Hubs and Authorities.

The HITS algorithm assigns each node in a network two measures.

- $x(v)$ is the authority centrality.
- $y(v)$ is the hub centrality.

Motivation:

Web pages can serve two purposes:

- authority: contains useful information that is likely to be consulted frequently.
- hub: contains links to pages that contain useful information.

These are related: good hubs should have (many?) links to good authorities. We can recognize an authoritative page because of all the authorities that link to it.

So we want $x: V(G) \to \mathbb{R}$ and $y: V(G) \to \mathbb{R}$ to satisfy.

$$x(v) = \alpha \cdot \sum_{w \in V(G)} a(w,u) \cdot y(w).$$

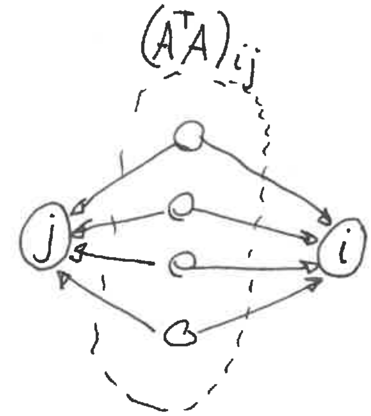
$$y(v) = \beta \cdot \sum_{w \in V(G)} a(v,w) \cdot x(w)$$

$$x = \alpha A^T y \qquad y = \beta A x$$

Combine: $\lambda x = A^T A x \quad \lambda y = A A^T y$, where $\lambda = 1/\alpha\beta$

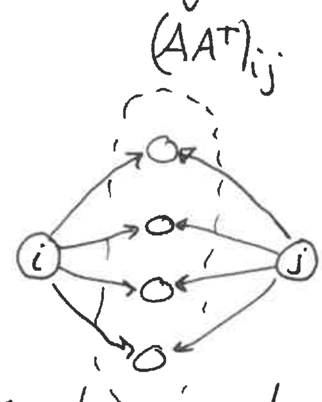
λ is an eigenvalue of AA^T and of A^TA.

- x is an eigenvector of A^TA
- y is an eigenvector of AA^T



$(A^TA)_{ij}$ = # ways to get from j to i by walking backward along one edge and then forward along another edge.

$= |N^-(j) \cap N^-(i)|$



$(AA^T) = |N^+(j) \cap N^+(i)|$

It is easy to cook up small examples where largest eigenvalue λ is not unique, but these don't seem to happen in large real-world networks.

$$C_a(v) = \frac{x(v)}{\max\{x(w) : w \in V(G)\}} \qquad C_h(v) = \frac{y(v)}{\max\{y(w) : w \in V(G)\}}$$

- Computation of Matrix-Based Centrality Measures.

~~Start~~

Usually we need to compute leading eigenvalue λ and corresponding eigenvector $\mathbf{c}$

$$\lambda c = A \cdot \mathbf{c}$$

Start with $\mathbf{c}^{(0)} = (\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n})$.

Iterate $$\mathbf{c}^{(k+1)} = \frac{A\mathbf{c}^{(k)}}{\underbrace{\|A\mathbf{c}^{(k)}\|}}$$

Any vector norm works.
Most easy is L_2 norm $\|\mathbf{v}\| = \sqrt{\sum_{i=1} (v_i)^2}$

- Each iteration requires doing the multiplication $A\mathbf{c}^{(k)}$.
- Trivial algorithm takes n^2 time is A is $n \times n$ matrix.
- Usually, A is sparse, so we can do this in $O(n + nz(A))$ time, where $nz(A)$ is the number of non-zero entries in matrix A.
- If A is the adjacency matrix of a directed graph G, then $nz(A) = |E(G)|$.
- Rate of convergence: $\|c - c^{(k)}\| = O((\lambda_2/\lambda_1)^k)$.
  λ_1 = largest eigenvalue.
  λ_2 = second largest eigenvalue

Distance-Based Centrality Measures

Closeness: $$C_C(v) = \frac{n-1}{\sum\limits_{w \in V(G)} \text{dist}_G(v,w)}.$$

Harmonic: $$C_H(v) = \frac{1}{n-1} \cdot \sum_{w \in V(G) \setminus \{v\}} \frac{1}{\text{dist}_G(v,w)}$$

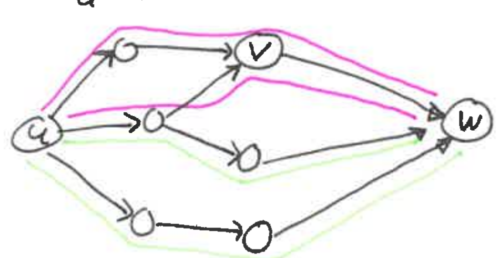
Eccentricity: $$C_E(v) = \frac{1}{\max\limits_{w \in V(G) \setminus \{v\}} \text{dist}_G(v,w)}$$

Betweenness: $$\frac{1}{(n-1)(n-2)} \sum_{u \in V(G) \setminus \{v\}} \sum_{w \in V(G) \setminus \{u,v\}} \frac{\ell(u,v,w)}{\ell(u,w)}$$

$\ell(u,w)$ = # shortest paths from u to w.

$\ell(u,v,w)$ = # shortest paths from u to w that include v.

$\text{dist}_G(u,w) = 3$



$\ell(u,w) = 4$

$\ell(u,v,w) = 2$

Efficiency.

For a connected graph G, the *efficiency* of G is

$$F(G) = \frac{1}{n(n-1)} \sum_{v \in V(G)} \sum_{w \in V(G) \setminus \{w\}} \left(\frac{1}{\mathrm{dist}_G(v,w)} \right)$$

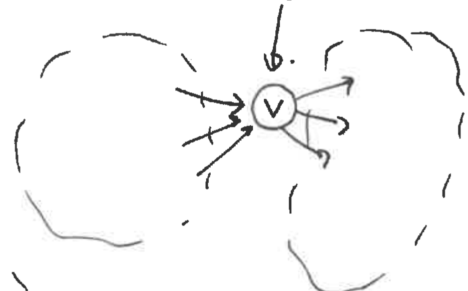
[If v and w are in different components then $\mathrm{dist}_G(v,w) = \infty$. and $\frac{1}{\mathrm{dist}_G(v,w)} = 0$. by convention

- Efficiency $= 1$ iff $\mathrm{dist}_G(v,w) = 1$ for each $\{v,w\} \in \binom{V(G)}{2}$, i.e. G is a clique.

- The Delta-centrality of a vertex v of G measures the drop in efficiency that occurs if we remove edges incident to v.

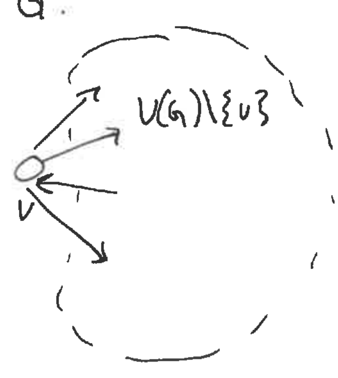
$$C_\Delta(v) = \frac{F(G) - F(G_v)}{F(G)}.$$

big $C_\Delta(v)$.

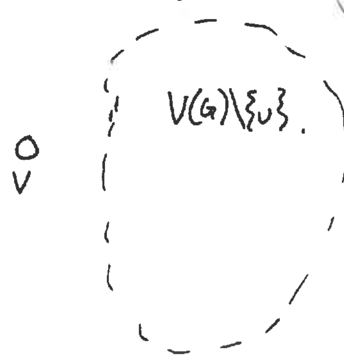


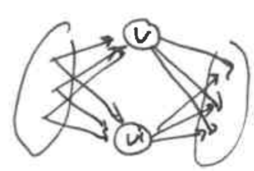
Eg

G.



G_v





$C_\Delta(v) \cancel{=0}$ is small.

$C_\Delta(v)$ is big when v is used in many shortest paths and for many of those paths there is no good alternative that avoids v.