8/31/20

- ☐ HW 1 due Friday
- ☐ HW 2 posted today.

Geodesic Paths.

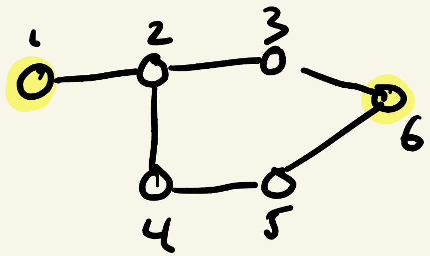
Path: seq of vertices

$x \to y \to \ldots \to z$

such that each pair is connected, i.e $i \to j$ $\Rightarrow$ i connected to j.

Geodesic: "Shortest Path"

shortest possible path between 2 vertices



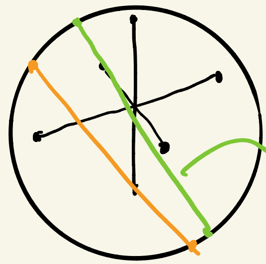
geodesic path.

$1 \to 2 \to 3 \to 6$

$1 \to 2 \to 4 \to 5 \to 6$ paths

Diameter: longest of all pairwise shortest paths.

intuition.



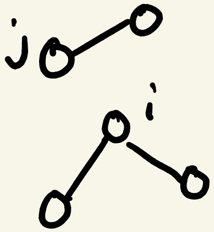
longest possible shortest path!

Dumb Algorithm:

A.P.S.P.

S.S.S.P

1. for all nodes i
2. and for all nodes j
3. compute length of S.P. $(i \to j)$
4. return $\max_{i,j} \{ SP(i \to j) \}$



Diameter?

Reachability.

If there is no path from i to j, then j is unreachable from i.

=> distance (in a path sense) is either undefined or infinite

<u>If</u> a network has multiple components, then let the diameter of the network be the diam. of its largest diameter component.

If the diameter of a network grows as $O(\log(n))$ then we say that the network has the **"small world"** property.

nodes n

Stanley Milgram.

#1 small worlds

#2 Authority, Pain, Electric Shock.

9/2/20

- ☐ HW1 due Fri @ Canvas
- ☐ HW2 posted
- ☐ Slack → #linkstream
  HW Q+A
- ☐ Wrap up Lec 1 Notes
  - Reciprocity
  - Clustering
- ☐ Centrality/Importance.

Defined by the fraction of links that are reciprocated.

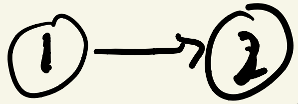
$$r = \frac{1}{m} \sum_{i,j} A_{ij} A_{ji}$$

recip: $A_{ij} = 1$ AND $A_{ji} = 1$

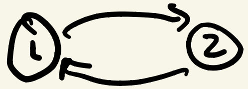
not: $A_{ij} = 1$ but $A_{ji} = 0$

Hint: think about indicator variables!

Ex:



$$\begin{array}{c|cc} & 1 & 2 \\ \hline 1 & 0 & 1 \\ 2 & 0 & 0 \end{array}$$

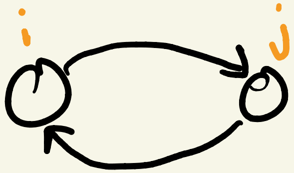


$$\begin{array}{c|cc} & 1 & 2 \\ \hline 1 & 0 & 1 \\ 2 & 1 & 0 \end{array}$$

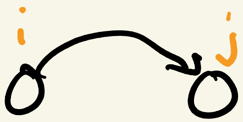
$$\frac{2}{2} = 1$$

Reciprocity (directed only)

Asks: are directed edges reciprocated?



reciprocated



not

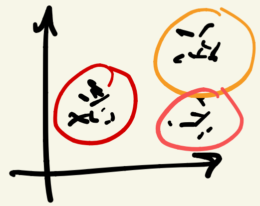
$$r = \frac{\sum_{i,j} A_{ij} A_{ji}}{\sum_{i,j} A_{ij}}$$

Trick! $M\vec{1} = \sum_j M_{ij} 1_j = \sum_j M_{ij}$ summing over j!

$\Rightarrow 1^T M$ ... summing over i

Clustering

What is the density of 3-cliques aka triangles?



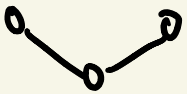
NO!

namespace is crowded!

$$C = \frac{\#\text{ of closed paths of } l=2}{\#\text{ of path of } l=2} \in [0,1]$$

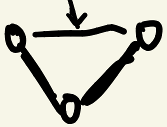
how often is this edge there to close the triangle?

path $l=2$



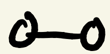
path $l=2$

closed

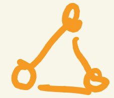


closed path $l=2$

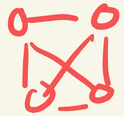
vocab: clique: a set of nodes connected "all to all"



2 clique



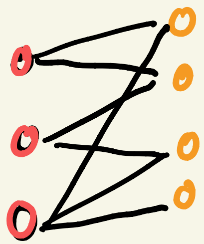
3 clique



4 clique

etc.

What is C for a bipartite network?



Always zero!

Why?

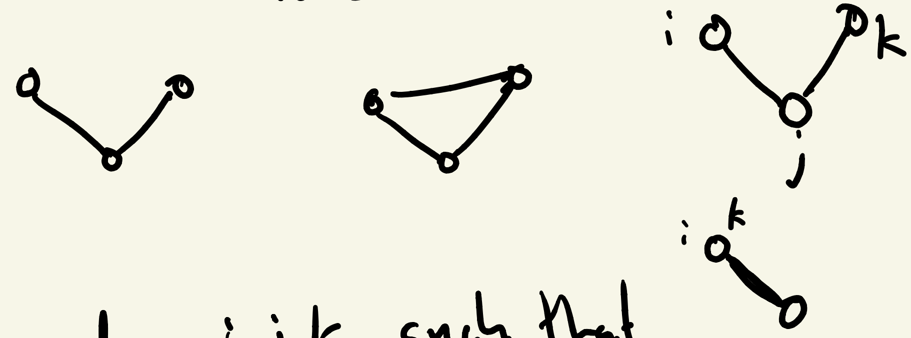
Because you can't connect two nodes of the same type!

$\Rightarrow$ no triangle can be closed

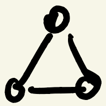
$\Rightarrow C = 0.$

Connected Triple:
Open Triad: Any i, j, k in which at least two pairs of nodes are connected.

Ex!

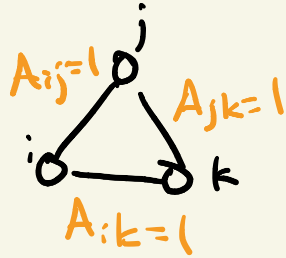


Closed Triad: i, j, k, such that all are connected.

$\Rightarrow$  a triangle.

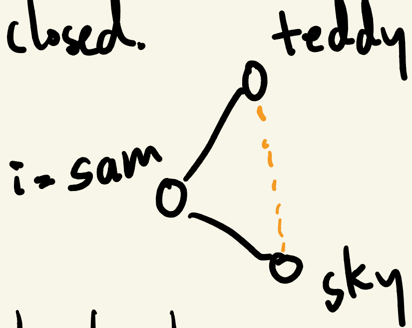
$$C = \frac{\text{\# triangles (or closed triads)} \times 3}{\text{\# connected triples}}$$

$$C = \frac{\sum_i \sum_j \sum_k A_{ij} A_{jk} A_{ki}}{\sum_i \sum_j \sum_{k \neq i} A_{ij} A_{jk}}$$



NB: "Nota bene" note well!

We can also define a vertex-level clustering: C_i = fraction of vertex i's potential triangles that are closed.



* with each other!

Local clustering:
How often are my friends also friends?*

- Soc. Networks tend to have high clustering.

- Soc. Networks may or may not have high reciprocity.

- Soc. Networks, like "Biological" Networks are diverse!

see p19 Table of all measures.

What is the actual question?

(underneath "which vertices are important?")

"Centrality"

Which Vertices Are Important?

Which bear is best?

Not well defined.

Two ways that ppl answer this Q:

① structural importance.

- input: graph G
- output: vector $\vec{v}$ whose entries are the rank or importance of each node

② dynamical importance

- input: graph G
- output: vector $\vec{v}$ whose entries are the rank or importance of each node

Assumes a dynamics.

Degree Centrality:

$$v_i = k_i = \sum_j A_{ij}$$

$(\vec{v} = A\vec{1})$ bonus!

Version 1: $v_i^{(1)} = \sum_j A_{ij}$

like an indicator!

Version 2: $v_i^{(2)} = \sum_j A_{ij} v_j^{(1)}$

$$= \sum_j A_{ij} \sum_k A_{jk}$$

version t: $v_i^{(t)} = \sum_j A_{ij} v_j^{(t-1)}$

limit as $t \to \infty$

$$\lambda \vec{v} = A\vec{v}$$

v, our importance/centrality, is an eigenvector of A.

Eigenvector Centrality.

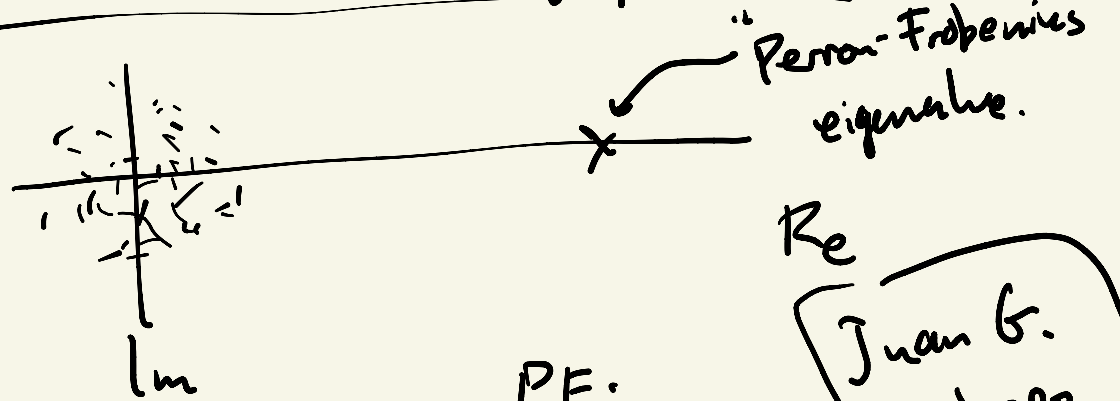
v_i = i^{th} entry of the eigenvector of A.

- $\vec{v}$ is the evec corresponding to λ, the largest eigenvalue of A. ("power method")
- evec cent ~ degree centrality.

left evec cent ~ out-degree centrality.
right evec cent ~ in-degree

be careful $A_{ij} = i \to j$ or $i \leftarrow j$

Q and A | ~~Spectrum~~ of random graph



PF:

complex => directed

- Re
- Unique
- larger in mag.
- positive, entrywise $v_i > 0$

Juan G. Restrepo
Ott 2007
Hunt
PRE

Req: $A_{ij} > 0$

Relaxation: Req: $A_{ij} \geq 0$

- lose uniqueness.
- $v_i \geq 0$
- larger in mag.

Let $\vec{x}$ be a random vector.

Let A be a network adj mtx.

Let $\{v^{(i)}\}_i$ be the set of A's eigenvectors, form a basis

$$Ax = A(\alpha_1 v^{(1)} + \alpha_2 v^{(2)} + \dots + \alpha_n v^{(n)})$$

$$\left[x = \alpha_1 v^{(1)} + \alpha_2 v^{(2)} + \dots + \alpha_n v^{(n)}\right]$$

$$= \alpha_1 A v^{(1)} + \alpha_2 A v^{(2)} + \dots \alpha_n A v^{(n)}$$

$$A v^{(i)} = \lambda_i v^{(i)}$$

$$= \alpha_1 \lambda_1 v^{(1)} + \alpha_2 \lambda_2 v^{(2)} + \dots \alpha_n \lambda_n v^{(n)}$$

$$A^2 x = \alpha_1 \lambda_1^2 v^{(1)} + \alpha_2 \lambda_2^2 v^{(2)} \dots$$

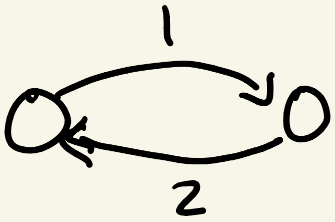
$$\frac{A^n x}{\lambda_1^n} = \alpha_1 \quad v^{(1)} + \alpha_2 \frac{\lambda_2^n}{\lambda_1^n} v^{(2)} + \dots$$

$$\left(\frac{\lambda_2}{\lambda_1}\right)^n \to 0 \text{ as } n \to \infty$$

Q and A

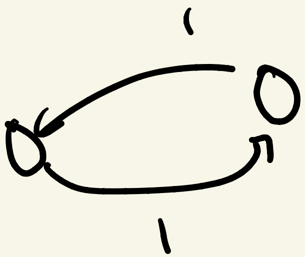
Can we use Recip. if the network is weighted?

What does reciprocity **mean** in a weighted network?



$$\begin{array}{cc} 1 & \begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix} \end{array}$$

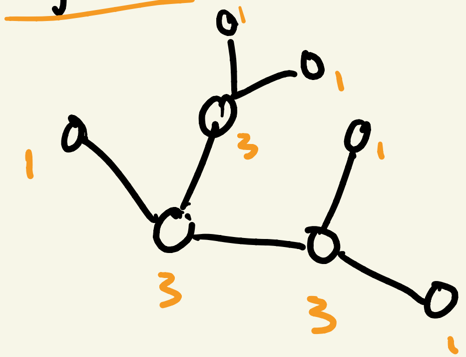
$$r = \frac{4}{3}$$



$$r = \frac{2}{2}$$

Office Hrs 9/3/20

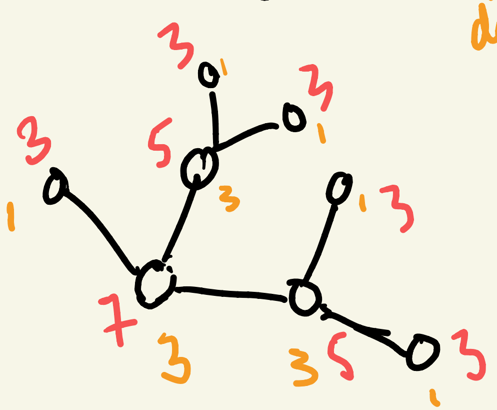
deg. cent. $v_i = k_i = \sum_j A_{ij}$



"double" degree centrality

$$x_i = \sum_j A_{ij} v_j$$

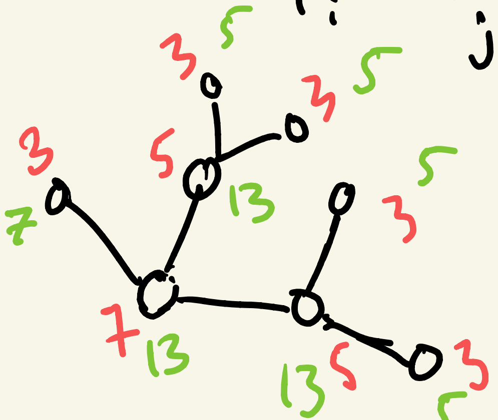
↑ degree centrality!



triple! degree centrality.

$$t_i = \sum_j A_{ij} x_j$$

↑ double degree cent.



$t_i = \sum_j A_{ij} x_j$ $\quad t = Ax \quad t = A A^2 1$, $t = A^3 1$

$x_i = \sum_j A_{ij} v_j$ $\quad x = Av \quad x = AA1$, $x = A^2 1$

$v_i = \sum_j A_{ij} 1_j$ $\quad v = A\vec{1}$

level q degree centrality... $c^{(q)} = A^q 1$

what about $q \to \infty$?

$$c^{(q+1)} \propto A c^{(q)}$$

$$c^{(q+1)} = \alpha A c^{(q)}$$

in the limit: $c = \alpha A c$
$q \to \infty$

$$\frac{1}{\alpha} c = Ac$$

$$Ac = \lambda c$$

an eigenvector (c)
eigenvalue (λ)
equation.

9/4/20

- HW1 due today → Canvas.
- HW2 posted
- Monday is a holiday!, No class.

"Centrality"

↓

① Degree. More important = Higher Degree.

Generalize.

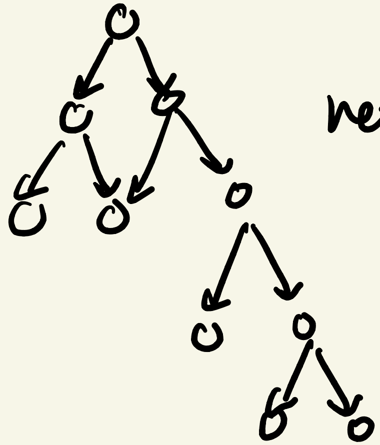
② Eigenvector Centrality.

More Important = Higher Entry in eigenvector $Av = \lambda v$

eval

evec

Eig. Centr. not so great for directed.



network w/ no cycles.

or nodes w/ zero out- or in-degree

↳ zero v_i

Lots of forms of Evec Centrality.

Page Rank. Larry Page.

↳ Google rank pages.

Page + Brin

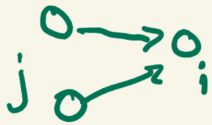
late 1990s.

papers.

every node gets a little centrality.

$$x_i = \beta + \alpha \sum_j \frac{A_{ij} x_j}{k_j^{out}}$$

add up the neighbor centralities.



$A_{ij} : j \to i$

$\max\{k_j^{out}, 1\}$

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

where D is a diagonal mtx with $D_{ii} = \max\{k_i^{out}, 1\}$

Typically $\beta = 1$

$$x = D(D - \alpha A)^{-1} \mathbf{1}$$

Alternative PageRank.

① Choose a node unif. at random.

② a) Follow a random link w.p. α

b) Get bored, start at a new page, chosen unif. at random, w.p. $1-\alpha$

Transition matrix.

Prob(go to i | I'm at j)

$$\Pi_{ij} = \frac{A_{ij}}{k_j^{out}} \qquad \text{mechanism a}$$

$$\frac{1}{N} \qquad \text{mechanism b}$$

$$Pr(\text{go to } i) = \alpha \sum_j \Pi_{ij} Pr(\text{at } j) + \frac{1-\alpha}{N}$$

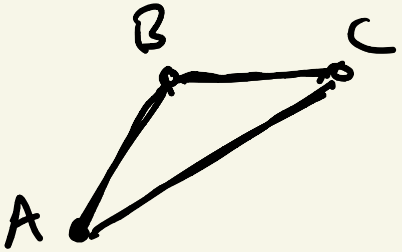
$$\vec{p} = \alpha \Pi \vec{p} + \left(\frac{1-\alpha}{N}\right)\mathbf{1}$$

$1-\alpha$ is sometimes called the "teleport" probability.

PR = Prob. that we find a random walker (web surfer) at page i, in steady state.

Geometric Centralities (paths!)

geodesics are not a distance metric. (No Δ inequality)



side AB   side BC   hypotenuse

$$d(A,B) + d(B,C) \overset{?}{\geq} d(A,C)$$

$\Rightarrow$ Applying your "Euclidean" intuition might lead you astray!

Closeness

most central point: the one closest to all others.

Let d_{ij} = geodesic distance $i \leftrightarrow j$

$$\ell_i = \frac{1}{n}\sum_{j=1}^{n} d_{ij}$$

$d_{ii} = 0$! divide by $\frac{1}{n-1}$?

yes... technically. But more conv. to ignore $\frac{n}{n-1}$ factor.

NB: this centrality is big when nodes are far...

Closeness: $$c_i = \frac{1}{\ell_i} = \frac{n}{\sum_j d_{ij}}$$

Problems w/ closeness:

1. multiple components
2. most networks have a small diam.

Recall: small world diam $= \Theta(\log n)$

$\Rightarrow$ values of C_i are in a narrow range. (sensitive to addition of a single edge)

Fixes + Patches for directed ...

Harmonic Centrality.

$$C_i = \frac{1}{n-1} \sum_{j \neq i} \frac{1}{d_{ij}}$$

Let $d_{ij} = \infty$ if $\nexists$ path from i to j. ← disconnected? no problem!

Between-ness Centrality.

Granovetter
"strength of weak ties"

Let's find bridge nodes!

1. map (write out) all the geodesic paths.
2. How many geodesics go through each node?

Let $b_i = \sum_{j,k} \#\{\text{geodesic } j \to \dots \to i \to \dots \to k\}$

$$\boxed{b_i = \sum_{j,k} \sigma_{jk}(i)}$$

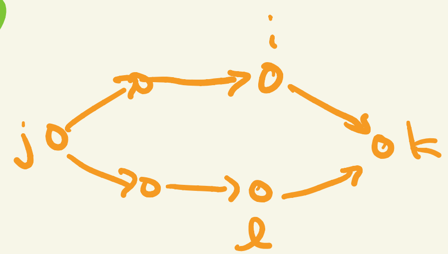
Alternative ①

$$b_i = \frac{\sum_{j,k} \#\{\text{geodesic } j \to \dots \to i \to \dots \to k\}}{\sum_{j,k} \#\{\text{geodesic } j \to \dots \to k\}}$$

$$\boxed{b_i = \frac{\sum_{j,k} \sigma_{jk}(i)}{\sum_{j,k} \sigma_{jk}}}$$

Define: $\frac{0}{0} = 0$

feature:



both i and l get equal credit.

Alternative ②

Normalize:

$$\boxed{b_i = \frac{1}{n^2} \sum_{jk} \frac{\sigma_{jk}(i)}{\sigma_{jk}}}$$

Note:

There is also an idea of edge betweeness! (roads, bottlenecks)