Network Centrality

Ya Qian
Wolfgang Karl Härdle

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. – Center for Applied Statistics
and Economics
Humboldt-Universität zu Berlin
lvb.wiwi.hu-berlin.de
case.hu-berlin.de
irtg1792.hu-berlin.de

International Trade Network

Figure: International trade picture

From Florida State Hispanic Chamber of Commerce

Production Network

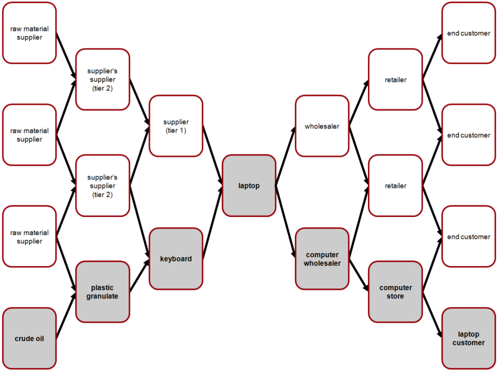


Figure: Supply chain

From Wikipedia

Social Network

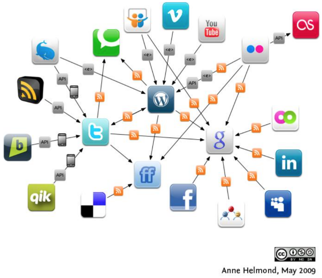


Figure: A sample of social network

From Webpage Kikolani

International Financial Network

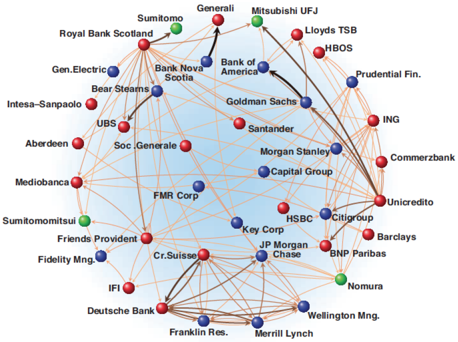


Figure: A sample of international financial network

From Frank Schweitzer, et al. (Science 325, 422 (2009))

Network of Coauthorship

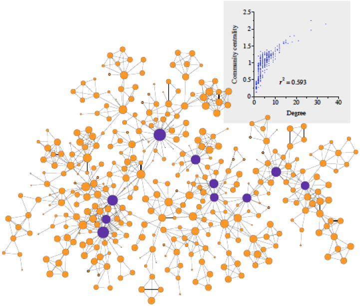


Figure: A network of coauthorships between 379 scientists whose research centers on the properties of networks of one kind or another

From M.E.J. Newman (Phys. Rev. E 74, 036104 (2006))

A More Specific Coauthorship Graph

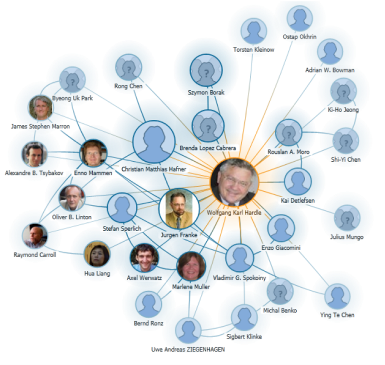


Figure: Coauthorship graph of W.K. Härdle

Source: Microsoft Academic

A More Specific Coauthorship Graph

Figure: Coauthorship path of W.K. Härdle

Source: Microsoft Academic

A More Specific Coauthorship Graph

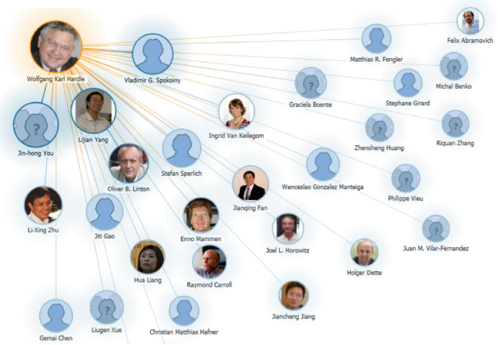


Figure: Citation path of W.K. Härdle

Source: Microsoft Academic

Outline

1. Motivation ✓
2. Definition
3. Relation to Adjacency Matrix
4. Node Centrality
5. Comparison of Various Node Centralities
6. Extension:Directed Cases
7. Graph Centrality
8. Examples of Various Graph Centralities

Notation & Terminology

An **undirected graph** $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of a list of vertices $\mathcal{V} = \{1, 2, \dots, N\}$ and a set of edges $\mathcal{E} = \{(i,j),(k,l),\dots\}$ for $i, j, k, l \in \mathcal{V}$ where

- vertices $\mathcal{V}$: node, individual, agent,
- edges $\mathcal{E}$: links, connections, ties.

Notation & Terminology

A graph could be represented by its **adjacency matrix** $A = [a_{i,j}]$ where

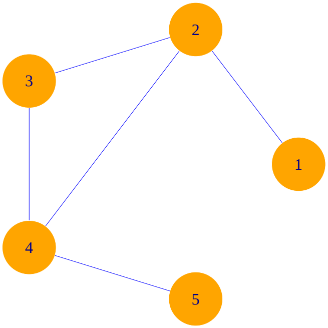
$$a_{i,j} = \begin{cases} 1 & \text{if } (i,j) \in \mathcal{V} \\ 0 & \textit{otherwise} \end{cases}$$

- A: symmetric binary matrix with diagonal to be zero.

Some Terms and Concepts of Graph Theory

A graph with five nodes and five edges:

A sample symmetric, undirected network of 5 nodes



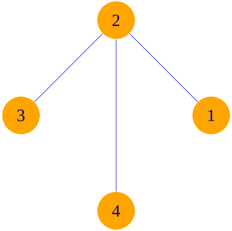
- Degree: the number of other nodes to which a given node is adjacent,
- Path: unordered pair of nodes, (P_i, P_j) each is reachable,
- Distance: the number of edges in that path,
- Geodesics: the shortest paths linking a given pair of nodes.

Adjacency Matrix & Network

Example 1: symmetric, unweighted adjacency matrix with 4 nodes

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

A sample symmetric, undirected network of 4 nodes

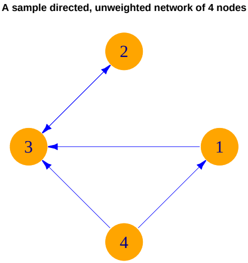


METISNET-adjtonet

Adjacency Matrix & Network

Example 2: asymmetric, unweighted adjacency matrix with 4 nodes

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$



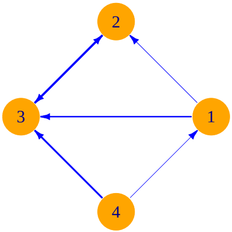
METISNET-adjtonet

Adjacency Matrix & Network

Example 3: asymmetric, weighted adjacency matrix with 4 nodes

$$A = \begin{bmatrix} 0 & 0.3 & 0.7 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0.1 & 0 & 0.9 & 0 \end{bmatrix}$$

A sample directed, weighted network of 4 nodes



METISNET-adjtonet

Three Distinct Structural Properties

- The position has the maximum possible degree,
- It falls on the geodesics between the largest possible number of other nodes,
- It is maximally close to them.

Node Centrality

- Degree centrality
- Closeness centrality
- Betweenness centrality
- Eigenvector centrality
- Katz centrality
- PageRank centrality
- Percolation centrality
- Cross-clique centrality
- Freeman Centrality

Degree Centrality

Nieminen (1974) has introduced a simple, natural and perfectly general measure of centrality based upon degree. For a network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, degree centrality of a node v_k equals

$$C_D(v_k) = \sum_{i=1}^{n} a(v_i, v_k)$$

where

$$a(v_i, v_k) = \begin{cases} 1 & \text{if } v_i \text{ and } v_k \text{ are connected by a line,} \\ 0 & \text{otherwise.} \end{cases}$$

The magnitude of $C_D(v_k)$ is partly a function of the size of the network on which it is calculated. To compare the relative centrality of nodes from different graphs, for example, we need a measure from which the effect of network size has been removed.

A given node, v_k can at most be adjacent to $n-1$ other nodes in a graph. The maximum of $C_D(v_k)$, therefore, is $n-1$. Then

$$C'_D(v_k) = \frac{\sum_{i=1}^{n} a(v_i, v_k)}{n-1}$$

Assumptions

- a measure of immediate effects(t+1) only
- models the frequency of visits by something taking an infinitely long random walk through a network

Applicable processes

- parallel duplication flow processes
- walk-based transfer processes such as the money exchange process

Closeness Centrality

Closeness-based measures of node centrality have been developed by Bavelas (1950), Beauchamp (1965), Sabidussi (1966), Moxley (1974) and Rogers (1974). The simplest and most natural of these measures is Sabidussi's (1966). For a network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, closeness centrality equals

$$C_C(v_k) = \frac{1}{\sum_{i=1}^{n} d(v_i, v_k)} \tag{1}$$

- $d(v_i, v_k)$: the number of edges in the geodesic linking v_i and v_k

Beauchamp (1965) derived the relative centrality based on closeness measure.

$$C'_C(v_k) = \frac{n-1}{\sum_{i=1}^{n} d(v_i, v_k)} \tag{2}$$

Assumptions

- flow along shortest paths or flow by parallel duplication
- shortest path assumptions: only works on connected graphs; taking shortest paths implies taking valid paths

Applicable processes

- geodesic paths
- parallel duplication flow processes

Betweenness Centrality

Shaw (1954) included betweenness counts in a complex empirically based measure of centrality, but he did not develop a measure of betweenness. Direct measures were developed independently by Anthonisse (1971) and Freeman (1977). For a network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, betweenness centrality equals

$$C_B(v_k) = \sum_{i<j}^{n} \sum_{j}^{n} \frac{g_{ij}(v_k)}{g_{ij}}$$

- g_{ij}: the number of geodesics linking v_i and v_j
- $g_{ij}(v_k)$: the number of geodesics linking v_i and v_j that contain v_k

Freeman (1977) proved that the maximum value taken by $C_B(v_k)$ is achieved only by the central node in a star. It is $\frac{n^2-3n+2}{2}$.

Therefore, the relative centrality of any node in a graph may be expressed as a ratio,

$$C'_B(v_k) = \frac{2C_B(v_k)}{n^2 - 3n + 2}$$

Assumptions

- the traffic is indivisible(transfer)
- the traffic travels only along shortest paths

Applicable processes

- package delivery process

More about 'distance'

- related to C_C and C_B
- the number of edges in the shortest connecting path (in the sense of edge number) – unweighted
- the real length of shortest connecting path (in the sense of real length)– weighted
- considered as a 'cost'

Example is given at the end of this talk : Minnesota Road Networks

Eigenvector Centrality

For a network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, eigenvector centrality equals

$$C_E(v) = \frac{1}{\lambda} \sum_{t \in M(v)} C_E(t) = \frac{1}{\lambda} \sum_{t \in \mathcal{G}} a_{v,t} C_E(t)$$

$$\lambda C_E = A C_E$$

- $A = \{a_{v,t}\}_{v,t=1}^{N}$ is a 0-1 adjacency matrix
- λ: maximum eigenvalue of A
- $M(v)$: set of neighbors of v
- $a_{v,t}$: the vt_{th} element of A

Assumptions

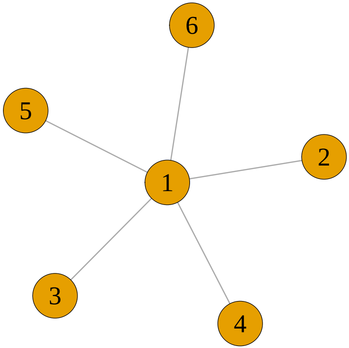
- the traffic is able to move via unrestricted walks
- each node affects all of its neighbors simultaneously

Applicable processes

- influence type processes

which is the central node?

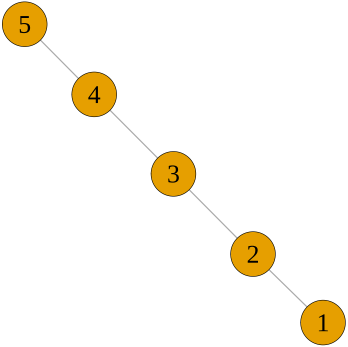
Example 4:



- $\mathbf{C_D(1) = 5}$, $C_D(2) = C_D(3) = C_D(4) = C_D(5) = C_D(6) = 1$
- $\mathbf{C_C(1) = 1}$, $C_C(2) = C_C(3) = C_C(4) = C_C(5) = C_C(6) = 0.56$
- $\mathbf{C_B(1) = 1}$, $C_B(2) = C_B(3) = C_B(4) = C_B(5) = C_B(6) = 0$
- $\mathbf{C_E(1) = 0.71}$, $C_E(2) = C_E(3) = C_E(4) = C_E(5) = C_E(6) = 0.32$

which is the central node?

Example 5:

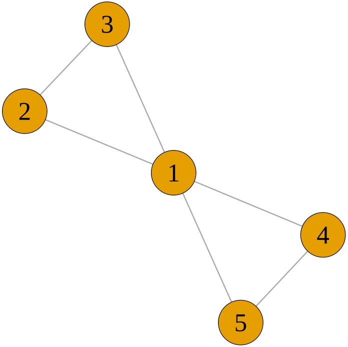


- $C_D(1) = C_D(5) = 1$, $\mathbf{C_D(2) = C_D(3) = C_D(4) = 2}$
- $C_C(1) = C_C(5) = 0.4$, $C_C(2) = C_C(4) = 0.57$, $\mathbf{C_C(3) = 0.67}$
- $C_B(1) = C_B(5) = 0$, $C_B(2) = C_B(4) = 0.50$, $\mathbf{C_B(3) = 0.67}$
- $C_E(1) = C_E(5) = 0.29$, $C_E(2) = C_E(4) = 0.50$, $\mathbf{C_B(3) = 0.58}$

METISNET-centralitymeasures

which is the central node?

Example 6:

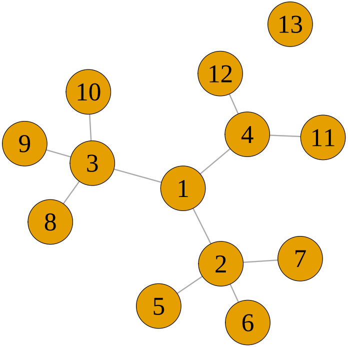


- $\mathbf{C_D(1) = 4}$, $C_D(2) = C_D(3) = C_D(4) = C_D(5) = 2$
- $\mathbf{C_C(1) = 1}$, $C_C(2) = C_C(3) = C_C(4) = C_C(5) = 0.67$
- $\mathbf{C_B(1) = 0.67}$, $C_B(2) = C_B(3) = C_B(4) = C_B(5) = 0$
- $\mathbf{C_E(1) = 0.62}$, $C_E(2) = C_E(3) = C_E(4) = C_E(5) = 0.39$

METISNET-centralitymeasures

which is the central node?

Example 7:



- $C_D(1) = C_D(4) = 3$, $\mathbf{C_D(2) = C_D(3) = 4}$
- $\mathbf{C_C(1) = 0.63}$, $C_C(2) = C_C(3) = 0.52$, $C_C(4) = 0.48$
- $\mathbf{C_B(1) = 0.61}$, $C_B(2) = C_B(3) = 0.36$, $C_B(4) = 0.27$
- $\mathbf{C_E(1) = 0.51}$, $C_E(2) = C_E(3) = 0.44$, $C_E(4) = 0.33$

METISNET-centralitymeasures

Minnesota Road Networks

- Minnesota Road Net data minnesota.mat is avaiable in Matlab R2016b and later versions
- Minnesota State of United States
- The data contains a G object which is a combination of Nodes (Coordinates of locations) and Edges (roads between each pair of nodes)

Minnesota Road Networks

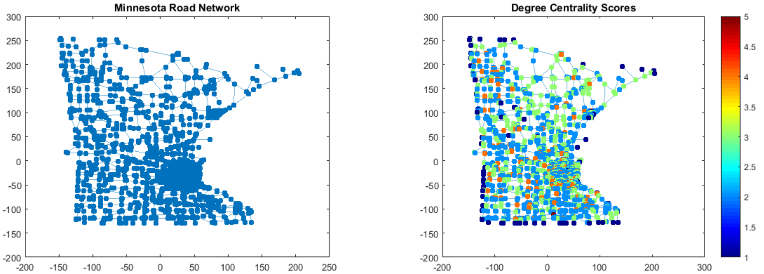


Figure: Display of Minnesota Road Net – no centrality displayed (left), degree centrality C_D (right)

METISNET-centralitycomparison

Minnesota Road Networks

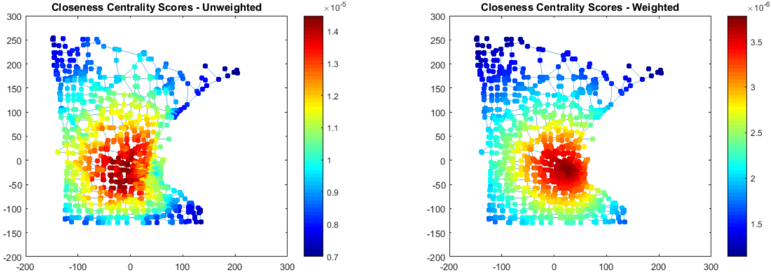


Figure: Display of Minnesota Road Net – closeness centrality unweighted (left), closeness centrality weighted (right)

METISNET-centralitycomparison

Minnesota Road Networks

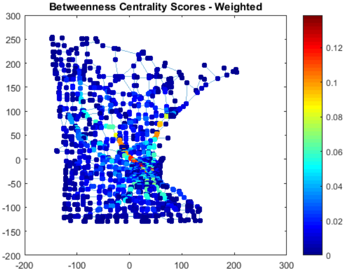


Figure: Display of Minnesota Road Net – betweenness centrality unweighted (left), betweenness centrality weighted (right)

METISNET-centralitycomparison

Minnesota Road Networks

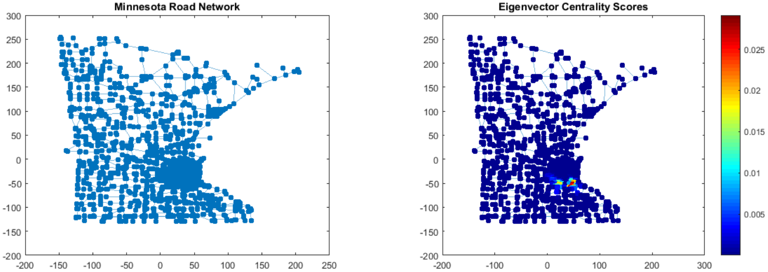


Figure: Display of Minnesota Road Net – no centrality displayed (left), eigenvector centrality C_E(right)

METISNET-centralitycomparison

Notation & Terminology

A **directed graph** $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of a list of vertices $\mathcal{V} = \{1, 2, \ldots, N\}$ and a set of **ordered edges** $\mathcal{E} = \{(i, j), (j, i), (k, l), (l, k) \ldots\}$ for $i, j, k, l \in \mathcal{V}$ where

- vertices $\mathcal{V}$: node, individual, agent,
- edges $\mathcal{E}$: arrows, directed edges, directed lines.

Notation & Terminology

A **directed graph** could be represented by its **adjacency matrix** $A = [a_{i,j}]$ where

$$a_{i,j} = \begin{cases} 1 & \text{if there is an arrow from i to j} \\ 0 & \textit{otherwise} \end{cases}$$

- there is no need to be an arrow from j to i simutaneouly.

Degree Centrality: Directed Cases

The **in-degree centrality** of node v_k is given by:

$$C_{InD}(v_k) = \sum_{i=1}^{n} a_{(v_i, v_k)}$$

- $a(v_i, v_k)$ equals 1 , if there is an arrow from v_i to v_k and 0 otherwise.

The **out-degree centrality** of node v_k is given by:

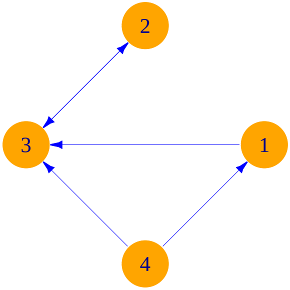
$$C_{OutD}(v_k) = \sum_{i=1}^{n} a_{(v_k, v_i)}$$

- $a(v_k, v_i)$ equals 1 , if there is an arrow from v_k to v_i and 0 otherwise.

Example:

Degree Centrality: Directed Cases

A sample directed, unweighted network of 4 nodes



- $C_{InD}(1) = C_{InD}(2) = 1$, $C_{InD}(3) = 3$, $C_{InD}(4) = 0$
- $C_{OutD}(1) = C_{OutD}(2) = 1$, $C_{OutD}(3) = 0$, $C_{OutD}(4) = 2$

Katz Centrality

The Katz centrality of node v_k is given by:

$$C_K(v_k) = \alpha \sum_{i=1}^{n} a_{(v_i, v_k)} \, C_K(v_i) + \beta$$

- α : scaling vector to normalize the score
- β : weight of the centrality of vertices ego is tied to .
  - if $\beta > 0$, ego has higher centrality when tied to vertices that are central
  - if $\beta < 0$, ego has higher centrality when tied to vertices that are not central
  - with $\beta = 0$, it is degree centrality

In matrix form :

$$C_K = \alpha C_K A + \beta$$

- where β is now a vector whose elements are all equal a given positive constant.

It follows that C_K can be computed as:

$$C_K = \beta(I - \alpha A)^{-1}$$

- The Katz status of a node is defined as the number of weighted paths reaching the node in the network plus an exogenous factor, a generalization of the in-degree measure which counts only paths of length one.

PageRank Centrality

The PageRank centrality of node v_k is given by:

$$C_{PR}(v_k) = \alpha \sum_{i=1}^{n} \frac{a_{k,i}}{d(v_i)} C_{PR}(v_i) + \beta$$

- α and β are constants
- $d(v_i)$ is the out-degree of node v_i if such degree is positive, or $d(v_i) = 1$ if the out-degree of v_i is null.

In matrix form :

$$C_{PR} = \alpha C_{PR} D^{-1} A + \beta$$

- β is now a vector whose elements are all equal a given positive constant
- D^{-1} is a diagonal matrix with i-th diagonal element equal to $1/d(v_i)$.

It follows that C_{PR} can be computed as:

$$C_{PR} = \beta(I - \alpha D^{-1}A)^{-1}$$

- The damping factor α and the personalization vector β have the same role seen for Katz centrality.

Two Certain Features of Graph Centralization

- It should index the degree to which the centrality of the most central node exceeds the centrality of all other nodes
- It should each be expressed as a ratio of that excess to its maximum possible value for a graph containing the observed number of nodes.

Then Freeman(1977) proposed a measure in graph centrality.

$$C_X = \frac{\sum_{i=1}^{n}\{C_X(v^*) - C_X(v_i)\}}{\max \sum_{i=1}^{n}\{C_X(v^*) - C_X(v_i)\}} \tag{3}$$

- n: number of nodes
- $C_X(v_i)$: one of the nodes centralities defined above
- $C_X(v^*)$: largest value of $Cx(v_i)$ for any node in the network
- $\max \sum_{i=1}^{n}\{C_X(v^*) - C_X(v_i)\}$: the maximum possible sum of differences in node centrality for a graph of n nodes

Degree-based measures of graph centrality

$$\begin{aligned} C_D &= \frac{\sum_{i=1}^{n}\{C_D(v^*) - C_D(v_i)\}}{\max\sum_{i=1}^{n}\{C_D(v^*) - C_D(v_i)\}} \\ &= \frac{\sum_{i=1}^{n}\{C_D(v^*) - C_D(v_i)\}}{n^2 - 3n + 2} \end{aligned}$$

Betweeness-based measures of graph centrality

$$
\begin{aligned}
C_B &= \frac{2\sum_{i=1}^{n}\{C'_B(v^*) - C'_B(v_i)\}}{n-1} \\
&= \frac{\sum_{i=1}^{n}\{C_B(v^*) - C_B(v_i)\}}{n^3 - 4n^2 + 5n - 2}
\end{aligned}
$$

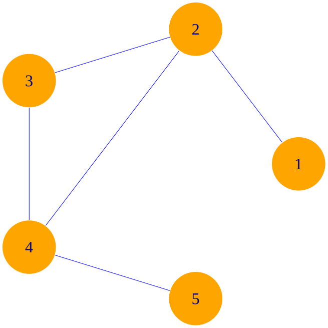
Closeness-based measures of graph centrality

$$C_C = \frac{\sum_{i=1}^{n}\{C'_C(v^*) - C'_C(v_i)\}}{(n^2 - 3n + 2)/(2n - 3)}$$

Example:

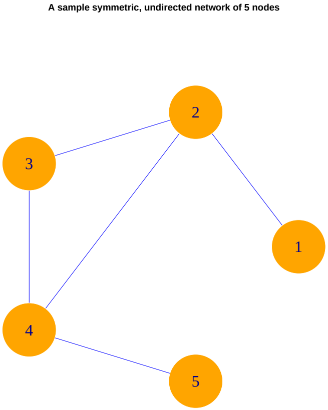
Degree Centrality

A sample symmetric, undirected network of 5 nodes



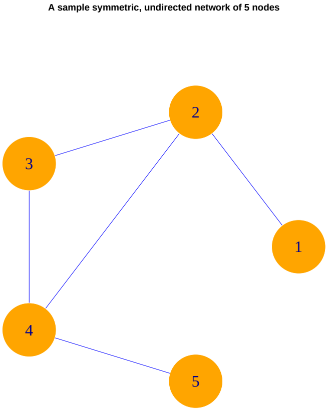
- $C_D = 0.42$
- $C_D(v_1) = 1, C_D(v_2) = 3, C_D(v_3) = 2, C_D(v_4) = 3, C_D(v_5) = 1;$
- $C'_D(v_1) = 0.25, C'_D(v_2) = 0.75, C'_D(v_3) = 0.5, C'_D(v_4) = 0.75, C'_D(v_5) = 0.25$

Betweenness Centrality



- $C_B = 0.38$
- $C_B(v_1) = 0, C_B(v_2) = 3, C_B(v_3) = 0, C_B(v_4) = 3, C_B(v_5) = 0;$
- $C'_B(v_1) = 0, C'_B(v_2) = 0.5, C'_B(v_3) = 0, C'_B(v_4) = 0.5, C'_B(v_5) = 0$

Closeness Centrality



- $C_C = 0.43$
- $C_C(v_1)^{-1} = 8, C_C(v_2)^{-1} = 5, C_C(v_3)^{-1} = 6, C_C(v_4)^{-1} = 5, C_C(v_5)^{-1} = 8;$
- $C'_C(v_1) = 0.5, C'_C(v_2) = 0.8, C'_C(v_3) = 0.67, C'_C(v_4) = 0.8, C'_C(v_5) = 0.5$

Reference

Anthonisse, J. M.
The Rush in a Graph
Mathematisch Centrum, Amsterdam, 1971.

Bavelas, A.
Communication patterns in task oriented groups
Journal of the Acoustical Society of America, 22:271-282, 1950.

Beauchamp, M. A.
An improved index of centrality
BehavioralScience, 10:161-163, 1965.

Reference

- Moxley, R. L. and N. F. Moxley
  Determining point-centrity in uncontrived social networks
  Sociometry, 37:122-130,1974.

- Nieminen, J.
  On centrality in a graph
  Scandinavian Journal of Psychology, 15:322-336, 1974.

- Rogers, D. L.
  Sociometric analysis of interorganizational relations: application of theory and measurement
  Rural Sociology, 39:487-503, 1974.

Reference

Shaw, M. E.
Group structure and the behavior of individuals in small groups
Journal Of Psychology, 38: 139-149, 1954.

Sabidussi, G.
The centrality index of a graph
Psychomatrika, 31 581-603, 1966.

Rogers, D. L.
A set of measures of centrality based on betweenness
Sociometry, 40:3541, 1977.

Network Centrality

Ya Qian
Wolfgang Karl Härdle

Ladislaus von Bortkiewicz Chair of Statistics
C.A.S.E. – Center for Applied Statistics
and Economics
Humboldt-Universität zu Berlin
lvb.wiwi.hu-berlin.de
case.hu-berlin.de
irtg1792.hu-berlin.de