CENTRALITY MEASURES

Measure the “importance” of a node in a network.
The Oracle of Bacon

Charles Chaplin

Picture People No. 3: Hobbies of the Stars (1941)

John Beal (I)

The Firm (1993)

Margo Martindale

Rails & Ties (2007)

Kevin Bacon

Kevin Bacon to Charles chaplin

Find link  More options »
Hollywood Revolves Around

Click on a name to see that person's table.

Steiger, Rod (2.678695)
Lee, Christopher (I) (2.684104)
Hopper, Dennis (2.698471)
Sutherland, Donald (I) (2.701850)
Keitel, Harvey (2.705573)
Pleasence, Donald (2.707490)
von Sydow, Max (2.708420)
Caine, Michael (I) (2.720621)
Sheen, Martin (2.721361)
Quinn, Anthony (2.722720)
Heston, Charlton (2.722904)
Hackman, Gene (2.725215)
Connery, Sean (2.730801)
Stanton, Harry Dean (2.737575)
Welles, Orson (2.744593)
Mitchum, Robert (2.745206)
Gould, Elliott (2.746082)
Plummer, Christopher (I) (2.746427)
Coburn, James (2.746822)
Borgnine, Ernest (2.747229)
### Most Connected Actors in Hollywood
(measured in the late 90’s)

<table>
<thead>
<tr>
<th>Actor</th>
<th>Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mel Blanc</td>
<td>759</td>
</tr>
<tr>
<td>Tom Byron</td>
<td>679</td>
</tr>
<tr>
<td>Marc Wallice</td>
<td>535</td>
</tr>
<tr>
<td>Ron Jeremy</td>
<td>500</td>
</tr>
<tr>
<td>Peter North</td>
<td>491</td>
</tr>
<tr>
<td>TT Boy</td>
<td>449</td>
</tr>
<tr>
<td>Tom London</td>
<td>436</td>
</tr>
<tr>
<td>Randy West</td>
<td>425</td>
</tr>
<tr>
<td>Mike Horner</td>
<td>418</td>
</tr>
<tr>
<td>Joey Silvera</td>
<td>410</td>
</tr>
</tbody>
</table>

DEGREE CENTRALITY

$K = \text{number of links}$

\[ k_i = \sum_{j=1}^{n} A_{ij}. \]

Where $A_{ij} = 1$ if nodes $i$ and $j$ are connected and 0 otherwise.
BETWEENNESS CENTRALITY

BC = number of shortest Paths that go through a node.

BC(G) = 0
BC(D) = 9 + 7/2 = 12.5
BC(A) = 5 * 5 + 4 = 29
BC(B) = 4 * 6 = 24

A set of measures of centrality based on betweenness
LC Freeman - Sociometry, 1977 - jstor.org
CLOSENESS CENTRALITY

C = Average Distance to neighbors

\[ C(G) = \frac{1}{10} (1 + 2 \times 3 + 2 \times 3 + 4 + 3 \times 5) \]

\[ C(G) = 3.2 \]

\[ C(A) = \frac{1}{10} (4 + 2 \times 3 + 3 \times 3) \]

\[ C(A) = 1.9 \]

\[ C(B) = \frac{1}{10} (2 + 2 \times 6 + 2 \times 3) \]

\[ C(B) = 2 \]

N = 11
Consider the adjacency matrix $A_{ij} = 1$ if node $i$ is connected to node $j$ and 0 otherwise. Now, measure the centrality of a node, as the sum over the centralities of all nodes.

\[ x_i = \frac{1}{\lambda} \sum_{j \in M(i)} x_j = \frac{1}{\lambda} \sum_{j=1}^{N} A_{i,j} x_j \]

This is equivalent to eigenvalue problem:

\[ Ax = \lambda x \]

Then the eigenvector centrality of node (i) is defined as:

\[ x_i \]

where $\lambda$ is the largest eigenvalue associated with $A$ and $x$ is its associated eigenvector.
EIGENVECTOR CENTRALITY

Adjacency Matrix

A  -0.5727  -0.2829  -0.3286  -0.4357  -0.4357  -0.3286  -0.1297  -0.0423  -0.0423  -0.0423  -0.0423
B  -0.2829  -0.3286  -0.4357  -0.4357  -0.3286  -0.1297  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423
C  -0.3286  -0.4357  -0.4357  -0.3286  -0.1297  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423
D  -0.4357  -0.4357  -0.3286  -0.1297  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423
E  -0.3286  -0.3286  -0.4357  -0.1297  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423
F  -0.1297  -0.1297  -0.1297  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423
G  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423
H  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423
I  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423
J  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423
K  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423  -0.0423
Power and Centrality: A Family of Measures

Phillip Bonacich


\[ c(\alpha, \beta) = \alpha \sum_{k=0}^{\infty} \beta^k R^{k+1} 1 = \alpha(R1 + \beta R^2 1 + \beta^2 R^3 1 + \ldots). \]  

(5)
\[ c(\alpha, \beta) = \alpha \sum_{k=0}^{\infty} \beta^k R^{k+1} = \alpha (R1 + \beta R^2 1 + \beta^2 R^3 1 + \ldots). \]

\[ \beta = 0.5 \text{ (after 3 iterations)} \]
\[ \begin{align*}
A \rightarrow 22.5 & \quad K \rightarrow 5 \\
B \rightarrow 11.5 & \quad E \rightarrow 15.5 \\
H \rightarrow 15.5 \\
\end{align*} \]

\[ \beta = -0.5 \text{ (after 3 iterations)} \]
\[ \begin{align*}
A \rightarrow 10.5 & \quad K \rightarrow 0 \\
B \rightarrow 1.5 & \quad E \rightarrow 5.5 \\
H \rightarrow 9.5 \\
\end{align*} \]
\[ PR(A) = \frac{PR(B)}{4} + \frac{PR(C)}{3} + PR(D) + \frac{PR(E)}{2} \]

A random surfer eventually stops clicking

\[ PR(X) = \frac{(1-d)}{N} + d \left( \sum PR(y)/k(y) \right) \]
PAGE RANK

PR=Probability that a random Walker would visit that node. PR=Each page votes for its neighbors.

\[ R = \begin{bmatrix} PR(p_1) \\ PR(p_2) \\ \vdots \\ PR(p_N) \end{bmatrix} \]

\[ R = \begin{bmatrix} (1 - d)/N \\ (1 - d)/N \\ \vdots \\ (1 - d)/N \end{bmatrix} + d \begin{bmatrix} \ell(p_1, p_1) & \ell(p_1, p_2) & \cdots & \ell(p_1, p_N) \\ \ell(p_2, p_1) & \cdots & \ddots & \vdots \\ \vdots & \ddots & \ell(p_i, p_j) \\ \ell(p_N, p_1) & \cdots & \ell(p_N, p_N) \end{bmatrix} \]

\[ \sum_{i=1}^{N} \ell(p_i, p_j) = 1, \]
CLUSTERING MEASURES

Measure the density of a group of nodes in a Network
Clustering Coefficient, Transitivity

$$C_i = \frac{2\Delta}{k(k-1)}$$

$$C_A = \frac{2}{12} = \frac{1}{6}$$

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

$$C_E = \frac{4}{6} = \frac{2}{3}$$
Topological Overlap
Mutual Clustering

\[ TO(A,B) = \frac{\text{Overlap}(A,B)}{\text{NormalizingFactor}(A,B)} \]

\[ TO(A,B) = \frac{N(A,B)}{\max(k(A),k(B))} \]
\[ TO(A,B) = \frac{N(A,B)}{(k(A) \times k(B))^{1/2}} \]
\[ TO(A,B) = \frac{N(A,B)}{k(A)+k(B)} \]
Topological Overlap
Mutual Clustering

TO(A,B) = \frac{N(A,B)}{\max(k(A),k(B))}

TO(A,B) = 0
TO(A,D) = \frac{1}{4}
TO(E,D) = \frac{2}{4}
Global Measures
The Distribution of any of the previously introduced measures
Superfamilies of Evolved and Designed Networks

Ron Milo, Shalev Itzkovitz, Nadav Kashtan, Reuven Levitt, Shai Shen-Orr, Inbal Ayzenshtat, Michal Sheffer, Uri Alon

Triad Significance Profile

Normalized Z score

TRANSC-E.COLI
TRANSC-YEAST
TRANSC-YEAST-2
TRANSC-B.SUBTILIS

SIGNAL-TRANSDUCTION
TRANSC-DROSOPHILA
TRANSC-SEA-URCHIN
NEURONS

WWW-1 N=325,729
WWW-2 N=277,114
WWW-3 N=47,870
SOCIAL-1 N=67
SOCIAL-2 N=28
SOCIAL-3 N=32

LANGUAGES: ENGLISH
FRENCH
SPANISH
JAPANESE
BIPARTITE MODEL

subgraphs
Giant Component

Components

S = NumberOfNodesInGiantComponent / TotalNumberOfNodes
Diameter
Diameter = Maximum Distance Between Elements in a Set

Diameter = D(G,J) = D(C,J) = D(G,I) = ... = 5
Average Path Length

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ D(1)=8 \]
\[ D(2)=9 \]
\[ D(3)=4 \]

\[ L=(8+2\times9+3\times4)/(8+9+4) = 1.8 \]
Degree Correlations
Are Hubs Connected to Hubs?
Dynamical and correlation properties of the Internet

Romualdo Pastor-Satorras,¹ Alexei Vázquez,² and Alessandro Vespignani³

Assortative Mixing in Networks

M. E. J. Newman

Department of Physics, University of Michigan, Ann Arbor, MI 48109–1120
Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501

(Received 20 May 2002; published 28 October 2002)

A network is said to show assortative mixing if the nodes in the network that have many connections tend to be connected to other nodes with many connections. Here we measure mixing patterns in a variety of networks and find that social networks are mostly assortatively mixed, but that technological and biological networks tend to be disassortative. We propose a model of an assortatively mixed network, which we study both analytically and numerically. Within this model we find that networks percolate more easily if they are assortative and that they are also more robust to vertex removal.

the normalized correlation function is

\[ r = \frac{1}{\sigma^2} \sum_{jk} jk (e_{jk} - q_j q_k), \]  

(3)

which is simply the Pearson correlation coefficient of the degrees at either ends of an edge and lies in the range \(-1 \leq r \leq 1\). For the practical purpose of evaluating \( r \) on
<table>
<thead>
<tr>
<th>Network</th>
<th>$n$</th>
<th>$r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physics coauthorship (a)</td>
<td>52909</td>
<td>0.363</td>
</tr>
<tr>
<td>Biology coauthorship (a)</td>
<td>1520251</td>
<td>0.127</td>
</tr>
<tr>
<td>Mathematics coauthorship (b)</td>
<td>253339</td>
<td>0.120</td>
</tr>
<tr>
<td>Film actor collaborations (c)</td>
<td>449913</td>
<td>0.208</td>
</tr>
<tr>
<td>Company directors (d)</td>
<td>7673</td>
<td>0.276</td>
</tr>
<tr>
<td>Internet (e)</td>
<td>10697</td>
<td>−0.189</td>
</tr>
<tr>
<td>World-Wide Web (f)</td>
<td>269504</td>
<td>−0.065</td>
</tr>
<tr>
<td>Protein interactions (g)</td>
<td>2115</td>
<td>−0.156</td>
</tr>
<tr>
<td>Neural network (h)</td>
<td>307</td>
<td>−0.163</td>
</tr>
<tr>
<td>Marine food web (i)</td>
<td>134</td>
<td>−0.247</td>
</tr>
<tr>
<td>Freshwater food web (j)</td>
<td>92</td>
<td>−0.276</td>
</tr>
<tr>
<td>Random graph (u)</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Callaway et al. (v)</td>
<td></td>
<td>$\delta/(1+2\delta)$</td>
</tr>
<tr>
<td>Barabási and Albert (w)</td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

What’s a problem here?