measures

Fixing Eigenvector Centrality

- Solving

$$\mathbf{x} = \alpha \mathbf{A}\mathbf{x} + \beta \mathbf{1}$$

we get

$$\begin{aligned} \mathbf{x} &= (\mathbf{I} - \alpha \mathbf{A})^{-1}(\beta \mathbf{1}) = \beta (\mathbf{I} - \alpha \mathbf{A})^{-1}\mathbf{1} \\ &\propto (\mathbf{I} - \alpha \mathbf{A})^{-1}\mathbf{1}. \end{aligned}$$

- Since we only care which *direction* the centrality vector points, we define

$$\mathbf{x}^{\mathrm{kc}}(\alpha) = (\mathbf{I} - \alpha \mathbf{A})^{-1}\mathbf{1},$$

the **Katz centrality**.

Centrality Measures

PageRank

- Katz centrality, like degree centrality and eigenvector centrality, has its own strengths and weaknesses.
- Another common alternative is to use the update equation

$$\mathbf{x}(n) = \alpha \mathbf{A}\mathbf{D}^{-1}\mathbf{x}(n-1) + \beta \mathbf{1},$$

where $\mathbf{D} = \mathrm{diag}(k_1^{\mathrm{out}}, \ldots, k_n^{\mathrm{out}})$ is the diagonal matrix with the out degree of each vertex along its diagonal.

Centrality Measures

PageRank

- Katz centrality, like degree centrality and eigenvector centrality, has its own strengths and weaknesses.
- Another common alternative is to use the update equation

$$\mathbf{x}(n) = \alpha \mathbf{A}\mathbf{D}^{-1}\mathbf{x}(n-1) + \beta \mathbf{1},$$

  where $\mathbf{D} = \mathrm{diag}(k_1^{\text{out}}, \ldots, k_n^{\text{out}})$ is the diagonal matrix with the out degree of each vertex along its diagonal.
- That is,

$$x_i(n) = \alpha \sum_j a_{ij} \frac{x_j(n-1)}{k_j^{\text{out}}} + \beta,$$

  so the impact of an incoming vertex j on i's centrality is dependent on the number of outgoing connections j makes.

Centrality Measures

PageRank

- The solution to

$$\mathbf{x} = \alpha \mathbf{A}\mathbf{D}^{-1}\mathbf{x} + \beta \mathbf{1},$$

namely

$$\begin{aligned}\mathbf{x}^{\mathrm{pr}} &= \beta \mathbf{D}(\mathbf{D} - \alpha \mathbf{I})^{-1}\mathbf{1} \\ &\propto \mathbf{D}(\mathbf{D} - \alpha \mathbf{I})^{-1}\mathbf{1},\end{aligned}$$

gives the PageRank centrality $\mathbf{x}^{\mathrm{pr}}$.

Centrality Measures

Which centrality measure to choose?

- We have covered 4 centrality measures:
  - Degree centrality
  - Eigenvector centrality
  - Katz centrality
  - PageRank centrality

  all of which assign centrality to a vertex i based in some way on the vertices pointing to it.

Centrality Measures

Which centrality measure to choose?

- We have covered 4 centrality measures:
  - Degree centrality
  - Eigenvector centrality
  - Katz centrality
  - PageRank centrality

  all of which assign centrality to a vertex i based in some way on the vertices pointing to it.
- Which centrality measure you use depends on what you're asking about the network.

Centrality Measures

Which centrality measure to choose?

- We have covered 4 centrality measures:
  - Degree centrality
  - Eigenvector centrality
  - Katz centrality
  - PageRank centrality

  all of which assign centrality to a vertex i based in some way on the vertices pointing to it.
- Which centrality measure you use depends on what you're asking about the network.
- The two most commonly used centrality measures are Eigenvector centrality and PageRank centrality.

Centrality Measures

Other Centrality Measures, or 7.5 to 7.7

- Closeness Centrality
  - What is the average geodesic distance between a vertex i and all other vertices in the network?
  - **Rationale:** If you can get to any other vertex in the network via vertex i quickly, then perhaps it is important.

Centrality Measures

Other Centrality Measures, or 7.5 to 7.7

- Closeness Centrality
  - What is the average geodesic distance between a vertex i and all other vertices in the network?
  - **Rationale:** If you can get to any other vertex in the network via vertex i quickly, then perhaps it is important.
- Betweenness Centrality
  - How many geodesic paths pass through vertex i?
  - **Rationale:** If a lot of the shortest paths pass through vertex i, perhaps it is important.

Centrality Measures

Other Centrality Measures, or 7.5 to 7.7

- Closeness Centrality
  - What is the average geodesic distance between a vertex i and all other vertices in the network?
  - **Rationale:** If you can get to any other vertex in the network via vertex i quickly, then perhaps it is important.
- Betweenness Centrality
  - How many geodesic paths pass through vertex i?
  - **Rationale:** If a lot of the shortest paths pass through vertex i, perhaps it is important.
- These are very different from the in-degree-based centralities.