

# Centrality Measures

I. Makarov & L.E. Zhukov

**BigData Academy MADE from VK**

**Social Network Analysis and Machine Learning on Graphs**



# Lecture outline

- 1 Notion of centrality
- 2 Graph-theoretic measures
- 3 Node centralities
  - Degree centrality
  - Closeness centrality
  - Betweenness centrality
  - Eigenvector centrality
  - Katz and Bonacich centralities
- 4 Rank correlation

# Centrality

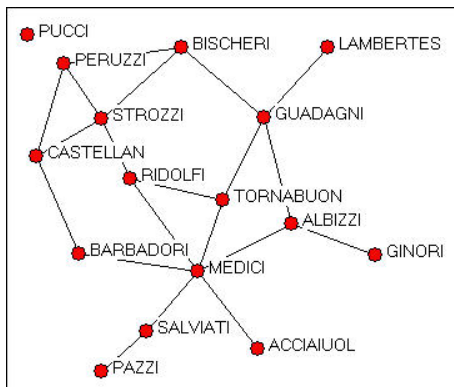
Which vertices are important?



M. Grandjean, 2014

# Centrality Measures

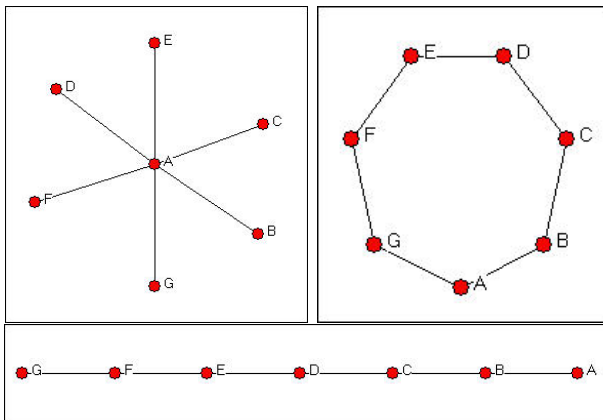
Sociology: determine the most "important" or "prominent" actors in the network based on actor location, involvement with other actors



Marriage alliances among leading Florentine families 15th century.

Padgett, 1993

# Three graphs



Star graph

Circle graph

Line Graph

# Degree centrality

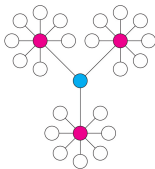
Degree centrality: number of nearest neighbours

$$C_D(i) = k(i) = \sum_j A_{ij} = \sum_j A_{ji}$$

Normalized degree centrality

$$C_D^*(i) = \frac{1}{n-1} C_D(i) = \frac{k(i)}{n-1}$$

High centrality degree -direct contact with many other actors



# Closeness centrality

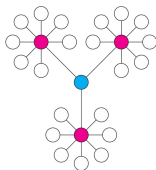
Closeness centrality: how close an actor to all the other actors in network

$$C_C(i) = \frac{1}{\sum_j d(i,j)}$$

Normalized closeness centrality

$$C_C^*(i) = (n-1)C_C(i) = \frac{n-1}{\sum_j d(i,j)}$$

High closeness centrality - short communication path to others, minimal number of steps to reach others



[\*\*\* Harmonic centrality  $C_H(i) = \sum_j \frac{1}{d(i,j)}$  \*\*\*]

Alex Bavelas, 1948

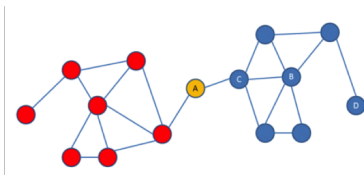
# Betweenness centrality

Betweenness centrality: number of shortest paths going through the actor  $\sigma_{st}(i)$

$$C_B(i) = \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

Normalized betweenness centrality

$$C_B^*(i) = \frac{2}{(n-1)(n-2)} C_B(i) = \frac{2}{(n-1)(n-2)} \sum_{s \neq t \neq i} \frac{\sigma_{st}(i)}{\sigma_{st}}$$



High betweenness centrality - vertex lies on many shortest paths

Probability that a communication from  $s$  to  $t$  will go through  $i$  (geodesics)

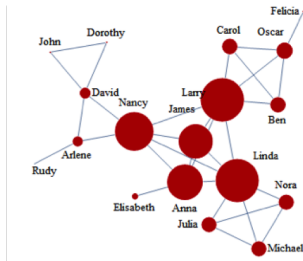
Linton Freeman, 1977



# Eigenvector centrality

Importance of a node depends on the importance of its neighbors  
(recursive definition)

$$v_i \leftarrow \sum_j A_{ij} v_j$$
$$v_i = \frac{1}{\lambda} \sum_j A_{ij} v_j$$
$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$$



Select an eigenvector associated with largest eigenvalue  $\lambda = \lambda_1$ ,  $\mathbf{v} = \mathbf{v}_1$

Phillip Bonacich, 1972.

# Centrality examples

## Closeness centrality



from [www.activenetworks.net](http://www.activenetworks.net)

# Centrality examples

## Betweenness centrality



from [www.activenetworks.net](http://www.activenetworks.net)

# Centrality examples

## Eigenvector centrality



from [www.activenetworks.net](http://www.activenetworks.net)

# Katz status index

Weighted count of all paths coming to the node: the weight of path of length  $n$  is counted with attenuation factor  $\beta^n$ ,  $\beta < \frac{1}{\lambda_1}$

$$k_i = \beta \sum_j A_{ij} + \beta^2 \sum_j A_{ij}^2 + \beta^3 \sum_j A_{ij}^3 + \dots$$

$$\mathbf{k} = (\beta \mathbf{A} + \beta^2 \mathbf{A}^2 + \beta^3 \mathbf{A}^3 + \dots) \mathbf{e} = \sum_{n=1}^{\infty} (\beta^n \mathbf{A}^n) \mathbf{e} = \left( \sum_{n=0}^{\infty} (\beta \mathbf{A})^n - \mathbf{I} \right) \mathbf{e}$$

$$\sum_{n=0}^{\infty} (\beta \mathbf{A})^n = (\mathbf{I} - \beta \mathbf{A})^{-1}$$

$$\mathbf{k} = ((\mathbf{I} - \beta \mathbf{A})^{-1} - \mathbf{I}) \mathbf{e}$$

$$(\mathbf{I} - \beta \mathbf{A}) \mathbf{k} = \beta \mathbf{A} \mathbf{e}$$

$$\mathbf{k} = \beta \mathbf{A} \mathbf{k} + \beta \mathbf{A} \mathbf{e}$$

$$\mathbf{k} = \beta (\mathbf{I} - \beta \mathbf{A})^{-1} \mathbf{A} \mathbf{e}$$

# Bonacich centrality

Two-parametric centrality measure  $c(\alpha, \beta)$

$\beta$  - radius of power,  $\alpha$  - normalization parameter,

$\beta > 0$  - tied to more central (powerful) people

$\beta < 0$  - tied to less central (powerful) people

$\beta = 0$  - degree centrality

$$c_i(\alpha, \beta) = \sum_j (\alpha + \beta c_j) A_{ij}$$

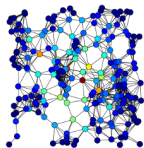
$$\mathbf{c} = \alpha \mathbf{A} \mathbf{e} + \beta \mathbf{A} \mathbf{c}$$

$$(\mathbf{I} - \beta \mathbf{A}) \mathbf{c} = \alpha \mathbf{A} \mathbf{e}$$

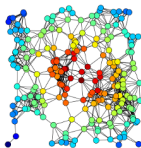
$$\mathbf{c} = \alpha (\mathbf{I} - \beta \mathbf{A})^{-1} \mathbf{A} \mathbf{e}$$

Normalizaton:  $\|\mathbf{c}\|_2 = \sum c_i^2 = 1$

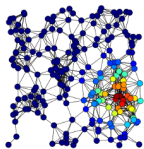
# Centrality examples



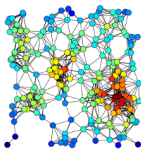
A



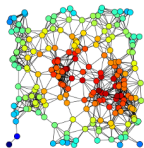
B



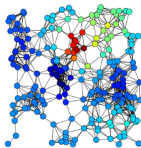
C



D



E



F

- A) Betweenness centrality
- B) Closeness centrality
- C) Eigenvector centrality
- D) Degree centrality
- E) Harmonic centrality
- F) Katz centrality

from Wikipedia

# Centralization

Centralization (network measure) - how central the most central node in the network in relation to all other nodes.

$$C_x = \frac{\sum_{i=1}^N [C_x(p_G^*) - C_x(p_G^i)]}{\max_{G'} \sum_{i=1}^N [C_x(p_{G'}^*) - C_x(p_{G'}^i)]}$$

$C_x$  - one of the centrality measures

$p_G^*$  - node in the graph  $G$  with the largest centrality value

max - is taken over all the graphs  $G'$  with the same number of nodes (for degree, closeness and betweenness the most centralized structure  $G'$  is the star graph)

Linton Freeman, 1979



- **Pearson correlation** coefficient

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

Shows linear dependence between variables,  $-1 \leq r \leq 1$   
(perfect when related by linear function)

- **Spearman rank correlation** coefficient (Sperman's rho):  
Convert raw scores to ranks - sort by score:  $X_i \rightarrow x_i$ ,  $Y_i \rightarrow y_i$

$$\rho = 1 - \frac{6 \sum_{i=1}^n (x_i - y_i)^2}{n(n^2 - 1)}$$

Shows strength of monotonic association  
(perfect for monotone increasing/decreasing relationship)

# Ranking comparison

- The Kendall tau rank distance is a metric that counts the number of pairwise disagreements between two ranking lists
- Kendall rank correlation coefficient, commonly referred to as Kendall's tau coefficient

$$\tau = \frac{n_c - n_d}{n(n-1)/2}$$

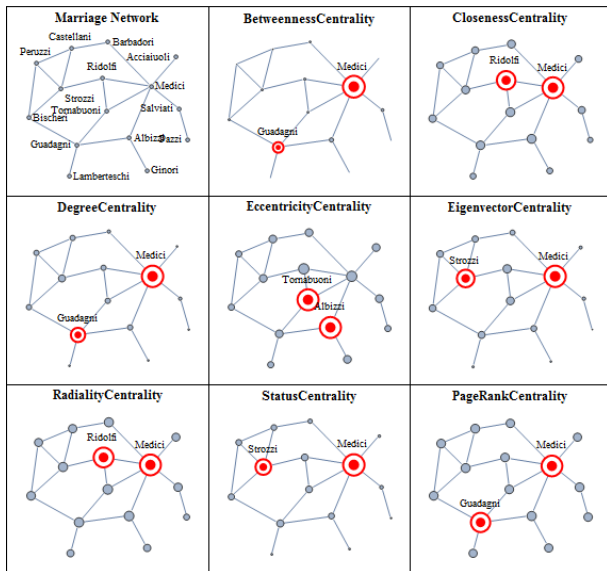
$n_c$  - number of concordant pairs,  $n_d$  - number of discordant pairs

- $-1 \leq \tau \leq 1$ , perfect agreement  $\tau = 1$ , reversed  $\tau = -1$
- Example

Rank 1	A	B	C	D	E
Rank 2	C	D	A	B	E

$$\tau = \frac{6 - 4}{5(5-1)/2} = 0.2$$

# Florentines families



- Centrality in Social Networks. Conceptual Clarification, Linton C. Freeman, Social Networks, 1, 215-239, 1979
- Power and Centrality: A Family of Measures, Phillip Bonacich, The American Journal of Sociology, Vol. 92, No. 5, 1170-1182, 1987
- A new status index derived from sociometric analysis, L. Katz, Psychometrika, 19, 39-43, 1953.
- Eigenvector-like measures of centrality for asymmetric relations, Phillip Bonacich, Paulette Lloyd, Social Networks 23, 191-201, 2001