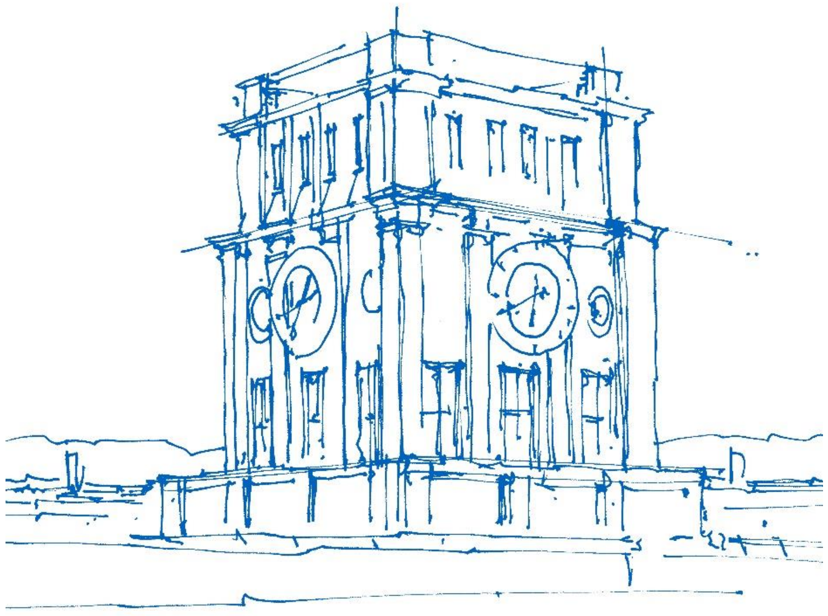
Zentralitätsmaße

Jürgen Pfeffer

Technical University of Munich

Bavarian School of Public Policy

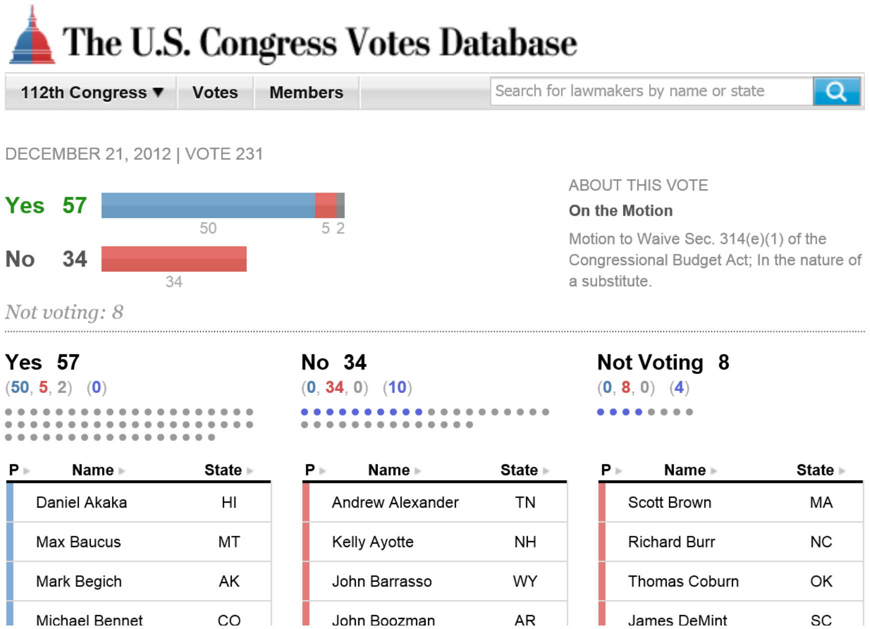
juergen.pfeffer@tum.de | @JurgenPfeffer

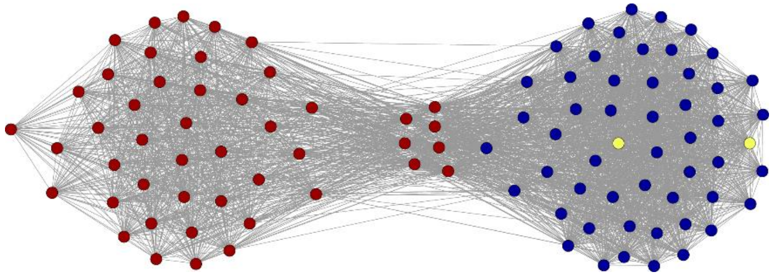


Uhrenturm der TUM

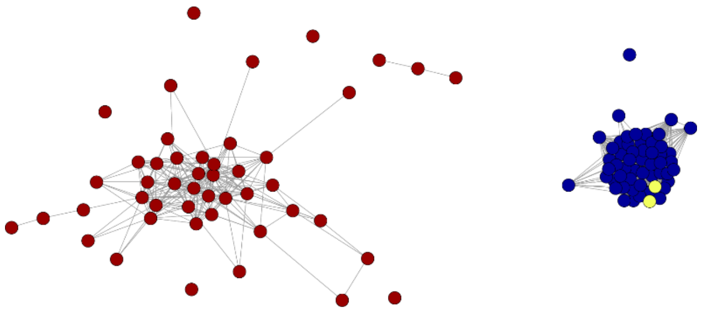
Social Network Analysis

112th Congress (2012), 100 Senators, 251 Vote





min 125 shared votes



min. 200 shared votes

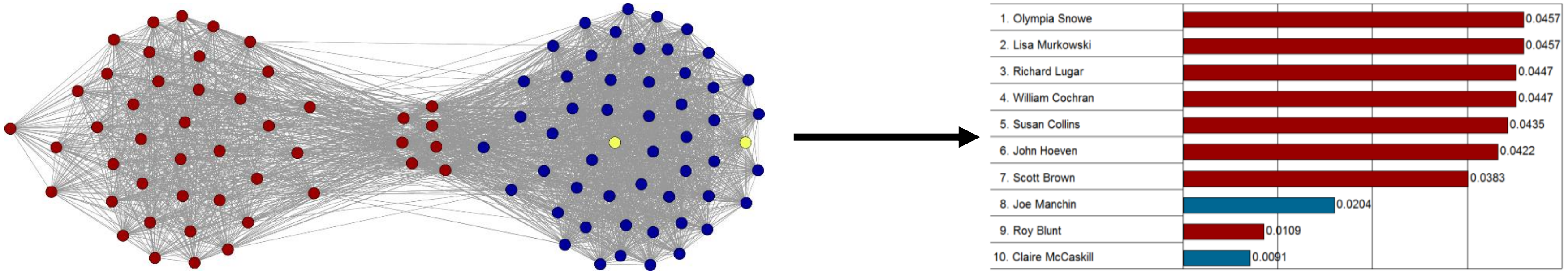
Network Metrics

Identify "important" nodes

Mapping of complex network structure to one value per node

Allows for rankings, comparisons, and other statistics

Different measures cover different interpretations of "importance"



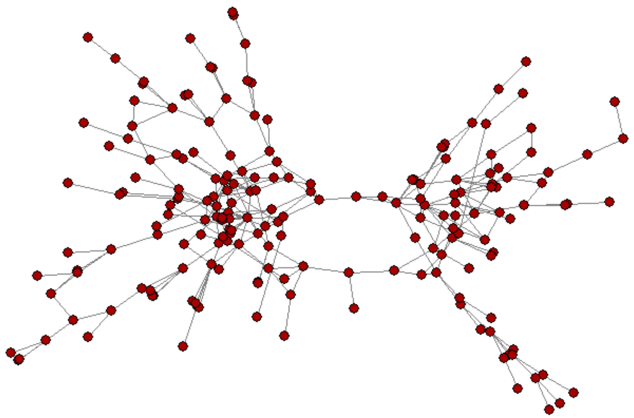
Node Level Metrics for Nodes

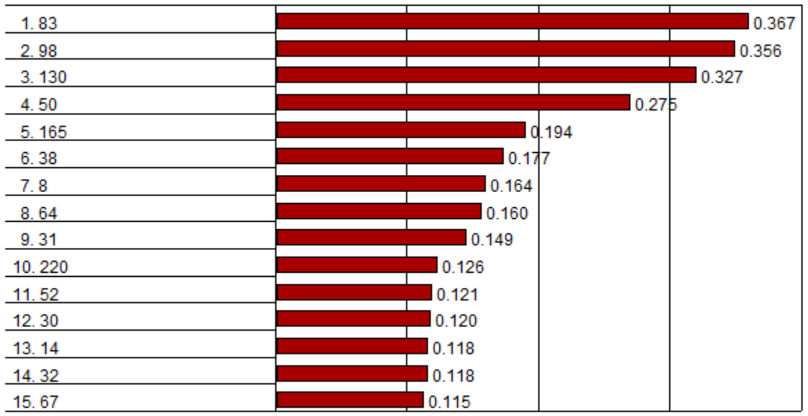
Centrality algorithms: Identify "important" nodes

Different metrics cover different interpretations of "importance":

- *Number of direct connections of a node (Freeman, 1979)*
- *All paths of length ≥ 2 (Stephenson and Zelen, 1989)*
- *Shortest paths (Anthonisse, 1971; Freeman, 1977, 1979)*
- *Neighbors of neighbors of … (Bonacich, 1972)*

Mapping to a one-value-for-each-node vector



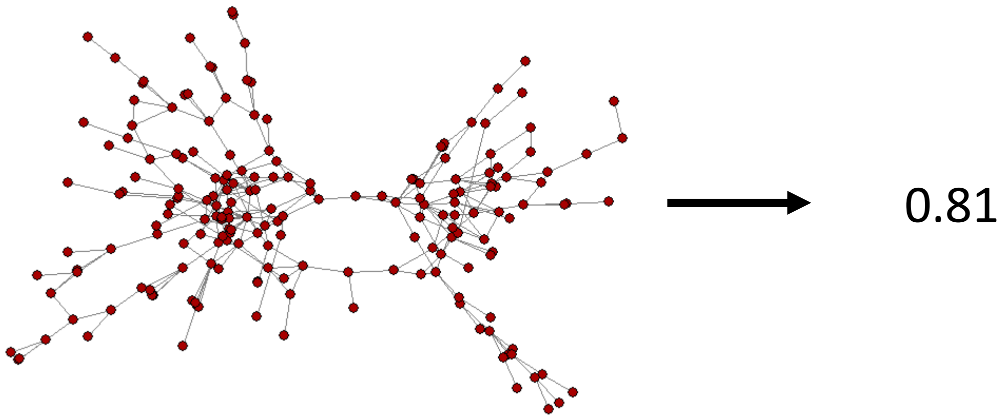


Network Level Metrics for Networks

A single value for a network

- Size, Density
- Centralization, distributions
- Fragmentation, modularity
- Path distances

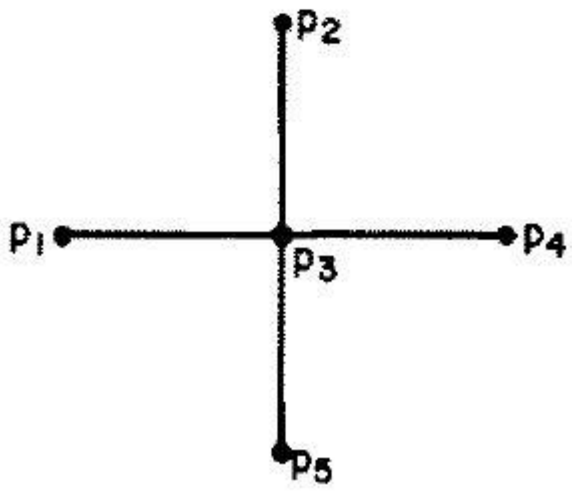
Compare networks

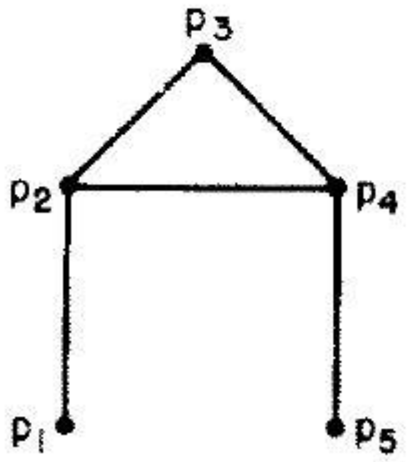


Zentralitätsmaße – Was ist “zentral”?

What is Central in a Network?

“The point at the center of a star or wheel […] is the most central position”





Freeman, L. C. (1979). Centrality in social networks: Conceptual clarification. *Social Networks*, *1*(3), 215–239.

What is Central? 3 Definitions

The central point has the largest degree

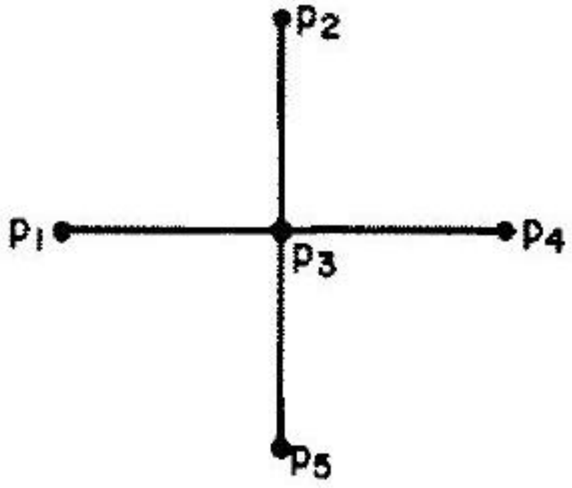
- Well connected to many other nodes

It is on the most geodesic paths

- Involved in many inter-network communications

It is as close to all points as possible

- Short path lengths



Definition 1: Largest Degree

Motivation

- Most contacts, highest activity

Simply measure the degree of each point

- Higher degree -> more central

$$C_D(p_k) = \sum_{i=1}^{n} a(p_i, p_k)$$

Normalization

Compare networks with different size

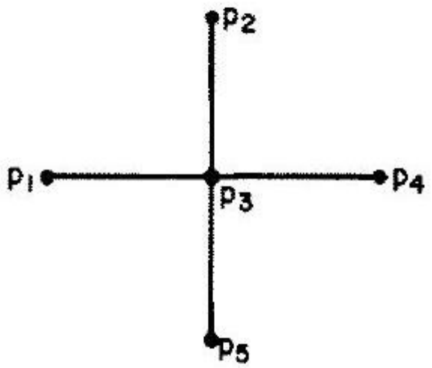
Normalize in range [0..1]

Divide by maximum possible value

Degree Centrality:

$$C_D(p_k) = \sum_{i=1}^{n} a(p_i, p_k)$$

$$C'_D(p_k) = \frac{C_D(p_k)}{n-1}$$



Definition 2: Most Geodesics (shortest paths)

Motivation: A person is central if he lies on many paths of communication

- Can withhold or distort information, control

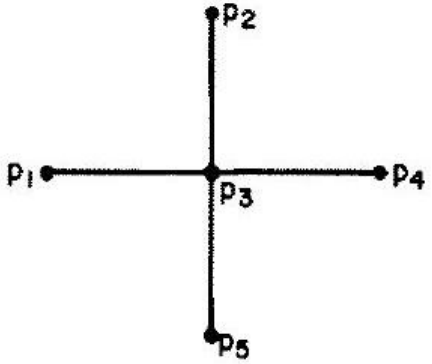
Number of shortest paths from *i* to *j* including *a*

Number of shortest paths from *i* to *j*

Betweenness Centrality:

$$C^B(n_a) = \sum_{i<j} \frac{g_{ij}(n_a)}{g_{ij}}$$

$$C^B_{max} = \frac{(N^2 - 3N + 2)}{2}$$



Definition 3: Shortest distance to all other points

Motivation: Being close to everybody else is good

- Close to information, resources, etc.

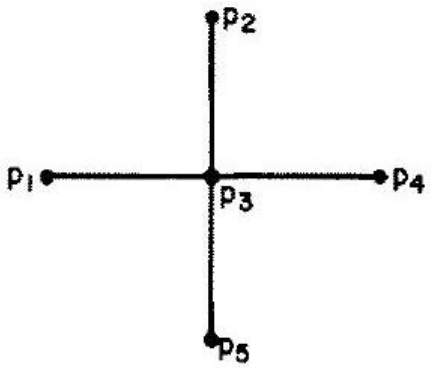
Sum up all geodesic paths from the point (k) to all other points

- Can't be calculated in unconnected graphs

Closeness Centrality:

$$C^C(n_a) = \frac{1}{\sum_{i \neq a} d(n_a, n_i)}$$

$$C^C_{max} = \frac{1}{(N-1)}$$

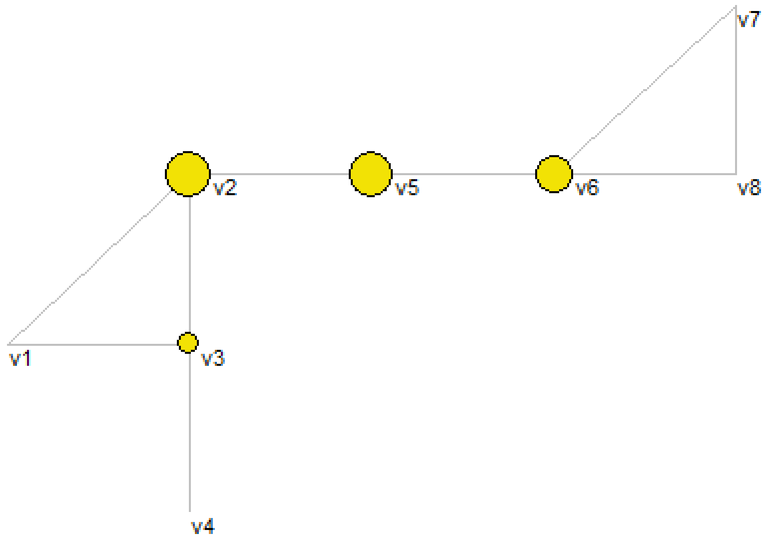


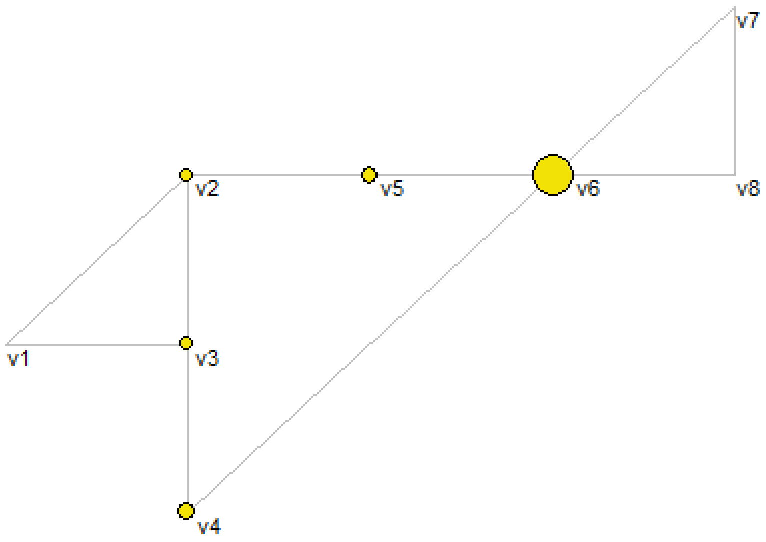
Robustness of Metrics

Shortest path metrics can change significantly with a single link

Positive: Sensitive to structural change

Negative: Not robust in case of missing data





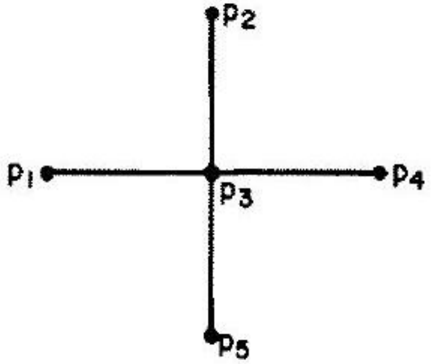
Centralization

How centralized is the network

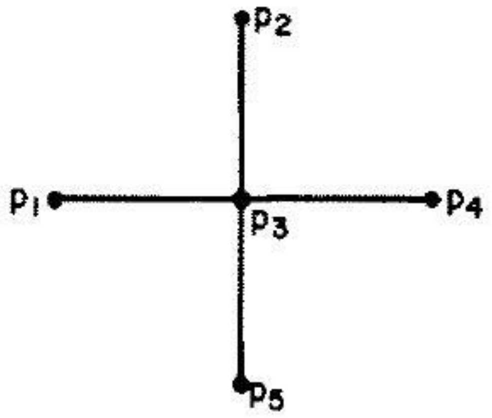
- 1 if one element completely dominates
- 0 if no point is any more central than any other

Normalize by dividing by the largest possible value of the quantity

$$C_x = \frac{\sum_{i=1}^{n} [C_X(p^*) - C_X(p_i)]}{\max \sum_{i=1}^{n} [C_X(p^*) - C_X(p_i)]}$$



Maximum Centralization



Degree centrality: $n^2 - 3n + 2$

Betweenness Centrality: $n^3 - 4n^2 + 5n - 2$

Closeness Centrality: $(n^2 - 3n + 2)/(2n - 3)$

$$C_x = \frac{\sum_{i=1}^{n} [C_X(p^*) - C_X(p_i)]}{\max \sum_{i=1}^{n} [C_X(p^*) - C_X(p_i)]}$$

Eigenvector Centrality

Eigenvector centrality is based on eigenvector calculation in linear algebra.

Agents have a high eigenvector score if they are important and connected to other important agents (Bonacich, 1972)

$$C_E\,(u) = \frac{1}{\lambda}\sum_{v=1}^{|V|} w_{u,v} C_E(U)$$

where λ is a constant. We can rewrite the equation as:

$$\lambda C_E = W \cdot C_E$$

Hubs and Authorities (HITS)

Important authorities are nodes at which important hubs point to:

$$x_i = \sum_j A_{ij} y_j$$
$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{y}$$

Important hubs are nodes that point to important authorities:

$$y_i = \sum_j A_{ji} x_j$$
$$\boldsymbol{y} = \boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{x}$$

Hubs and Authorities are the eigenvectors of $\mathbf{A}\mathbf{A}^T$ and $\mathbf{A}^T\mathbf{A}$ with the same eigenvalues.

Hubs and Authorities (HITS)