Network Centrality

Ya QIAN
Wolfgang Karl Härdle
Lucas Valentin Umann

Ladislaus von Bortkiewicz Professor of Statistics
Humboldt-Universität zu Berlin
Blockchain Research Center
lvb.wiwi.hu-berlin.de
Charles University, WISE XMU, NCTU 玉山學者

Figure 1: International trade

Florida State Hispanic Chamber of Commerce (2014)

Social Network

Figure 2: A social network

[medium.com]()

International Financial Network

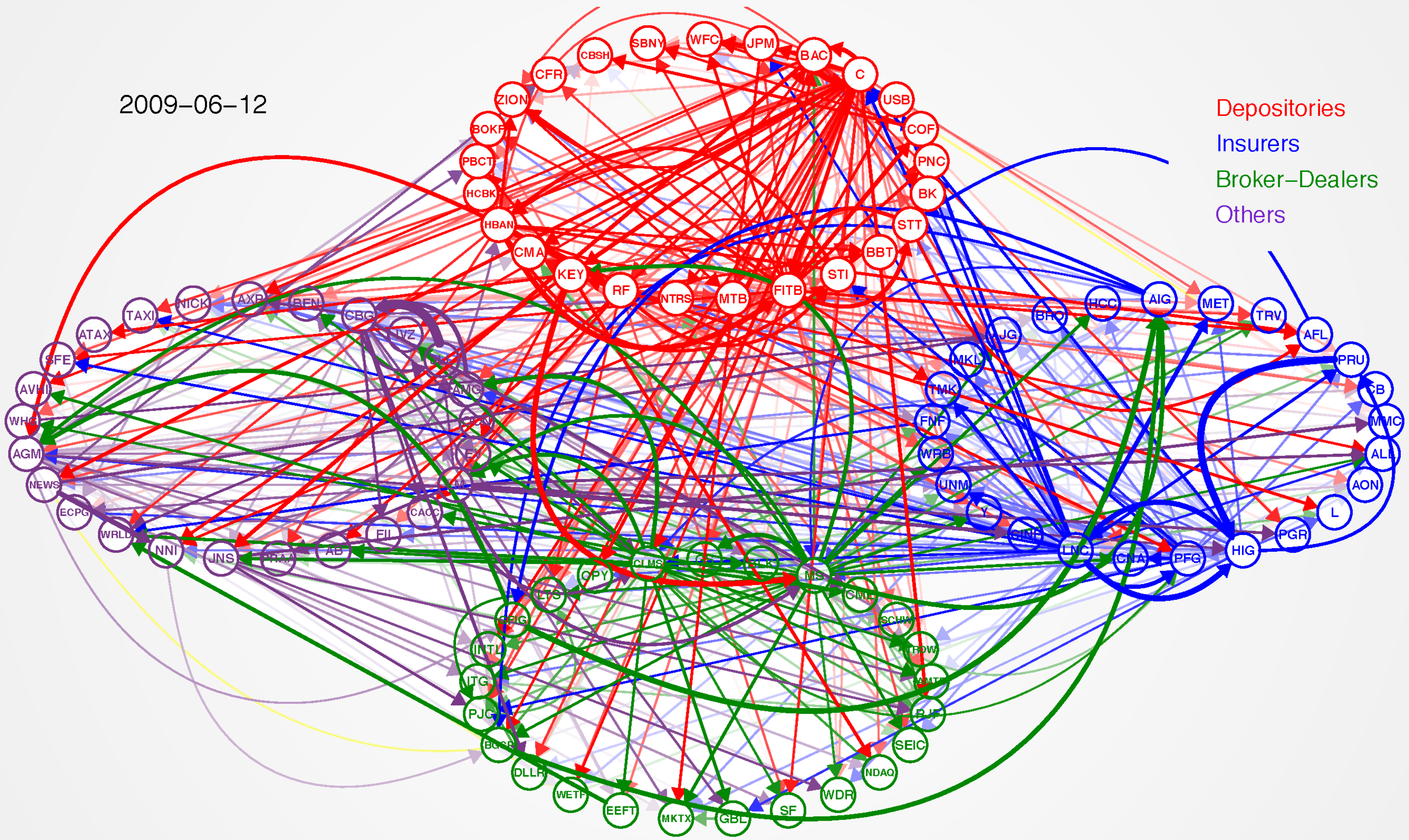


Figure 3: TENET complex international financial network

Härdle et al (2016)

Network of Coauthorship

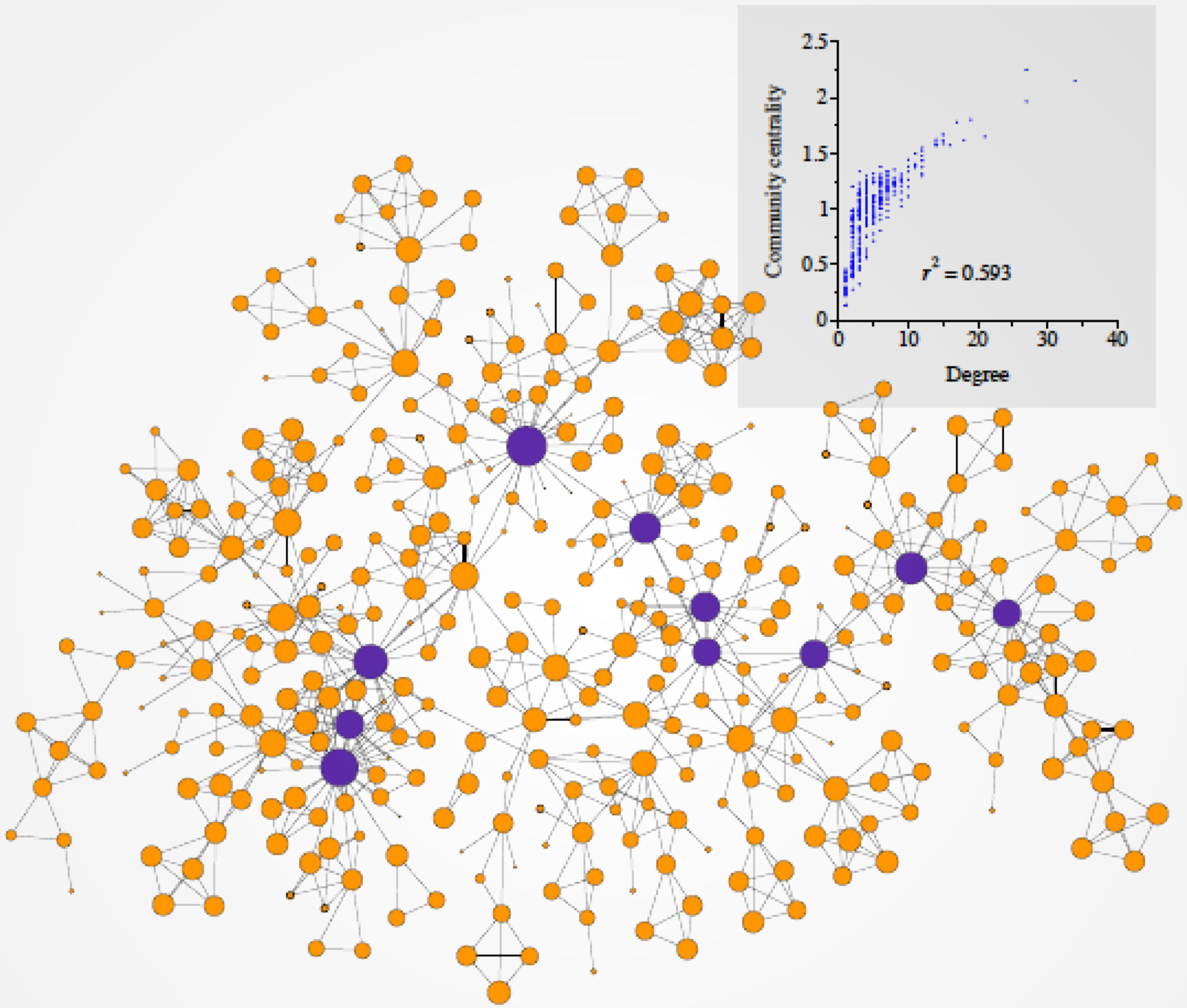


Figure 4: A network of coauthorships between 379 scientists whose research centers on the properties of networks

M.E.J. Newman (2006)

A More Specific Coauthorship Graph

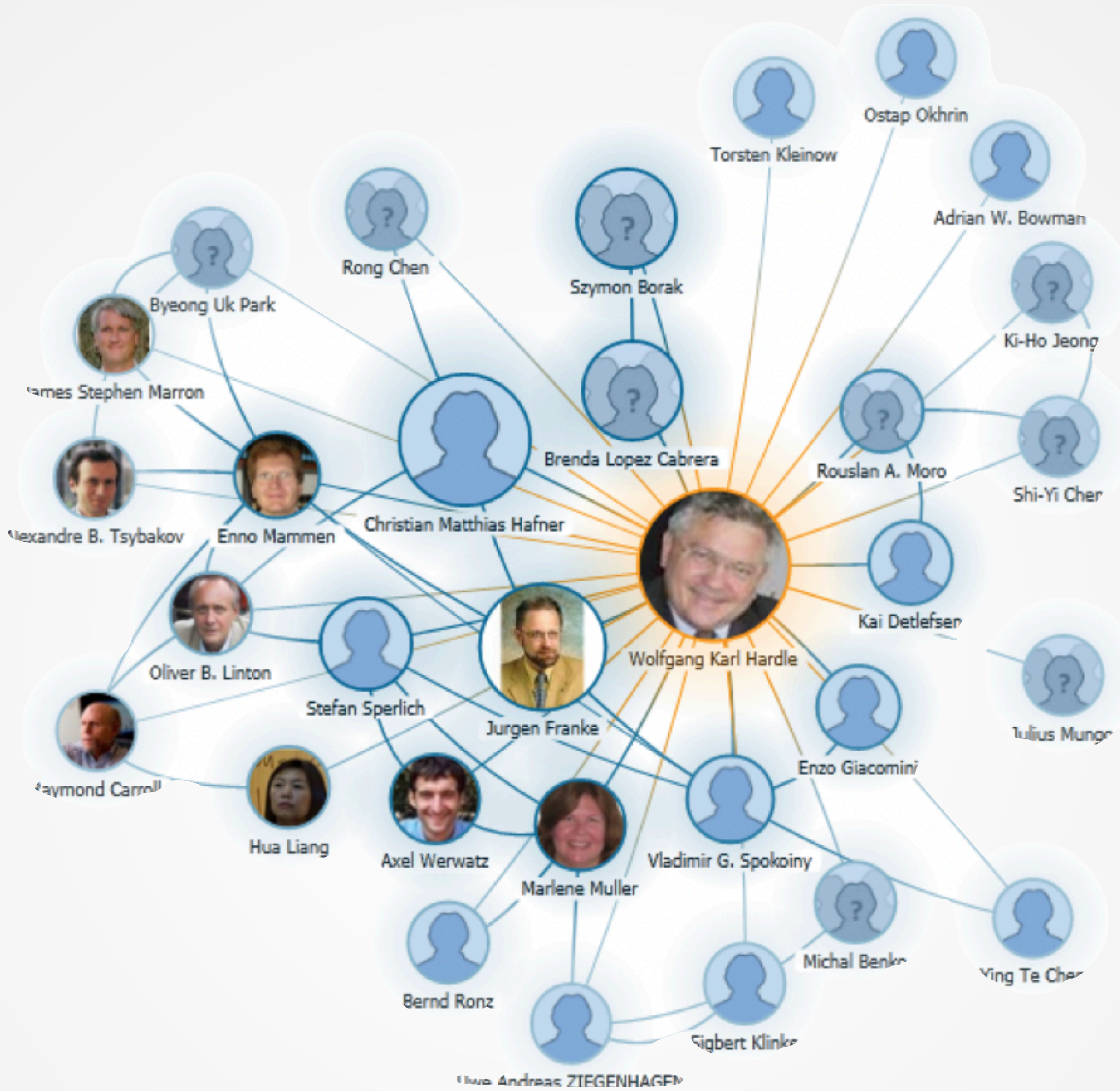


Figure 5: Coauthorship graph of W.K. Härdle

Source: Microsoft Academic

An *Even* More Specific Coauthorship Graph

Figure 6: Coauthorship graph of W.K. Härdle to Paul Erdős

Source: Microsoft Academic

The Berlin subway station network

Figure 7: An excerpt of the Berlin public transport network

Outline

1. Motivation ✓
2. Definition
3. Relation to Adjacency Matrix
4. Node Centrality
5. Comparison of Various Node Centralities
6. Directed Networks
7. Graph Centrality
8. Abstract Examples
9. Case Study: Minnesota Road Network
10. Case Study: Systemic Risk
11. Case Study: TENET
12. Case Study: CRIX Correlations
13. Case Study: CRIX & FRM

Notation & Terminology

Undirected graph $\mathscr{G} = (\mathscr{V}, \mathscr{E})$:

consists of a list of vertices $\mathscr{V} = \{v_1, v_2, \ldots, v_N\}$ and a set of edges

$\mathscr{E} = \{(v_i, v_j), (v_k, v_l), \ldots\}$ for $v_i, v_j, v_k, v_l, \ldots \in \mathscr{V}$

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

Adjacency Matrix

A graph is represented by its adjacency matrix $A = [a_{i,j}]$

where

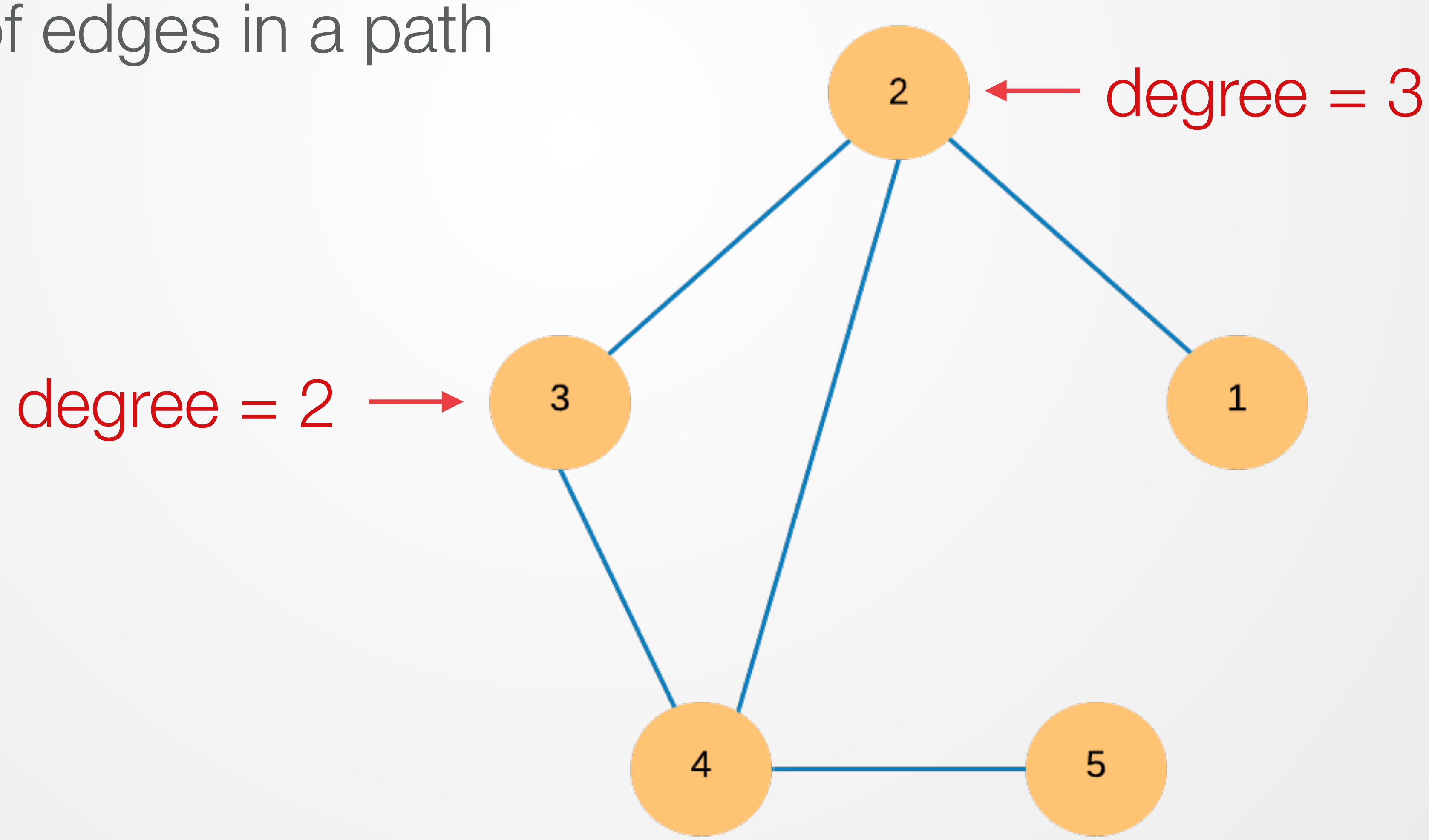
$$a_{i,j} = \mathbf{I}\{(v_i, v_j) \in \mathscr{E}\}$$

- A: symmetric binary matrix with zero diagonal
- $\mathbf{I}$: indicator function

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

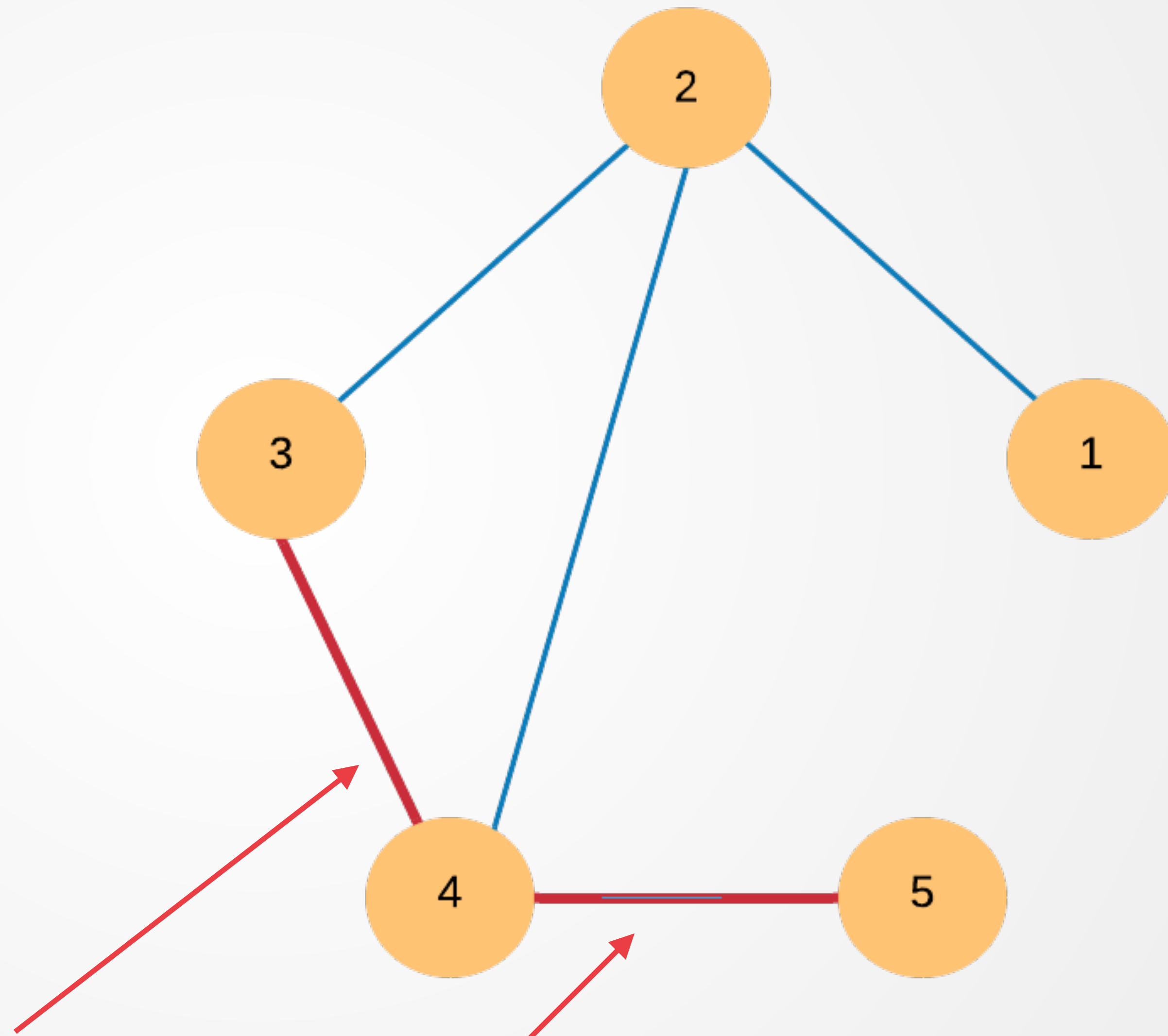
Frequently Used Terms and Concepts for Graphs

- Degree: # of other nodes adjacent to given node
- Path: sequence of edges linking two nodes (P_i, P_j)
- Distance: # of edges in a path



Frequently Used Terms and Concepts for Graphs

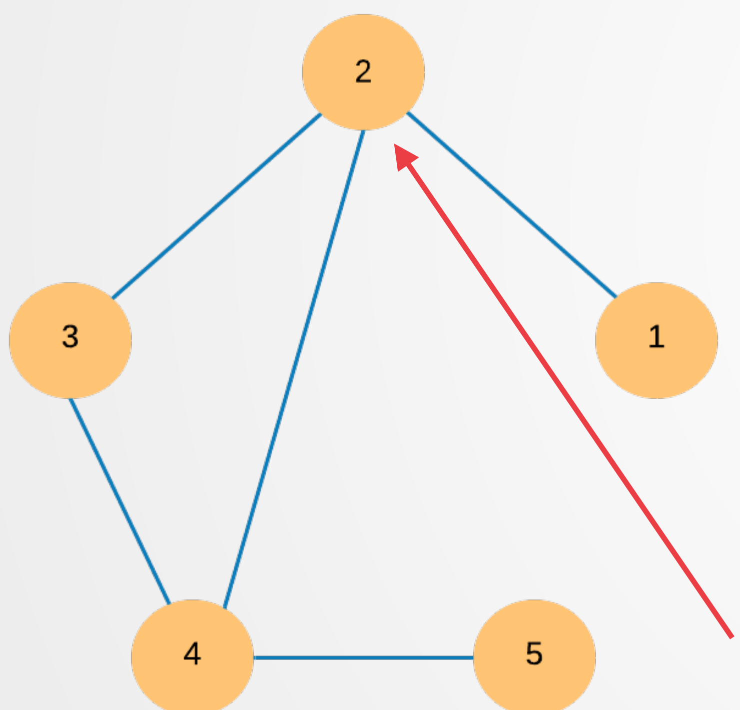
- Geodesic: shortest path linking two given nodes



Geodesic from node 3 to node 5

A Graph Illustration

Example 1: illustration of an undirected graph



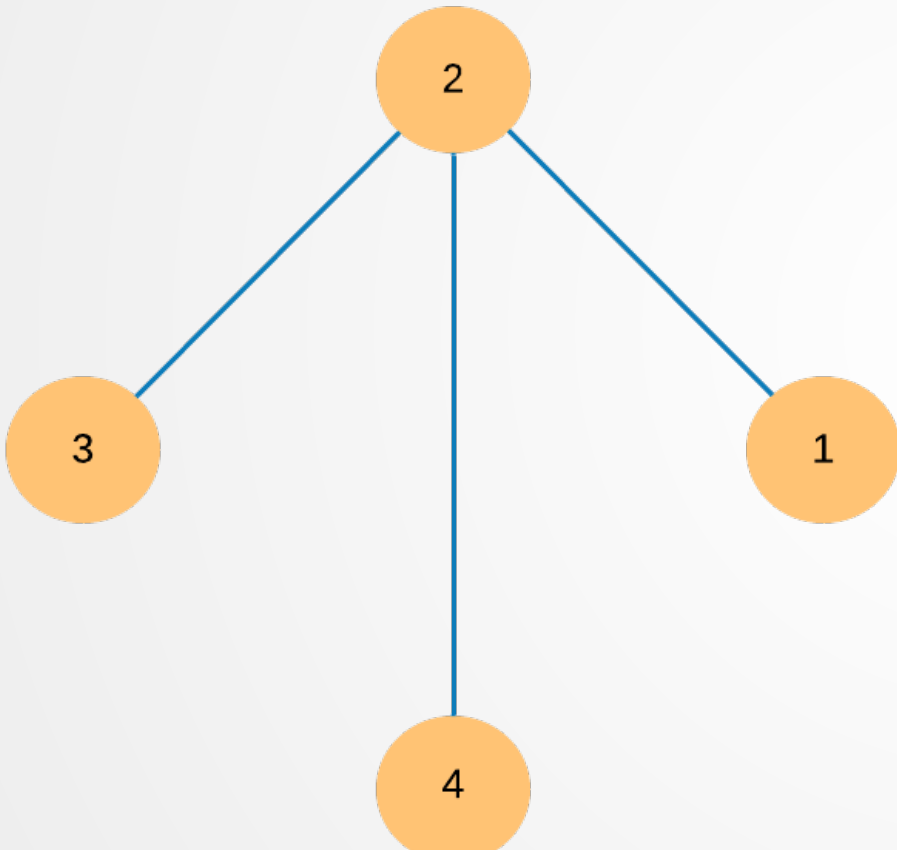
- number of nodes: 5
- number of edges: 5
- corresponding adjacency matrix:

degree = 3

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

Adjacency Matrix & Network

Example 2: symmetric, unweighted adjacency matrix with 4 nodes



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

METISNET-adjtonet

Node Centrality Measures

- Degree centrality
- Closeness centrality
- Betweenness centrality
- Eigenvector centrality
- Katz centrality
- PageRank centrality
- Freeman Centrality

Degree Centrality

Nieminen (1974): Measure of centrality based upon degree.

For a network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, degree centrality of node v_k equals

$$C_D(v_k) = \sum_{i=1}^{n} a_{i,k}$$

where

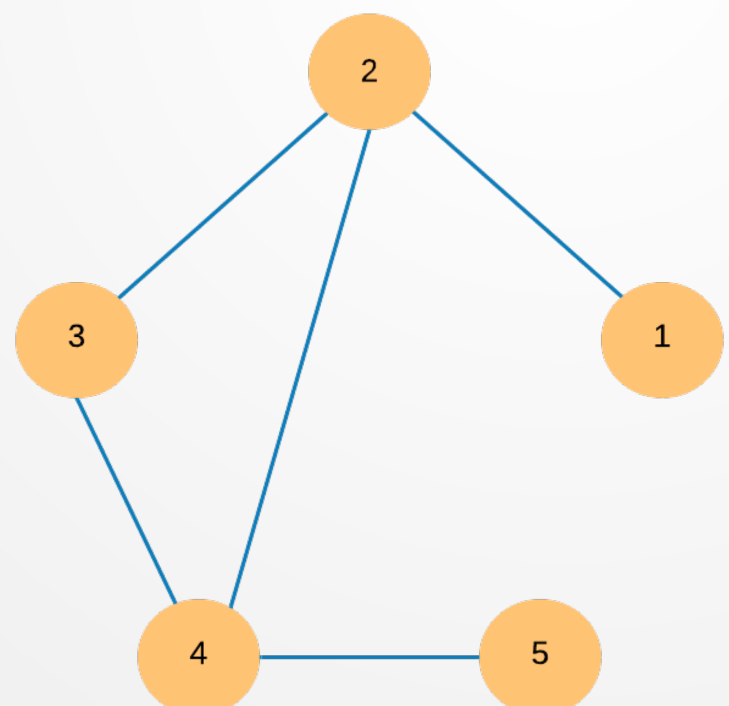
$$a_{i,j} = \mathbf{I}\{(i,j) \in \mathcal{E}\}$$

Magnitude of $C_D(v_k)$ depends on size of the network

Compare relative centrality of nodes from different graphs, remove network size. A given node v_k can, at most, be adjacent to *n–1* other nodes. The maximum value of $C_D(v_k)$, therefore, is *n–1*:

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

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



$$\frac{\sum_{i=1}^{5} a_{i,2}}{5-1} = 3/4$$

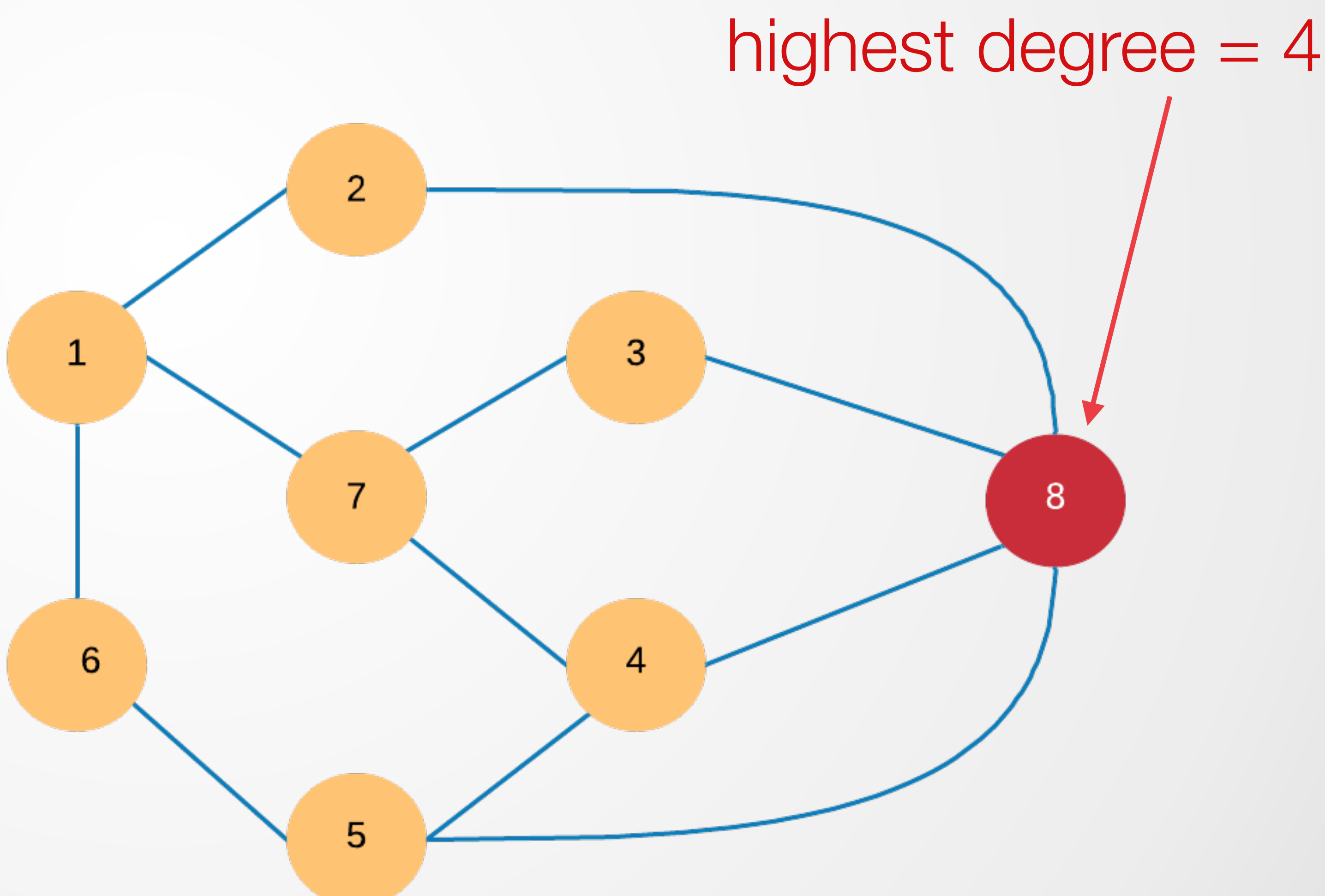
Pros & Cons

Pros

- intuitive
- easy to calculate

Cons

- node with highest degree may be peripheral, not in the center



Assumptions and Applications

Assumptions

- a measure of immediate effects (from time t to $t+1$) 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 e.g. a money exchange process

Closeness Centrality

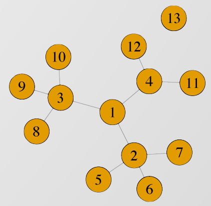
For a network $\mathscr{G} = (\mathscr{V}, \mathscr{E})$, closeness centrality equals

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

where $d(v_i, v_k)$: # of edges in the geodesic linking v_i, v_j

Relative centrality based on closeness measure:

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



Pros & Cons

Pros

- intuitive
- suitable to characterize a process in which information travels through the shortest distances

Cons

- the range of variation is too narrow due to the small diameter of networks

Assumptions and Applications

Assumptions

- flow along shortest paths
- 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

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

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

where $g_{i,j}$: # of geodesics linking v_i and v_j

and $g_{i,j}(v_k)$: # 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 $\dfrac{n^2 - 3n + 2}{2}$.

Therefore, the relative centrality of any node in a graph is the ratio:

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

Pros & Cons

Pros

- suitable for transportation networks
- considers the intermediate role taken by each node, useful for information transmission

Cons

- not general due to the limitation that it only considers shortest distances
- computationally expensive for large graphs

Assumptions and Applications

Assumptions

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

Applicable processes

- package delivery

More about ‘distance’

- related to C_C and C_B
- # of edges in the shortest connecting path – unweighted
- the real length of shortest connecting path – weighted
- considered as a ’cost’

Check Minnesota Road Case Study later.

Eigenvector Centrality

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

$$C_E(v_k) = \frac{1}{\lambda} \sum_{t \in M(v_k)} C_E(t) = \frac{1}{\lambda} \sum_{t \in \mathscr{G}} a_{k,t} C_E(v_t)$$

$$\lambda C_E = A C_E$$

where $A = \{a_{v,t}\}_{v,t=1}^{N}$ is the 0-1 adjacency matrix

λ: maximum eigenvalue of A

$M(v_t)$: set of neighbors of v_t

$a_{v,t}$: (v, t) element of A

Pros & Cons

Pros

- considers relative importance of each node

Cons

- importance of less central nodes may not be well quantified

Assumptions and Applications

Assumptions

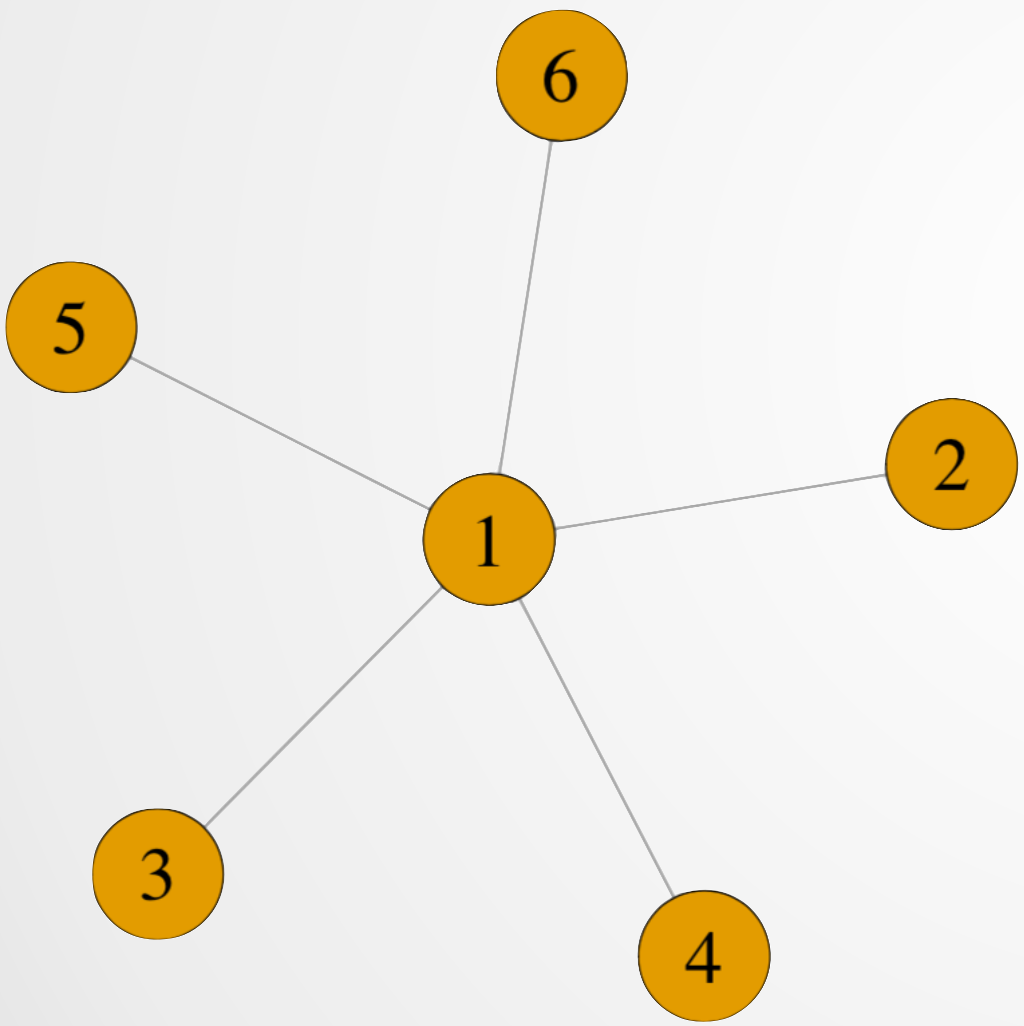
- traffic is able to move via unrestricted walks
- each node affects all of its neighbors simultaneously

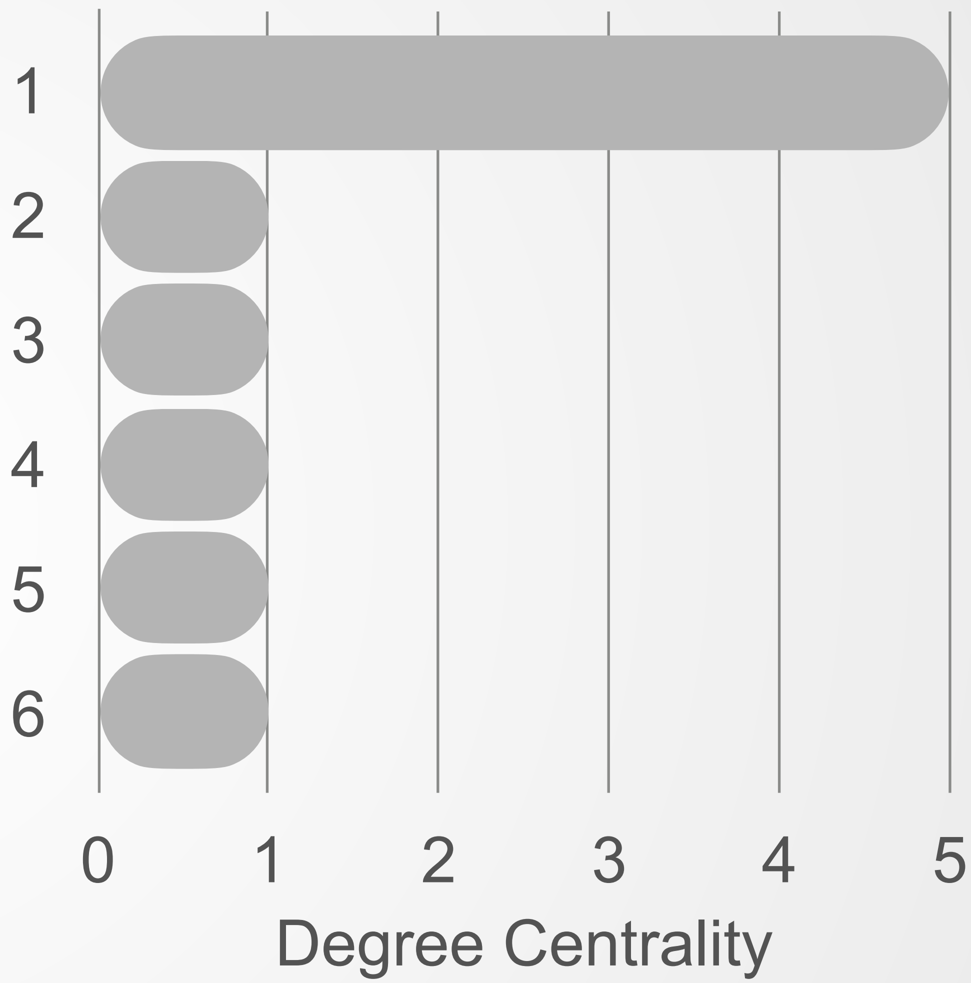
Applicable processes

- influence type processes

Which is the central node?

Example 3:

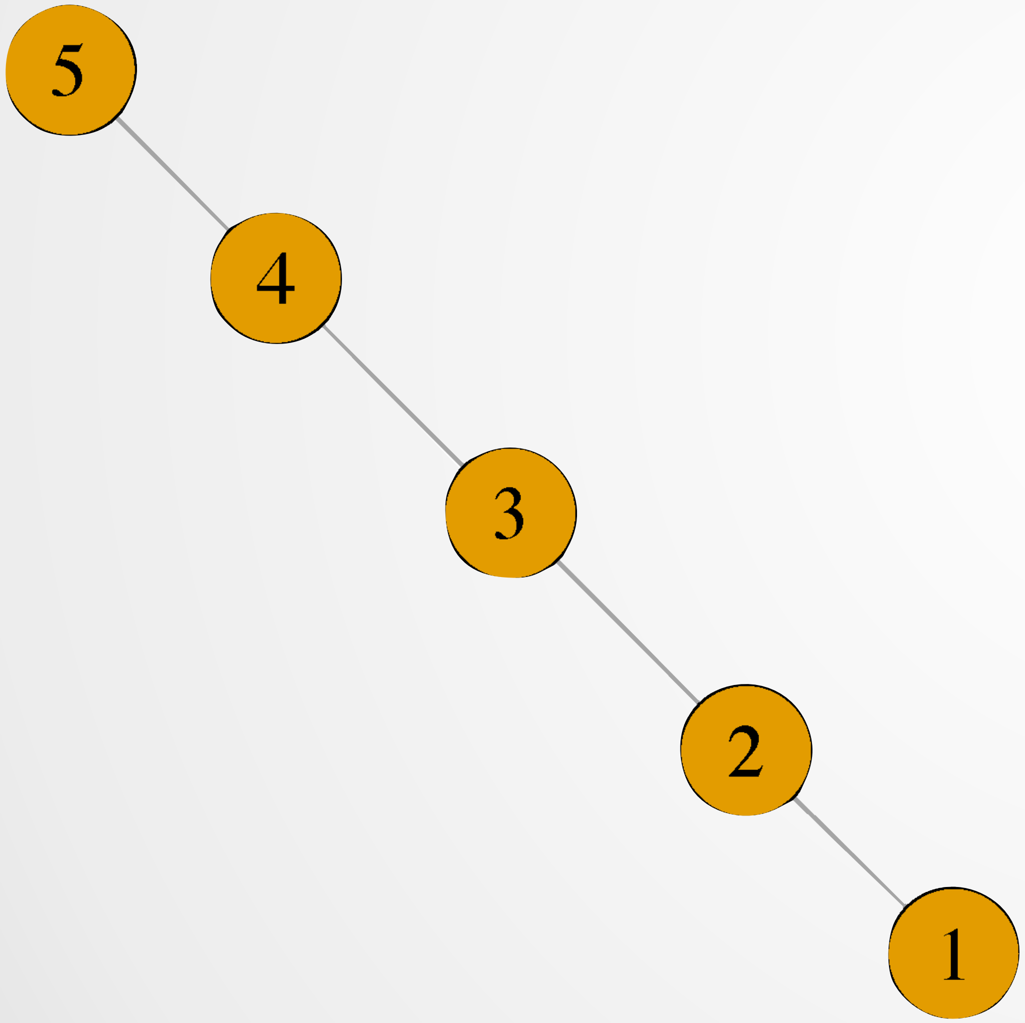


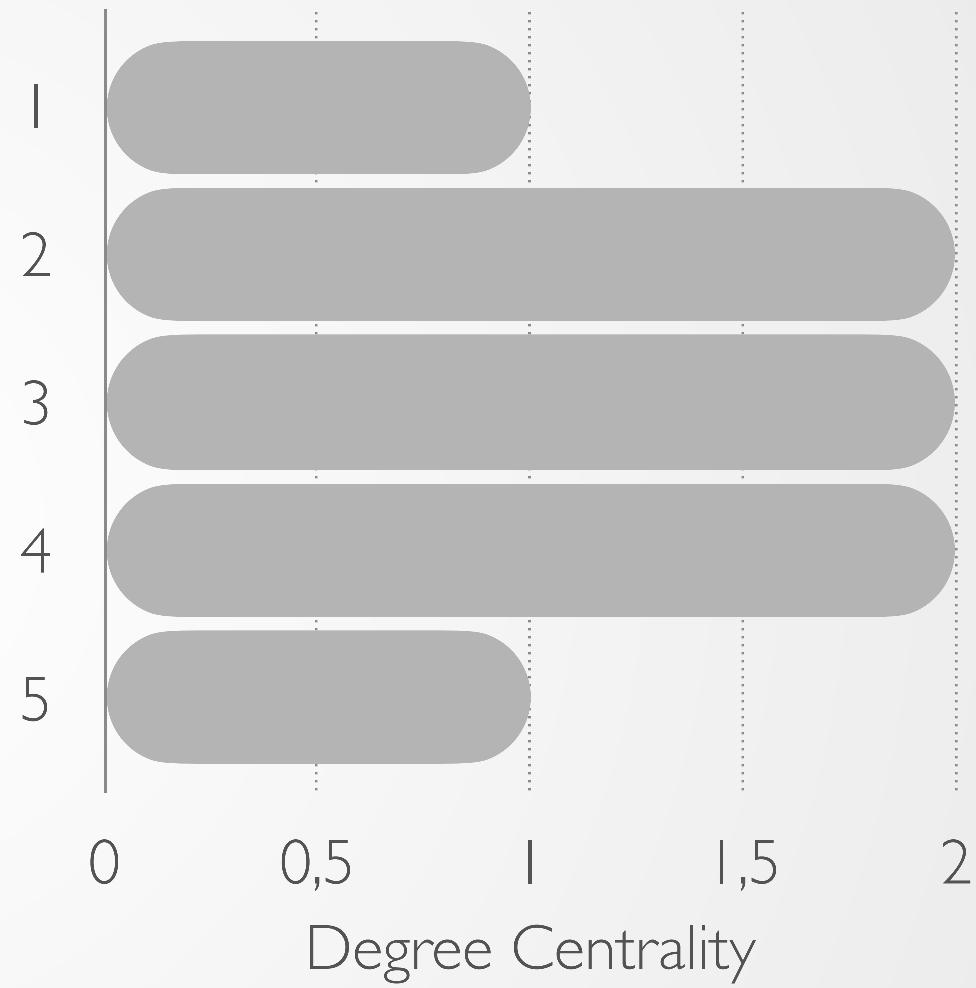


METISNET-centralitymeasures

Which is the central node?

Example 4:

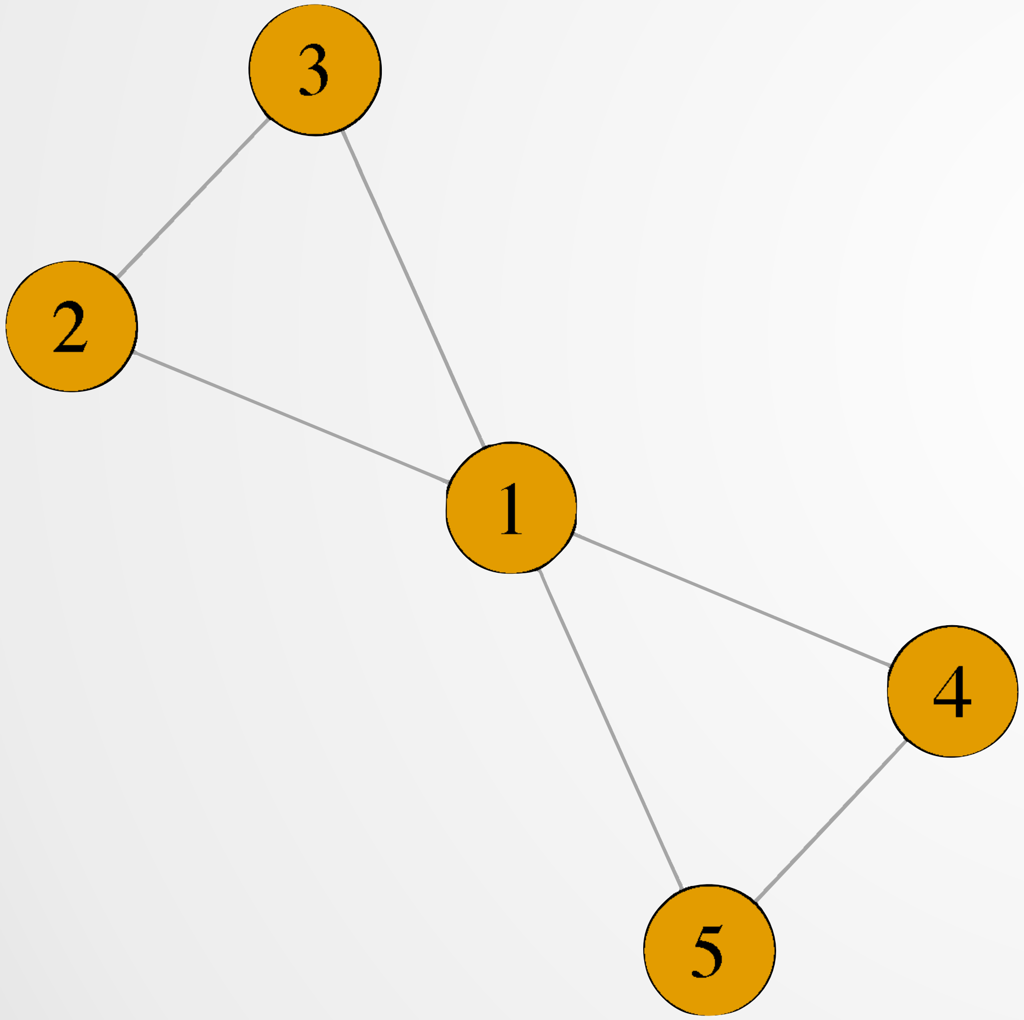


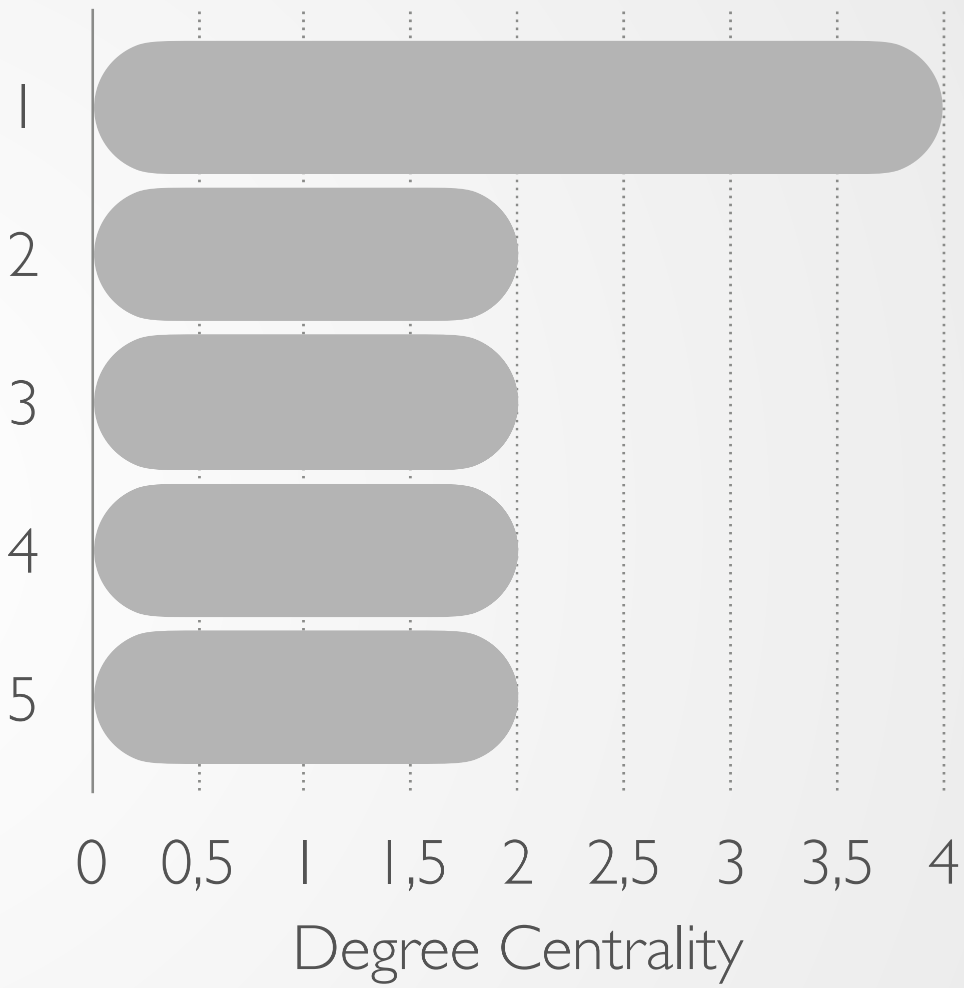


METISNET-centralitymeasures

Which is the central node?

Example 5:

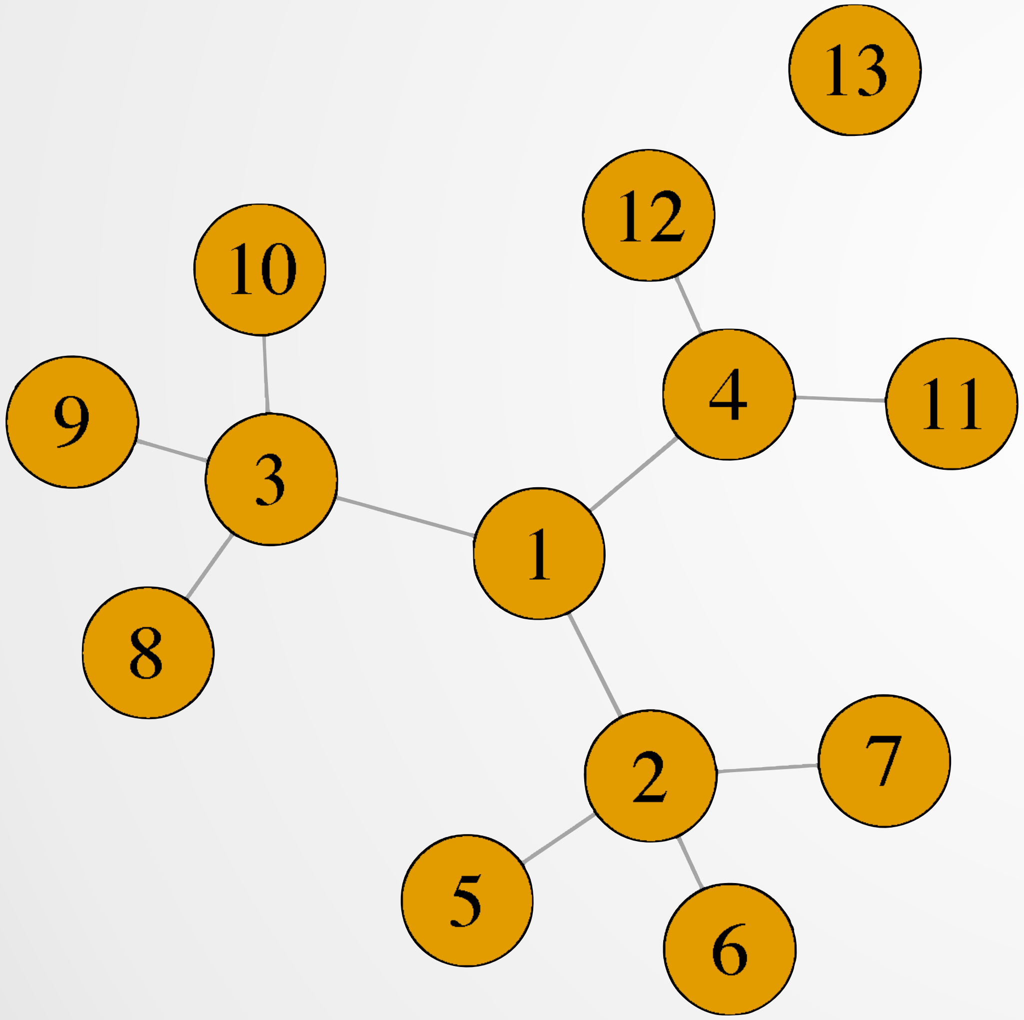


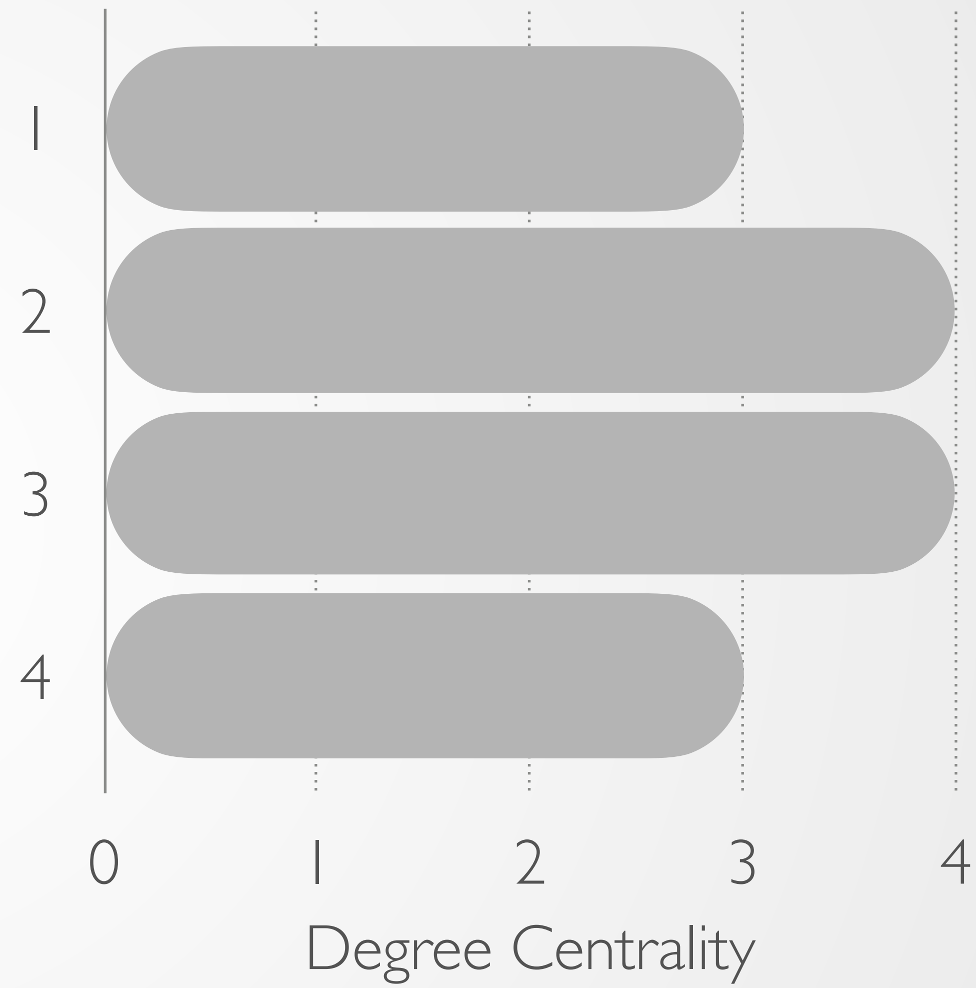


METISNET-centralitymeasures

Which is the central node?

Example 6:





METISNET-centralitymeasures

Notation & Terminology

Directed graph $\mathscr{G} = (\mathscr{V}, \mathscr{E})$:

consists of a list of vertices $\mathscr{V} = \{v_1, v_2, \ldots, v_N\}$ and a set of ordered edges $\mathscr{E} = \{(v_i, v_j), (v_j, v_i), (v_k, v_l), \ldots\}$ for $v_i, v_j, v_k, v_l \in \mathscr{V}$

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

Notation and Terminology

Important: In the undirected case we had undirected $\{v_i, v_j\}$ as edges, for directed graphs we have directed tuples (v_i, v_j)

(v_i, v_j) does not imply (v_j, v_i)!

Notation & Terminology

An unweighted directed graph is represented by its adjacency matrix $A = [a_{i,j}]$

where $a_{i,j} = \mathbf{I}\{(v_i, v_j) \in \mathscr{E}\}$

Notation & Terminology

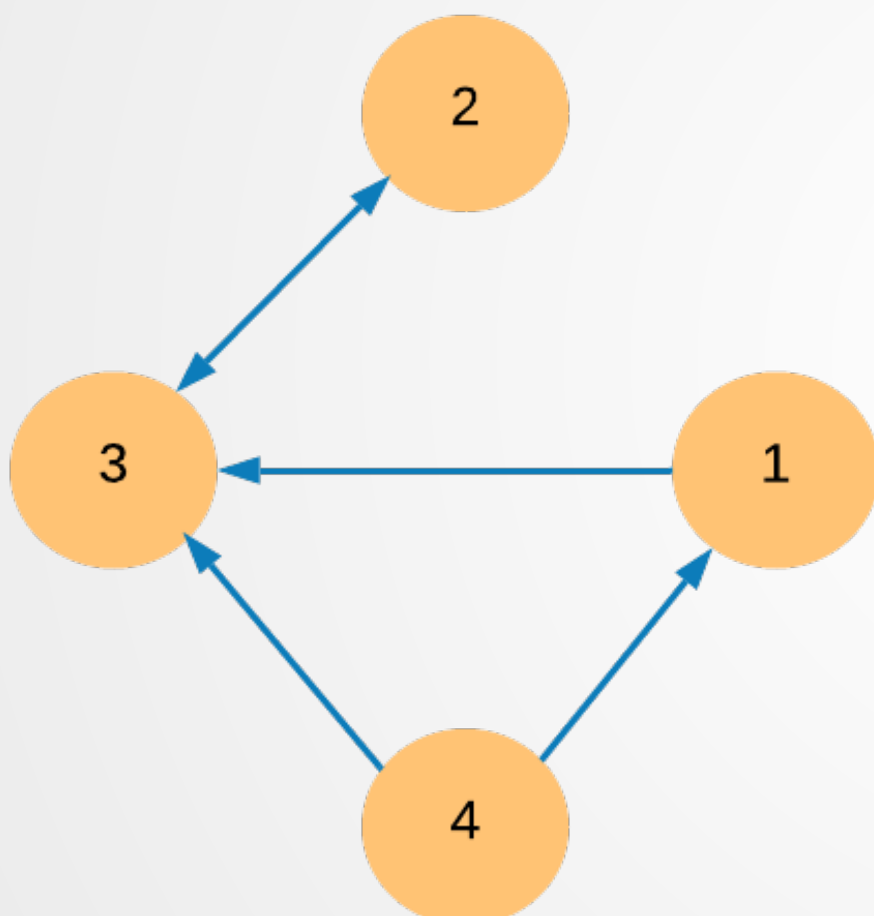
A weighted directed graph additionally incorporates weights for each edge in the adjacency matrix: $A = [a_{i,j}]$

where $a_{i,j} = w_{i,j} * \mathbf{I}\{(v_i, v_j) \in \mathscr{E}\}$

$w_{i,j}$: weighting coefficient

Asymmetric Adjacency Matrix & Directed Network

Example 7: asymmetric, unweighted adjacency matrix with 4 nodes

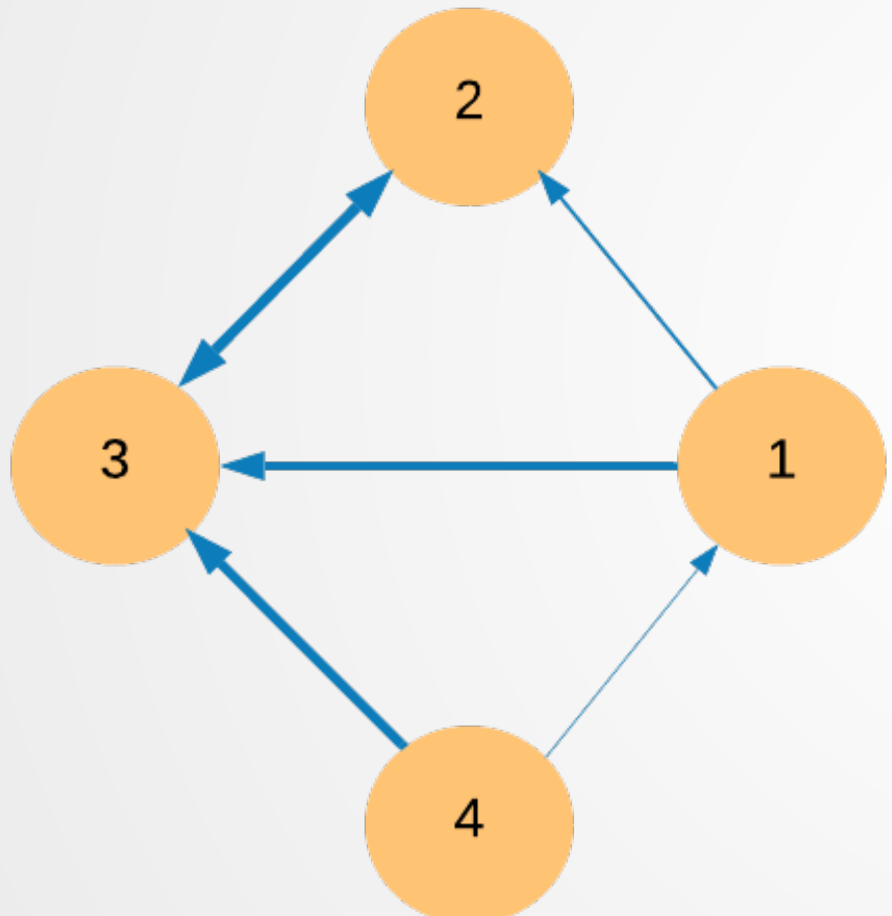


$$A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

METISNET-adjtonet

Asymmetric Adjacency Matrix & Directed Network

Example 8: asymmetric, weighted adjacency matrix with 4 nodes



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

METISNET-adjtonet

Degree Centrality: Directed Cases

in-degree centrality of node v_k:

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

where $a_{i,k} = \mathbf{I}\{(v_i, v_k) \in \mathscr{E}\}$

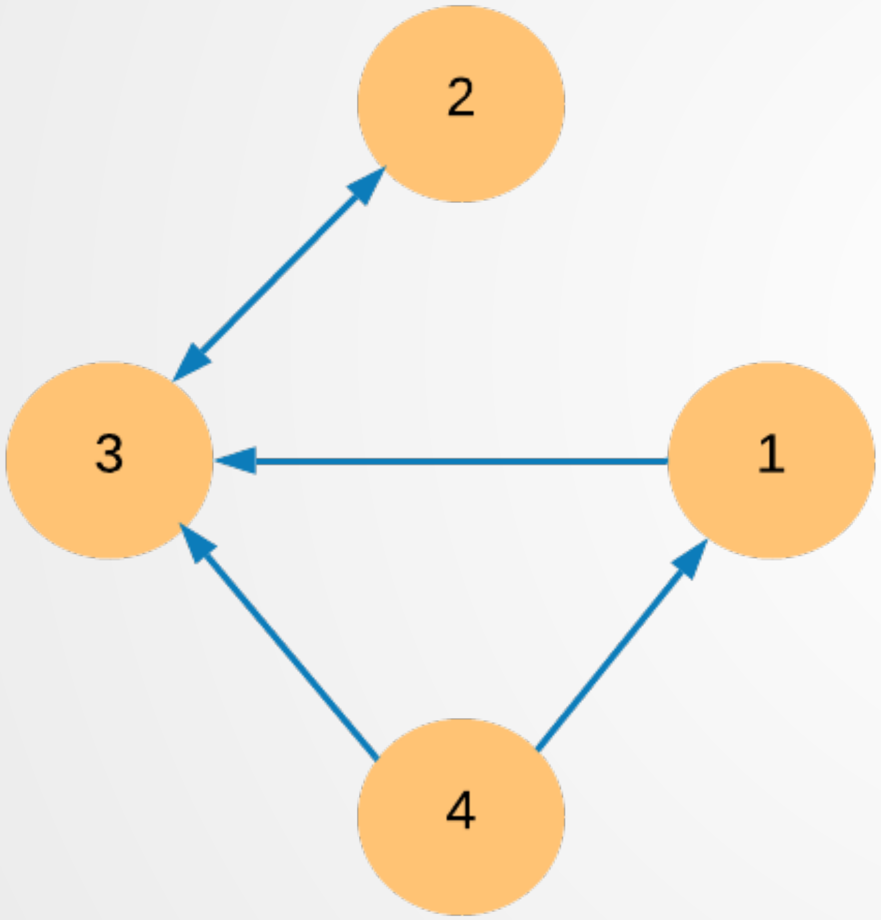
out-degree centrality of node v_k:

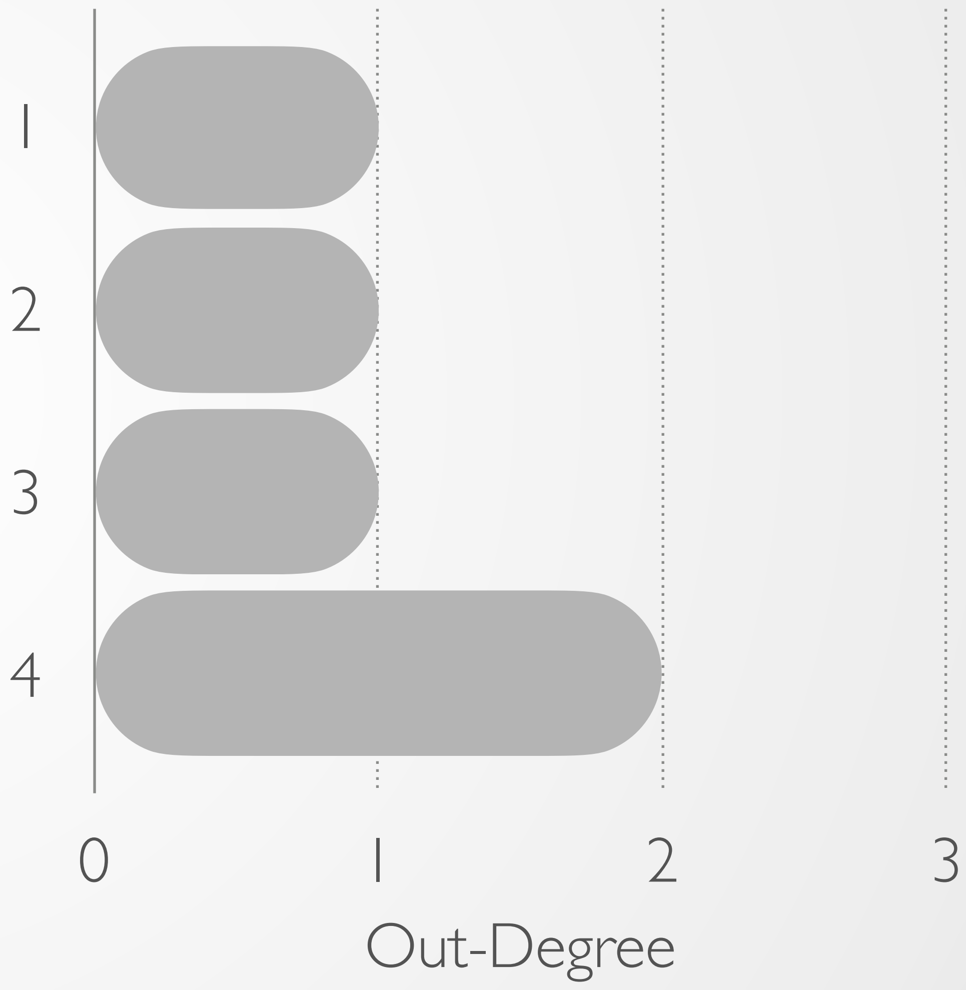
$$C_{OutD}(v_k) = \sum_{i=1}^{n} a_{(v_k, v_i)}$$

where $a_{k,i} = \mathbf{I}\{(v_k, v_i) \in \mathscr{E}\}$

Asymmetric Adjacency Matrix & Directed Network

Example 9: Degree Centrality - Directed Cases





METISNET-adjtonet

Two Views Regarding Graph Centralization

- A graph exhibiting centrality to the degree that all of its points are central
- A network is central to the degree that a single point can control its communication
  - This view is more general

Two Desirable 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 centrality to its maximum possible value for a graph containing the observed number of nodes

Freeman's (1977) measure in graph centrality

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

where n : number of nodes

$C_X(v_i)$: one of the nodes centralities defined above

$C_X(v^*)$: largest value of $C_X(v_i)$ for any node in the network

$\max \sum_{i=1}^{n} \{C_X(v^*) - C_X(v_i)\}$: maximum possible sum of differences in node centrality for a graph of n nodes

Degree-based measures of graph centrality

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

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

Betweenness-based measures of graph centrality

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

$$= \frac{\sum_{i=1}^{n} \{C_B(v^*) - C_B(v_i)\}}{n^3 - 4n^2 + 5n - 2}$$

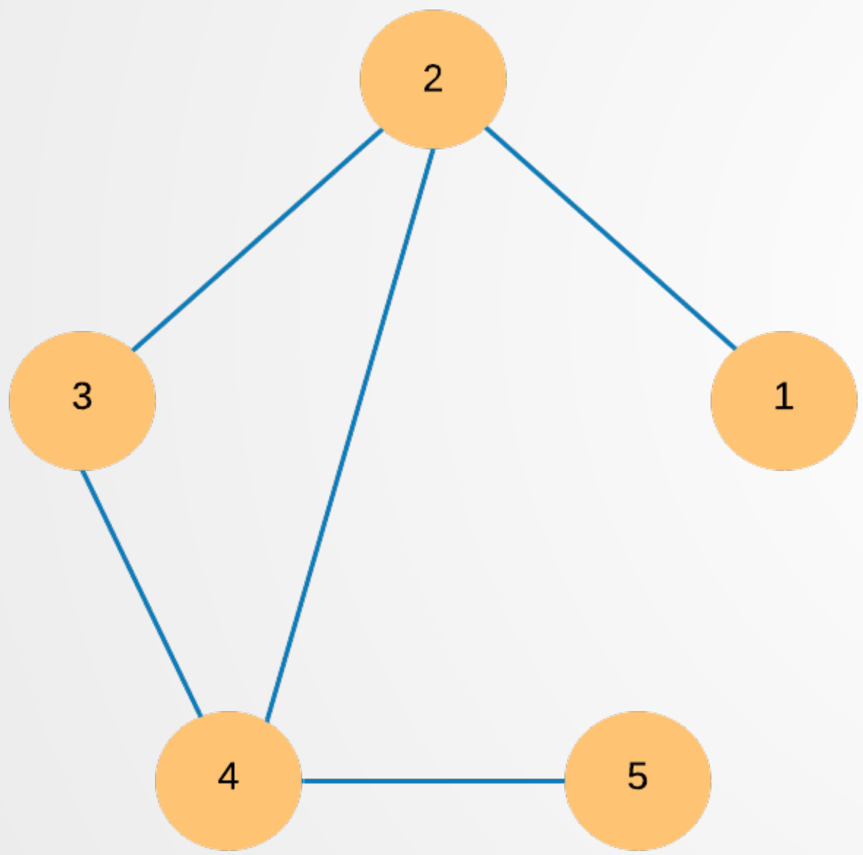
Closeness-based measures of graph centrality

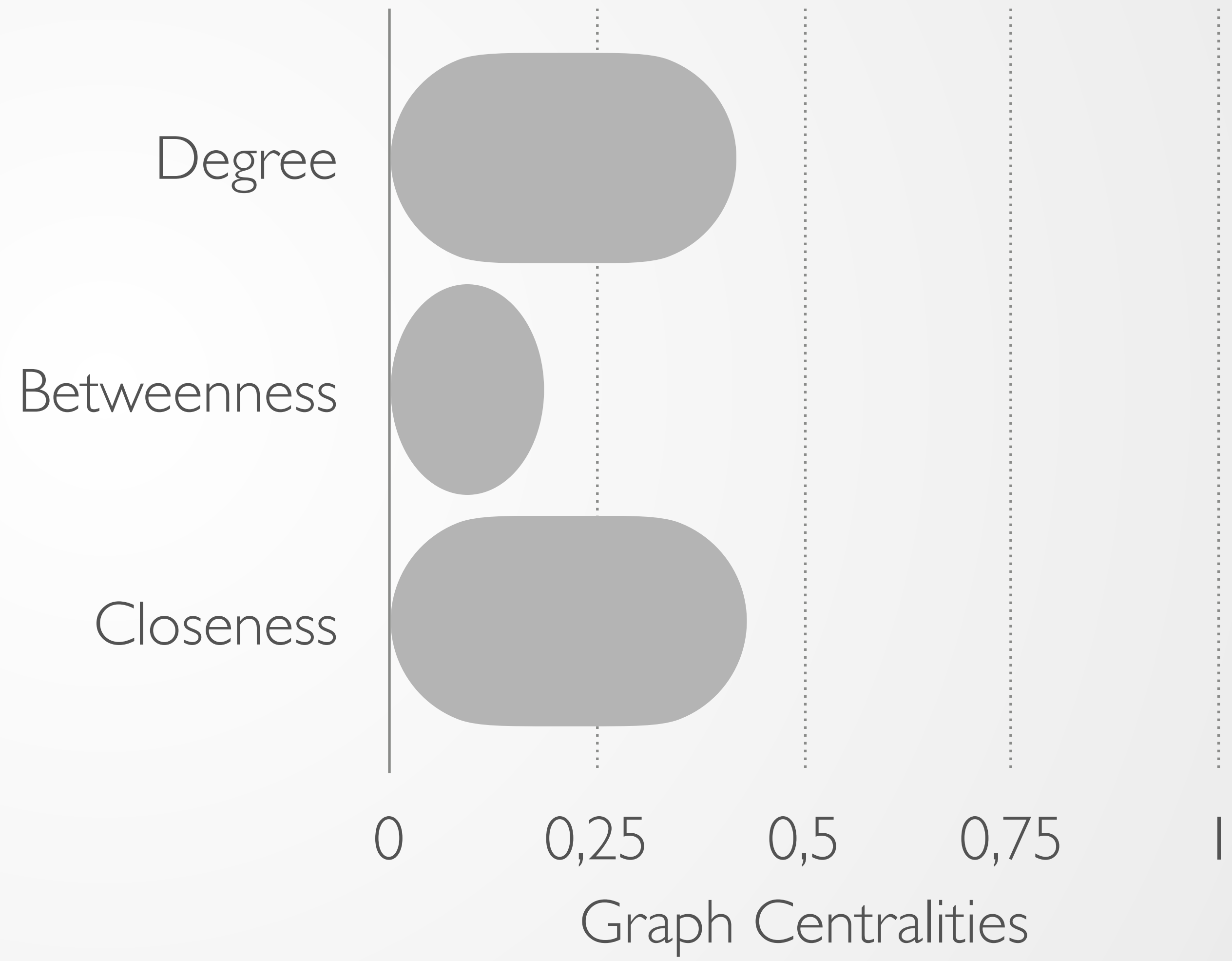
$$C_C = \frac{\sum_{i=1}^{n} \{C_C(v^*) - C_C(v_i)\}}{max \sum_{i=1}^{n} \{C_C(v^*) - C_C(v_i)\}}$$

$$= \frac{\sum_{i=1}^{n} \{C_C(v^*) - C_C(v_i)\}}{(n^2 - 3n + 2)/(2n^2 - 5n + 3)}$$

Examples

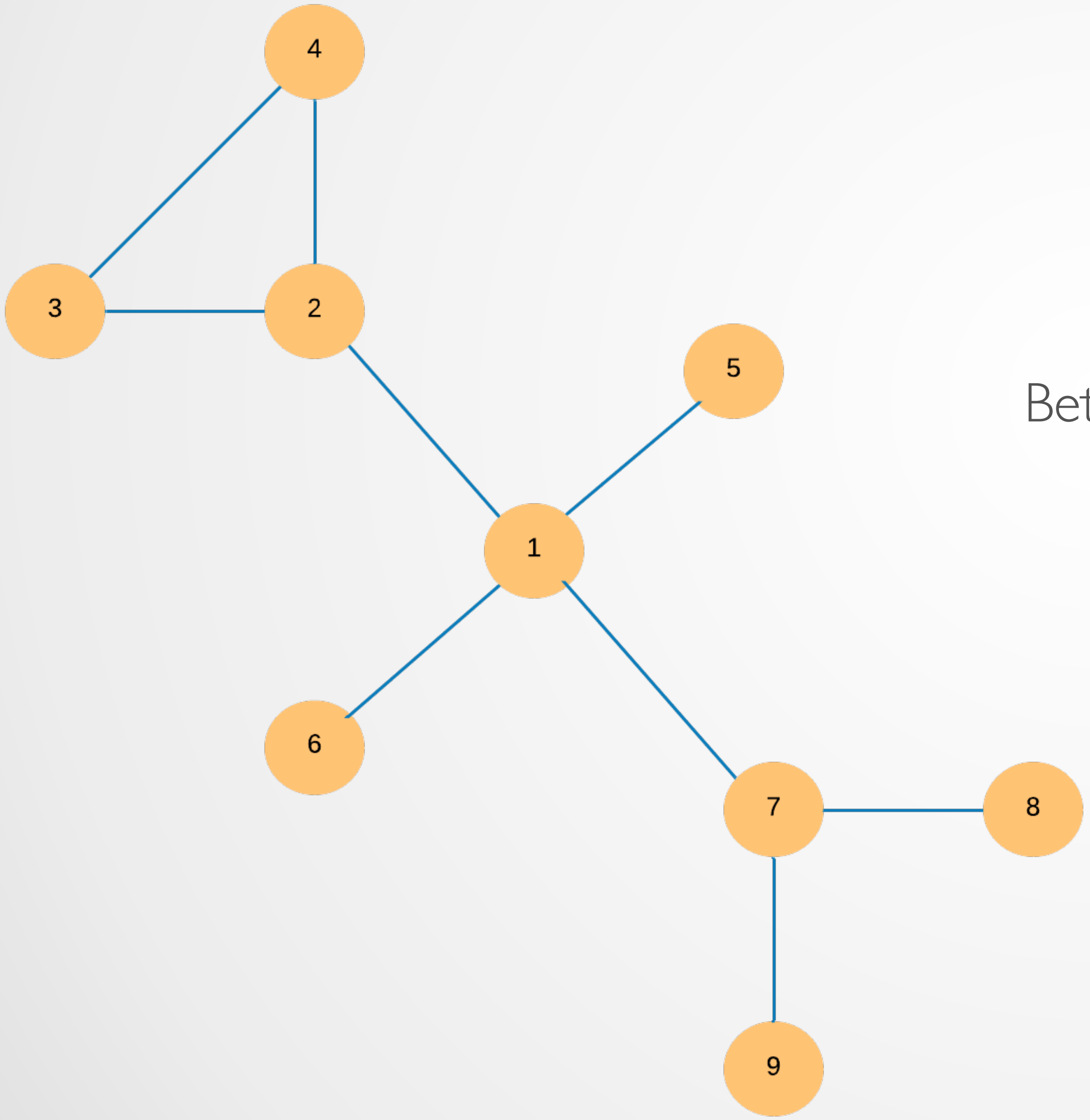
Example 10: Graph Centralities

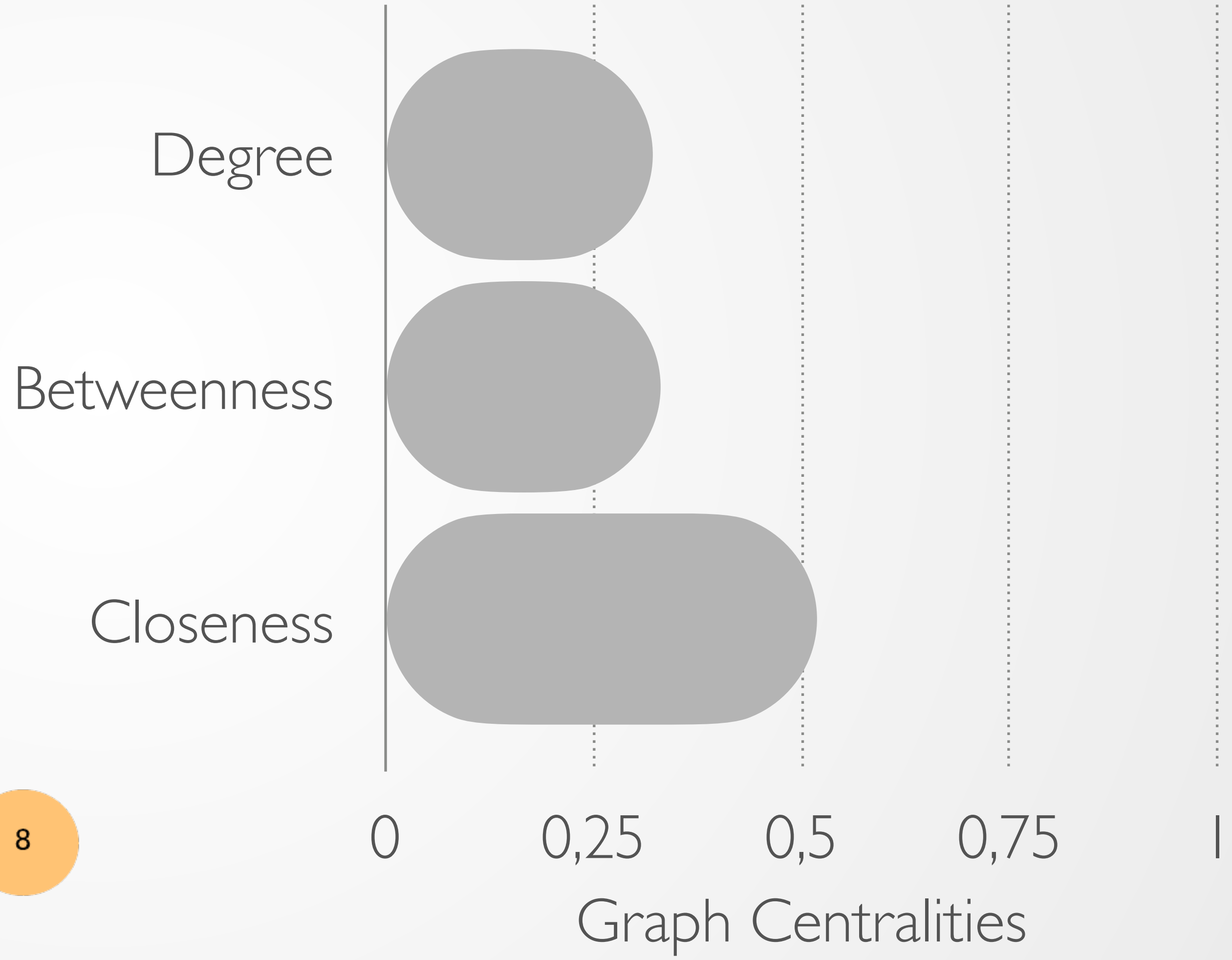




Examples

Example 10: Graph Centralities





Minnesota Road Networks

- Minnesota Road Net data minnesota.mat is available in Matlab R2016b and later versions
- State of Minnesota in the USA
- The data contains a graph object with information about nodes (coordinates of locations) and edges (roads connecting the nodes)

Minnesota Road Networks

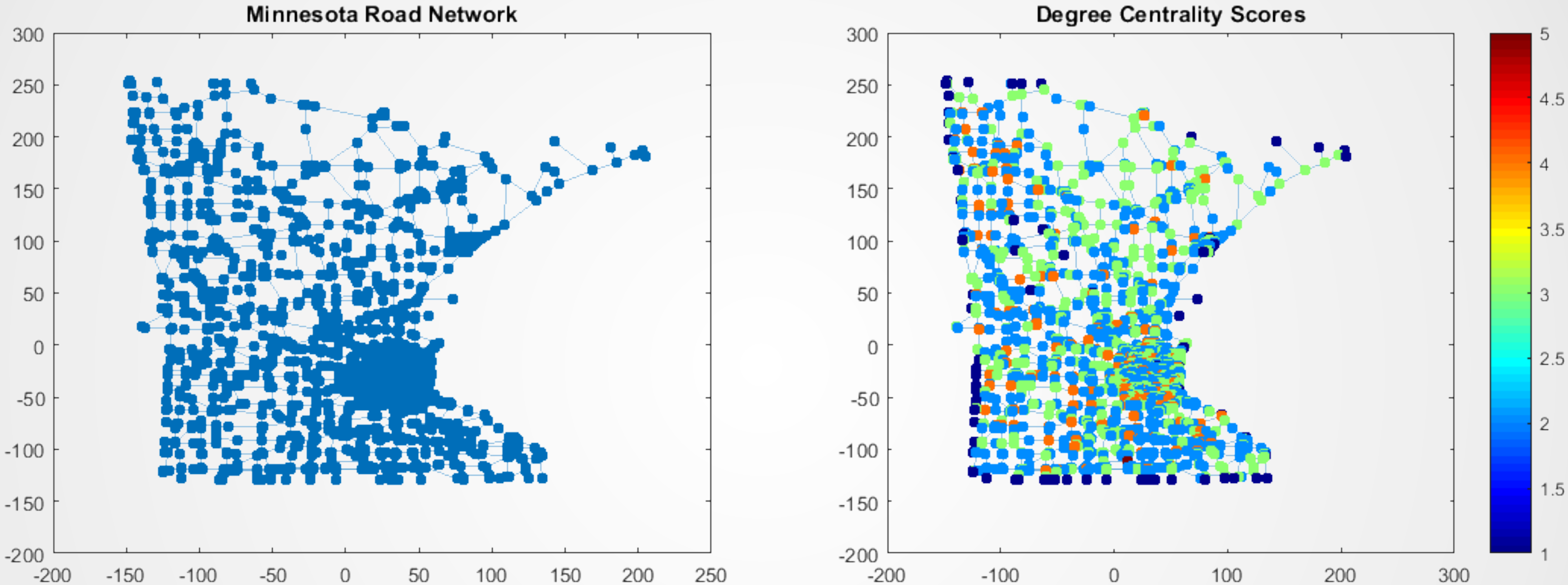


Figure 9: Minnesota Road Net - left: no centrality, right: degree centrality

METISNET-centralitycomparison

Minnesota Road Networks

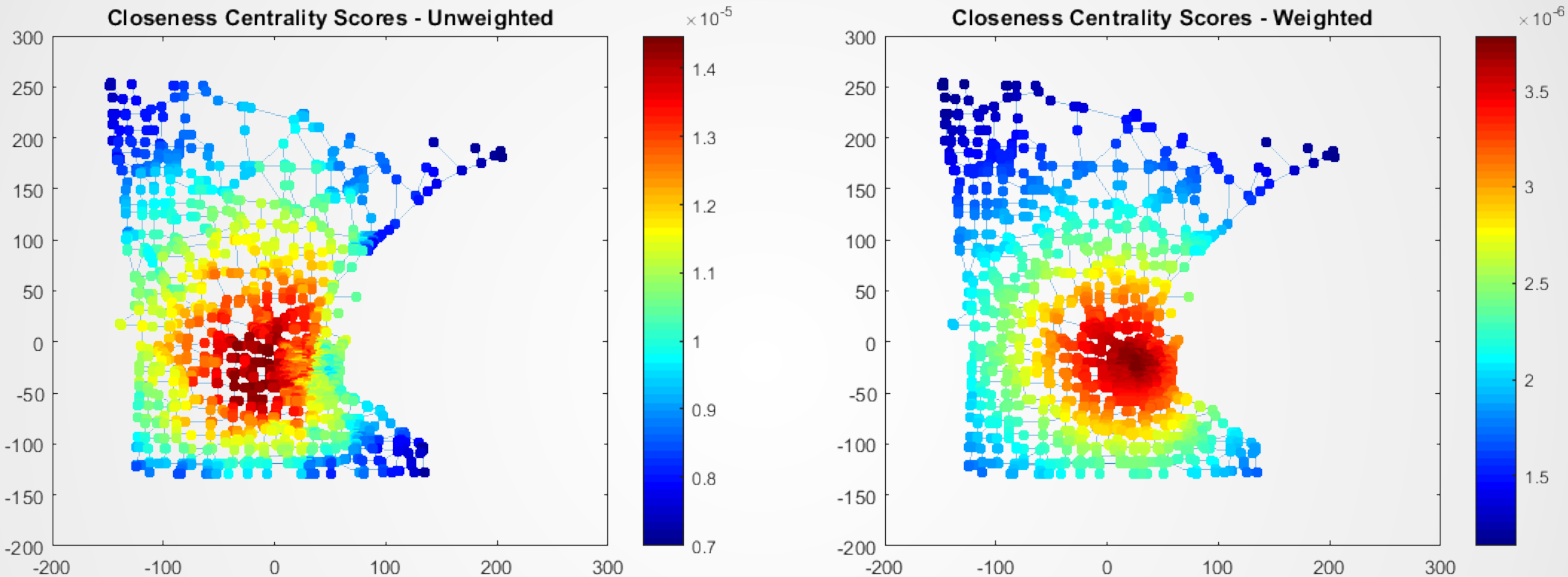


Figure 10: Minnesota Road Net - left: closeness centrality unweighted, right: closeness centrality weighted

METISNET-centralitycomparison

Minnesota Road Networks

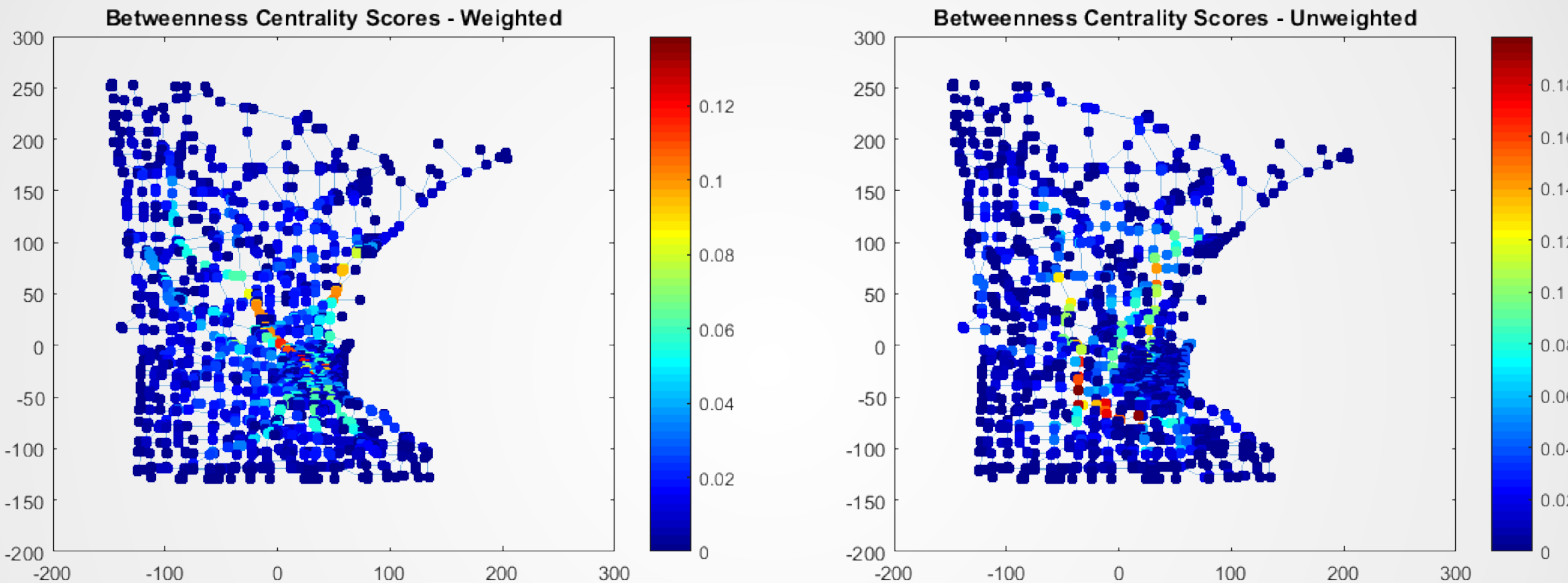


Figure 11: Minnesota Road Net - left: betweenness centrality unweighted, right: betweenness centrality weighted

METISNET-centralitycomparison

Minnesota Road Networks

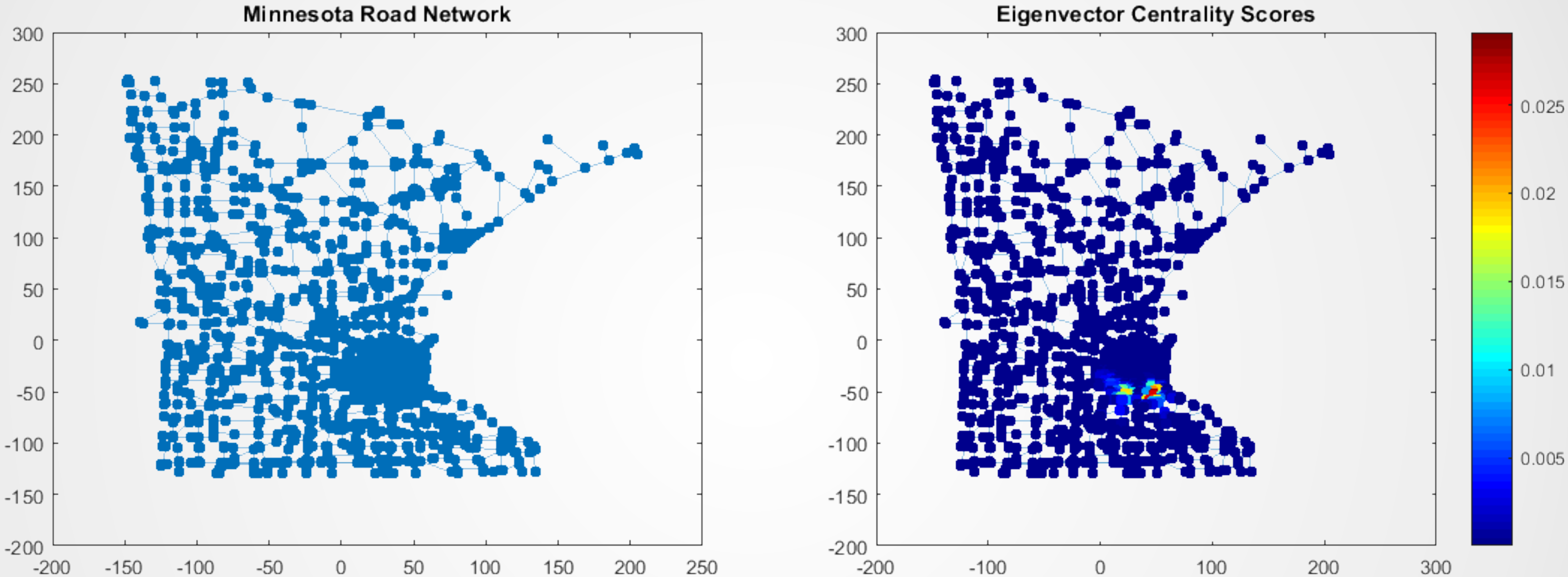


Figure 12: Minnesota Road Net - left: no centrality, right: eigenvector centrality

METISNET-centralitycomparison

Systemic Risk with Neural Network Quantile Regression

- Approach and data from Georg Keilbar (2018)
- Eight globally important banks are chosen in the analysis:

  Citibank (C), Bank of America Corporation (BAC), JPMorgan Chase & Co (JPM), Wells Fargo & Co (WFC), Morgan Stanley (MS), Goldman Sachs Group Inc (GS), Bank of New York Mellon (BK), State Street Corporation (STT)
- Daily data from 2008-01-02 to 2018-05-31
- node size: indicates centrality
- edge width: link strength

Financial Risk Contagion Networks

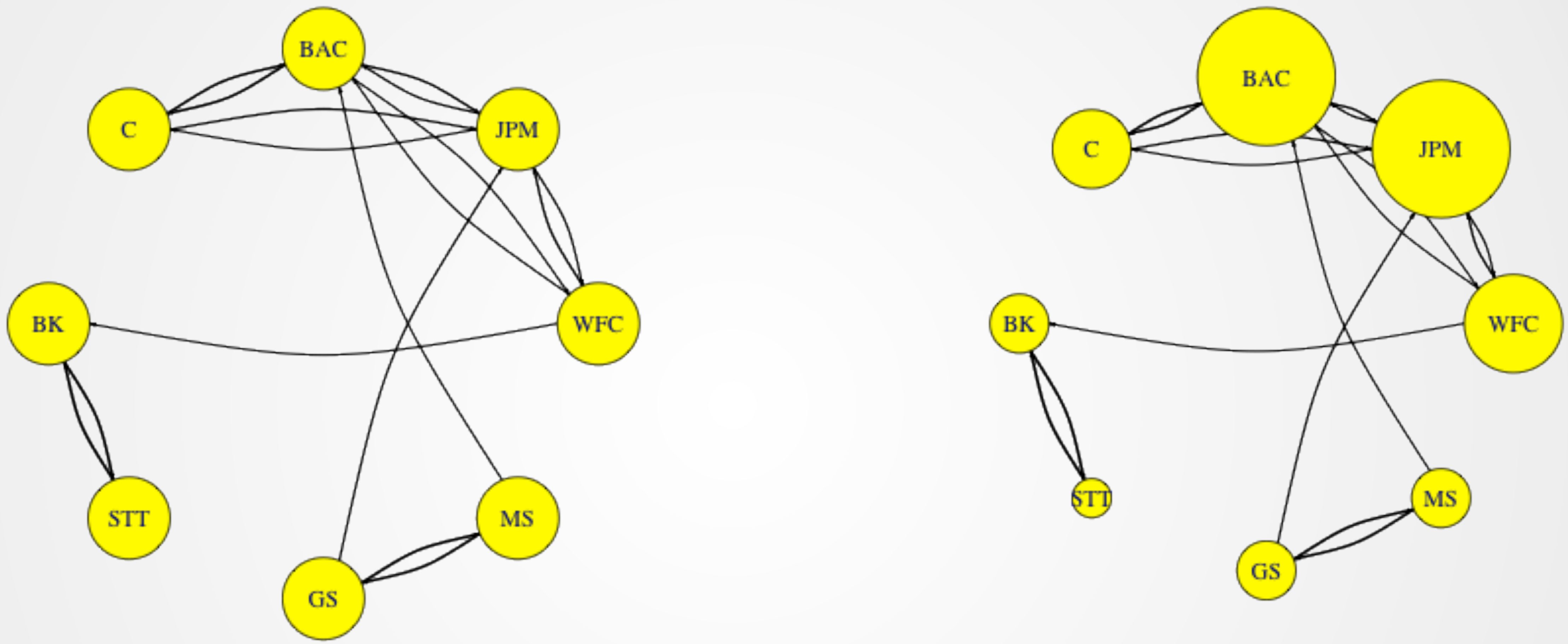


Figure 13: Financial Risk Contagion Network - left: no centrality, right: total degree centrality

METISNET-NNQR-centrality

Financial Risk Contagion Networks

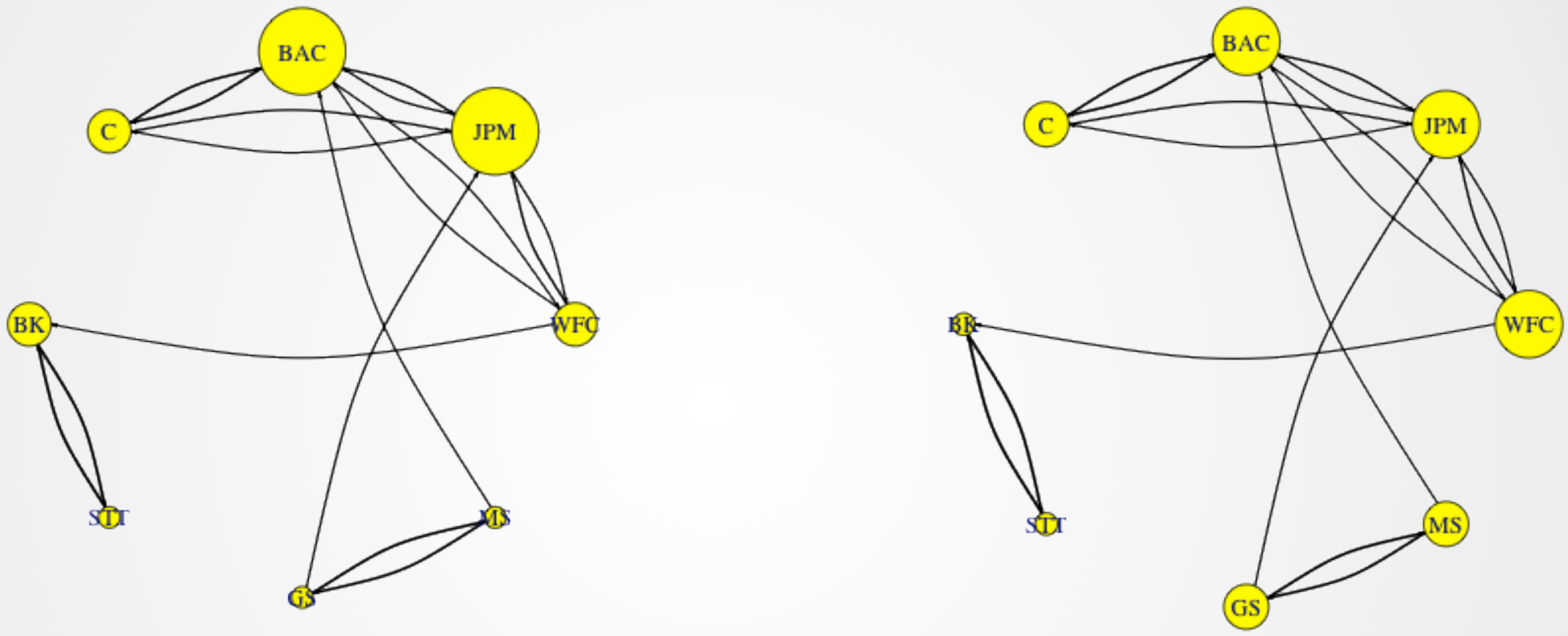


Figure 14: Financial Risk Contagion Network - left: in degree centrality, right: out degree centrality

METISNET-NNQR-centrality

Financial Risk Contagion Networks

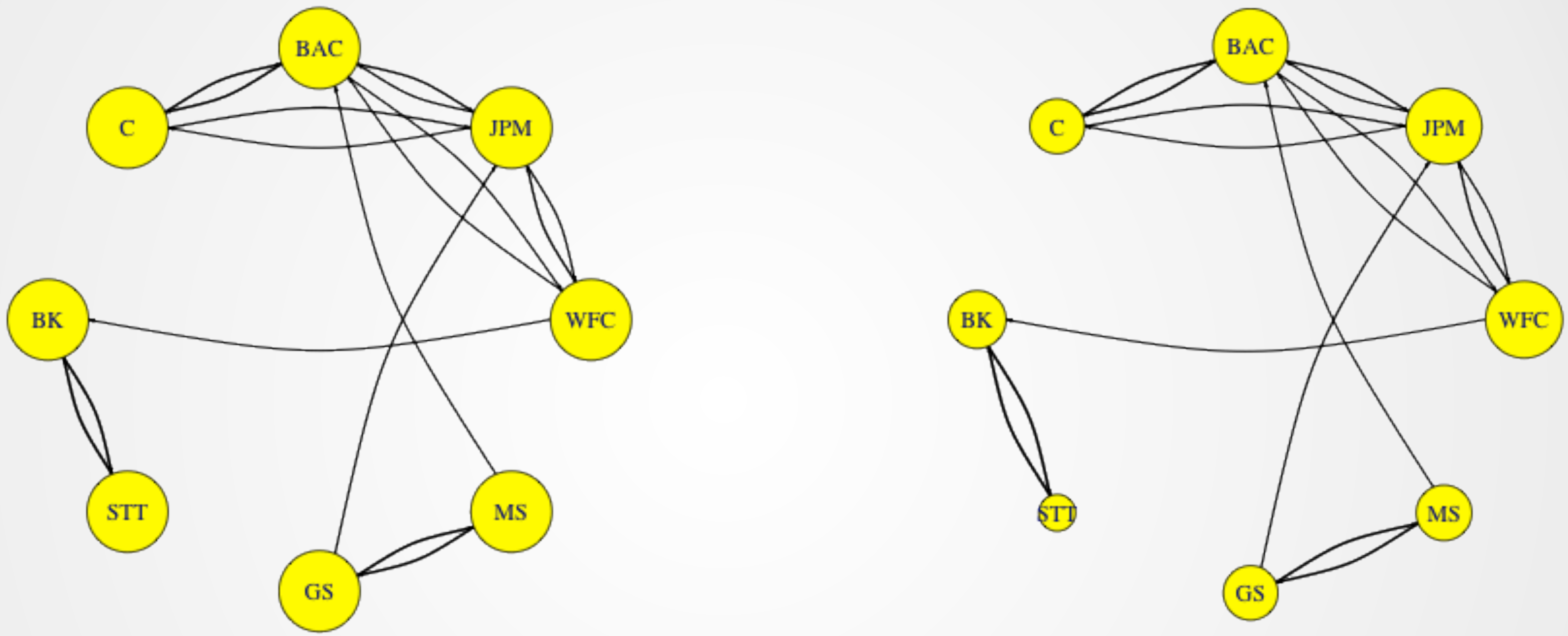


Figure 15: Financial Risk Contagion Network - left: no centrality, right: total closeness centrality

Financial Risk Contagion Networks

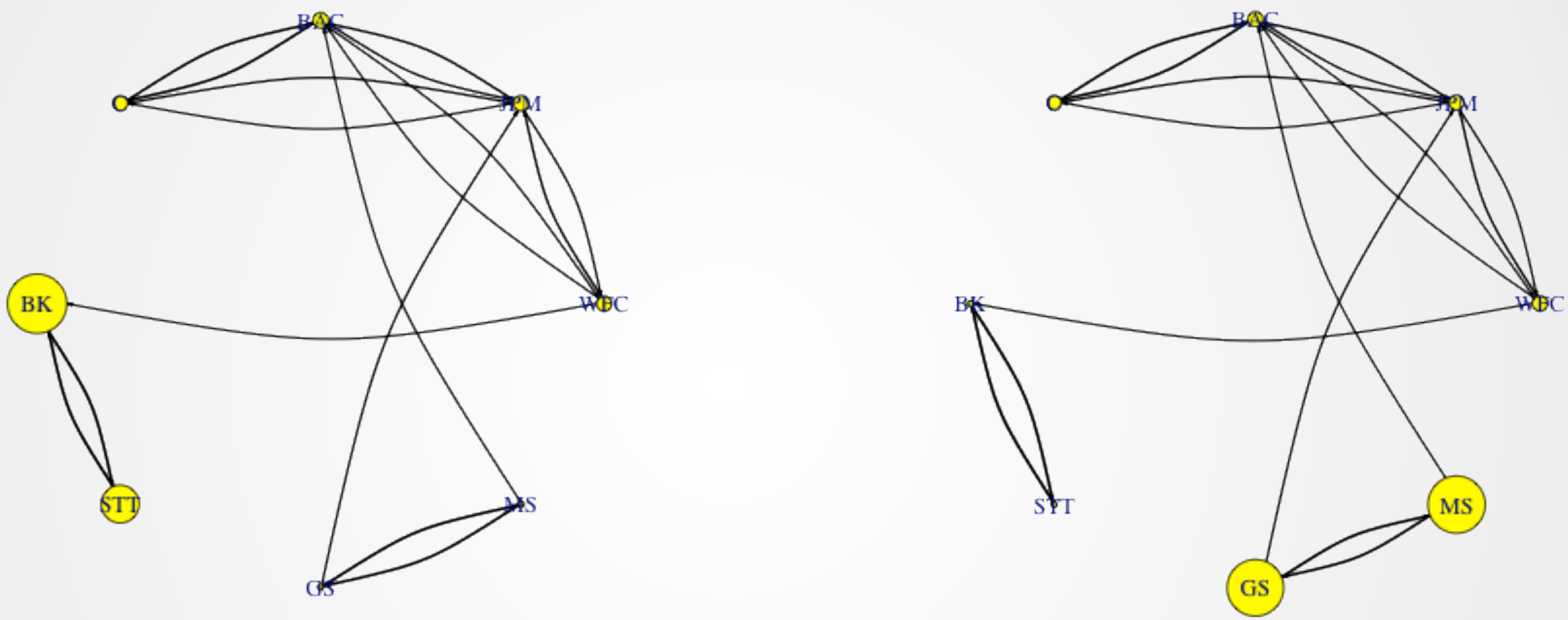


Figure 16: Financial Risk Contagion Network - left: in closeness centrality, right: out closeness centrality

Financial Risk Contagion Networks

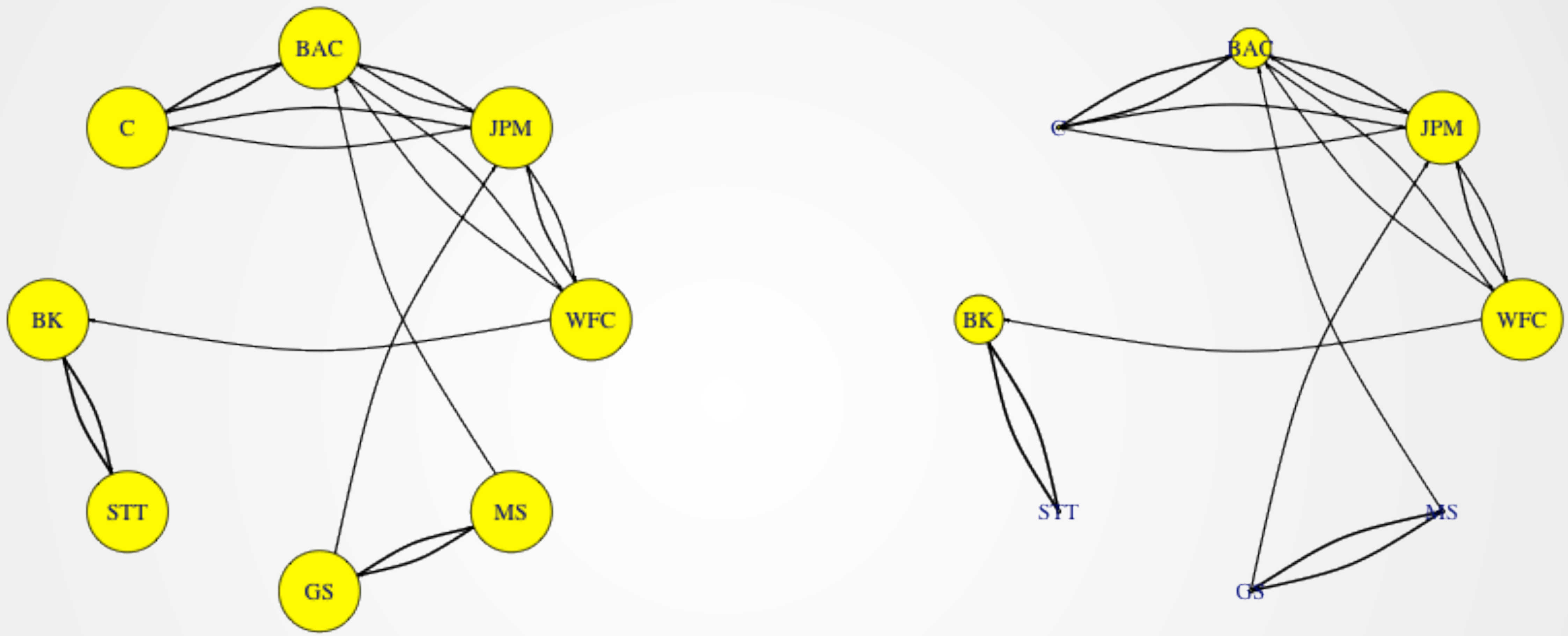


Figure 17: Financial Risk Contagion Network - left: no centrality, right: directed betweenness centrality

Financial Risk Contagion Networks

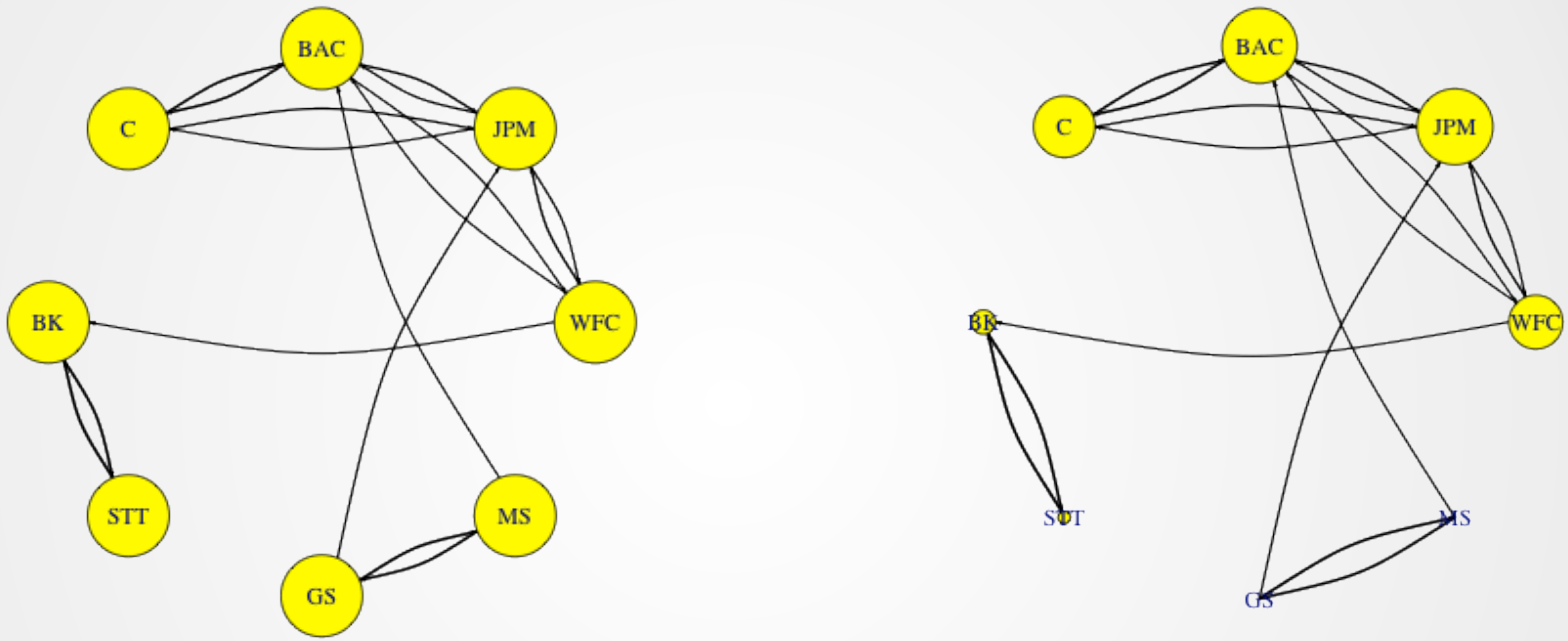


Figure 18: Financial Risk Contagion Network - left: no centrality, right: directed eigenvector centrality

TENET

- Data from TENET
- Nodes: Financial institutions
- Edges: risk spillover
- Edge width: edge weight
- Showing edges whose weight is at least the mean of the 100 heaviest edges
- Groups:
  - left: others
  - top: depositories
  - bottom: broker-dealers
  - right: insurers

TENET_group_network

TENET

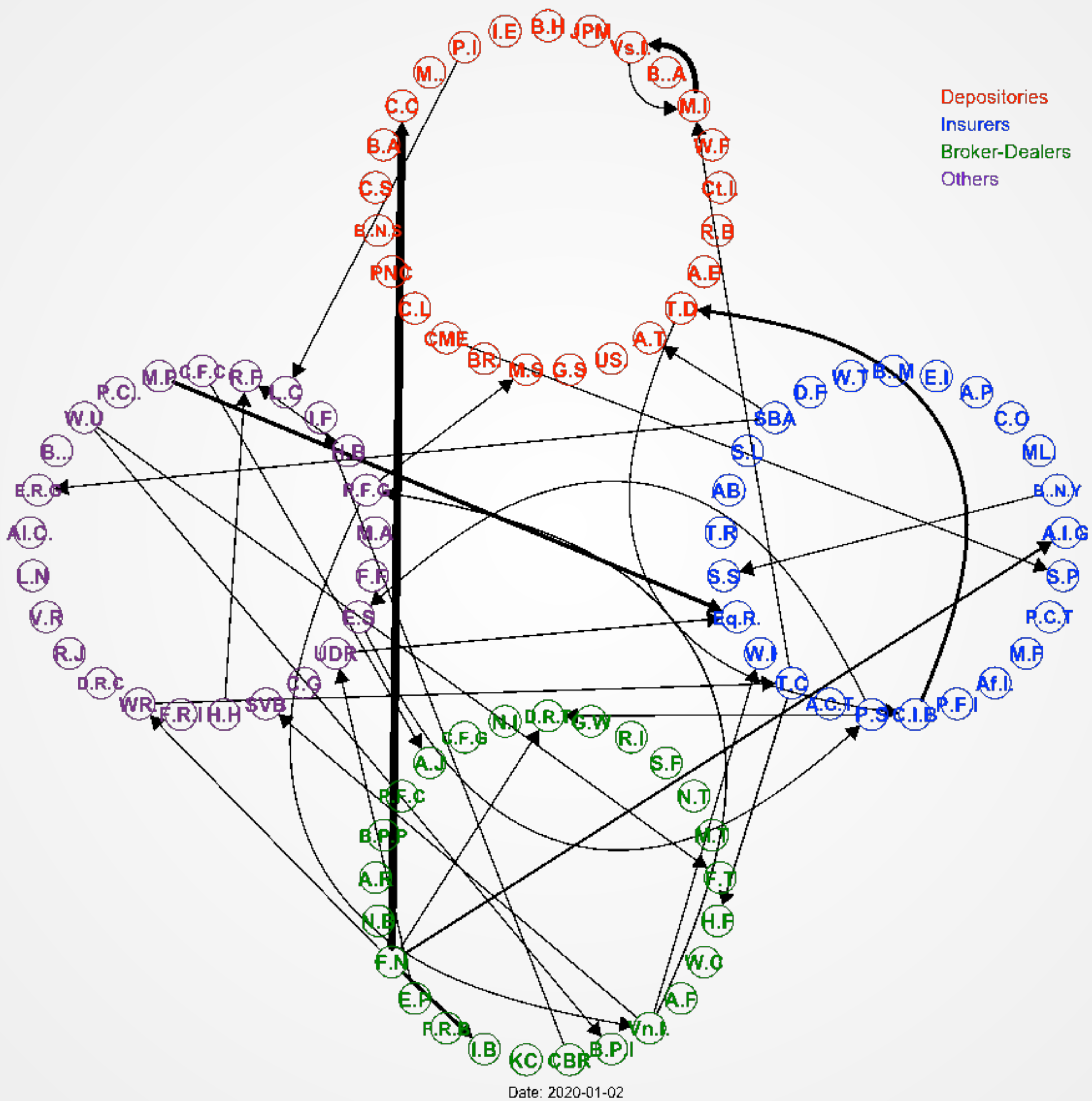


Figure 19: TENET - network movie

TENET

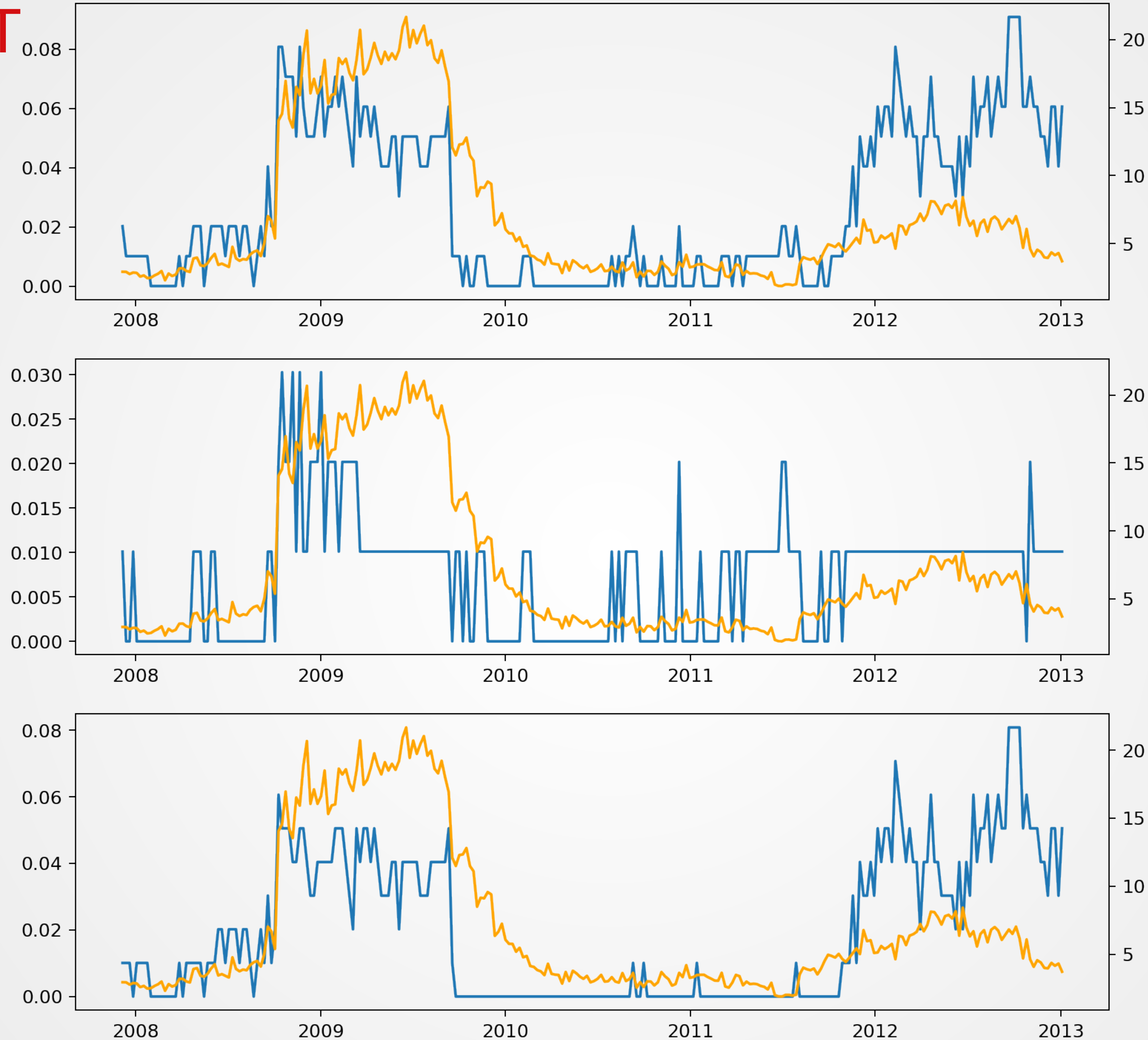


Figure 20: TENET - MS Morgan Stanley Centrality vs FRM Lambda

TENET

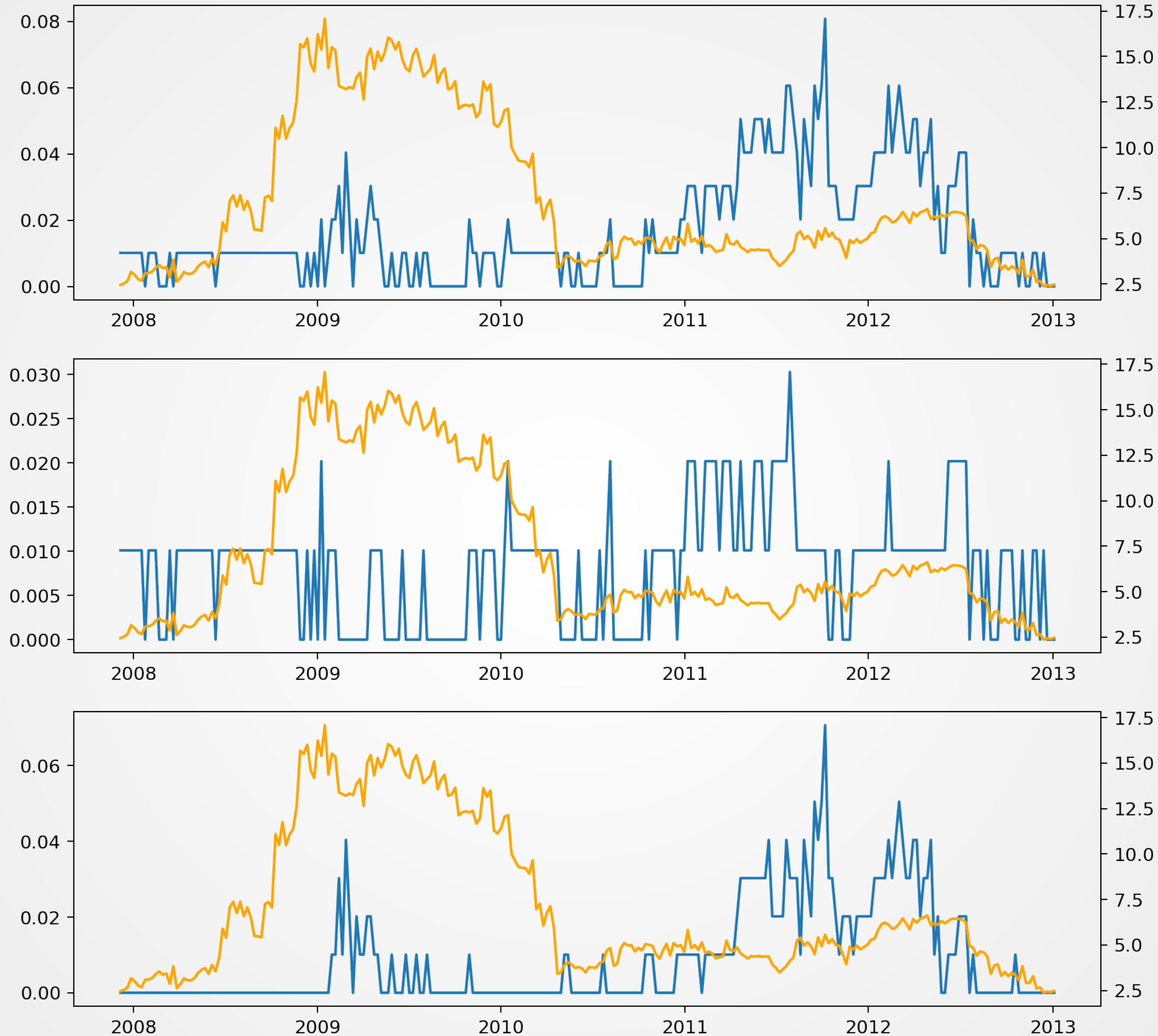
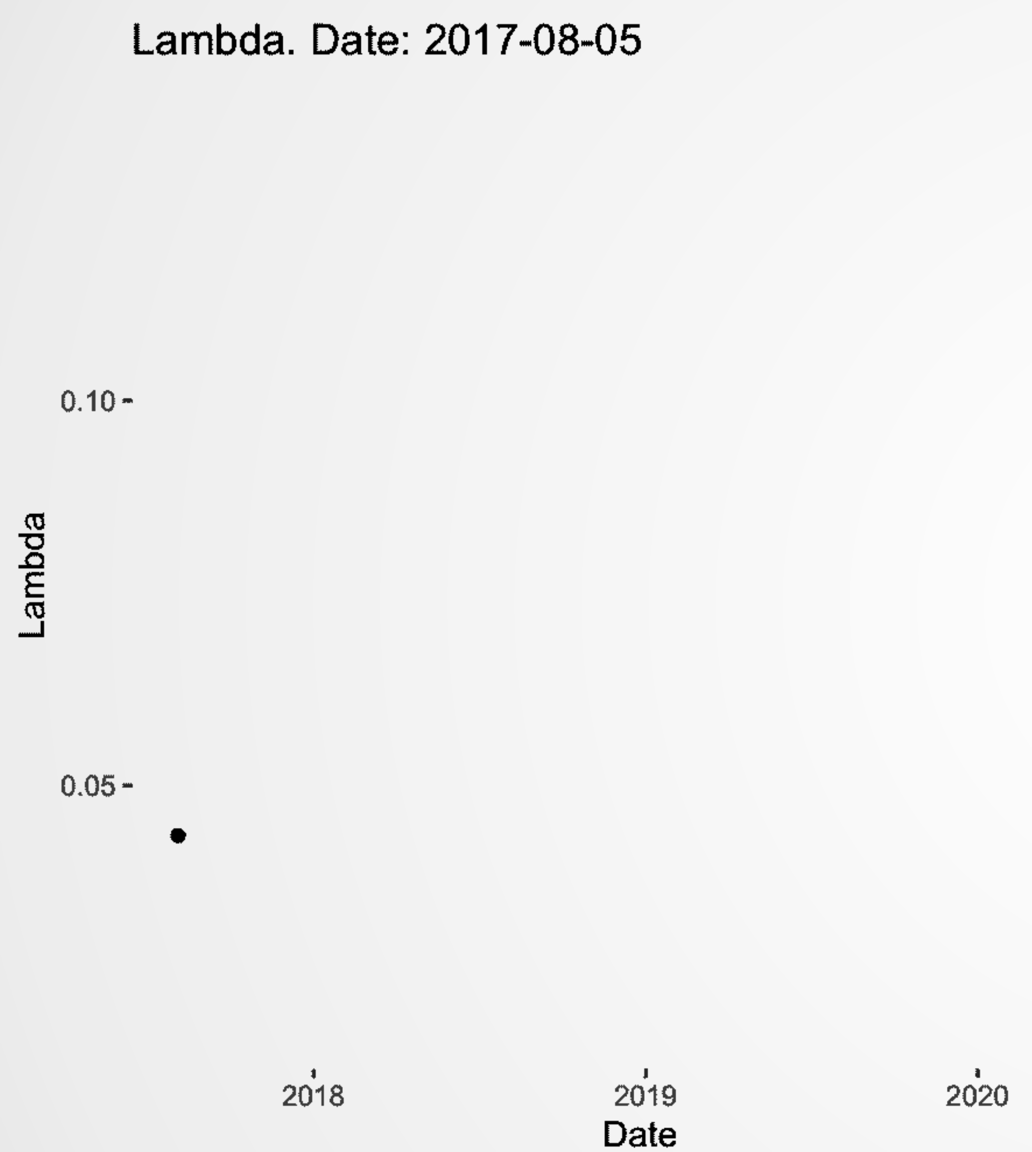


Figure 21: TENET - STI SunTrust Banks Centrality vs FRM Lambda

FRM Crypto



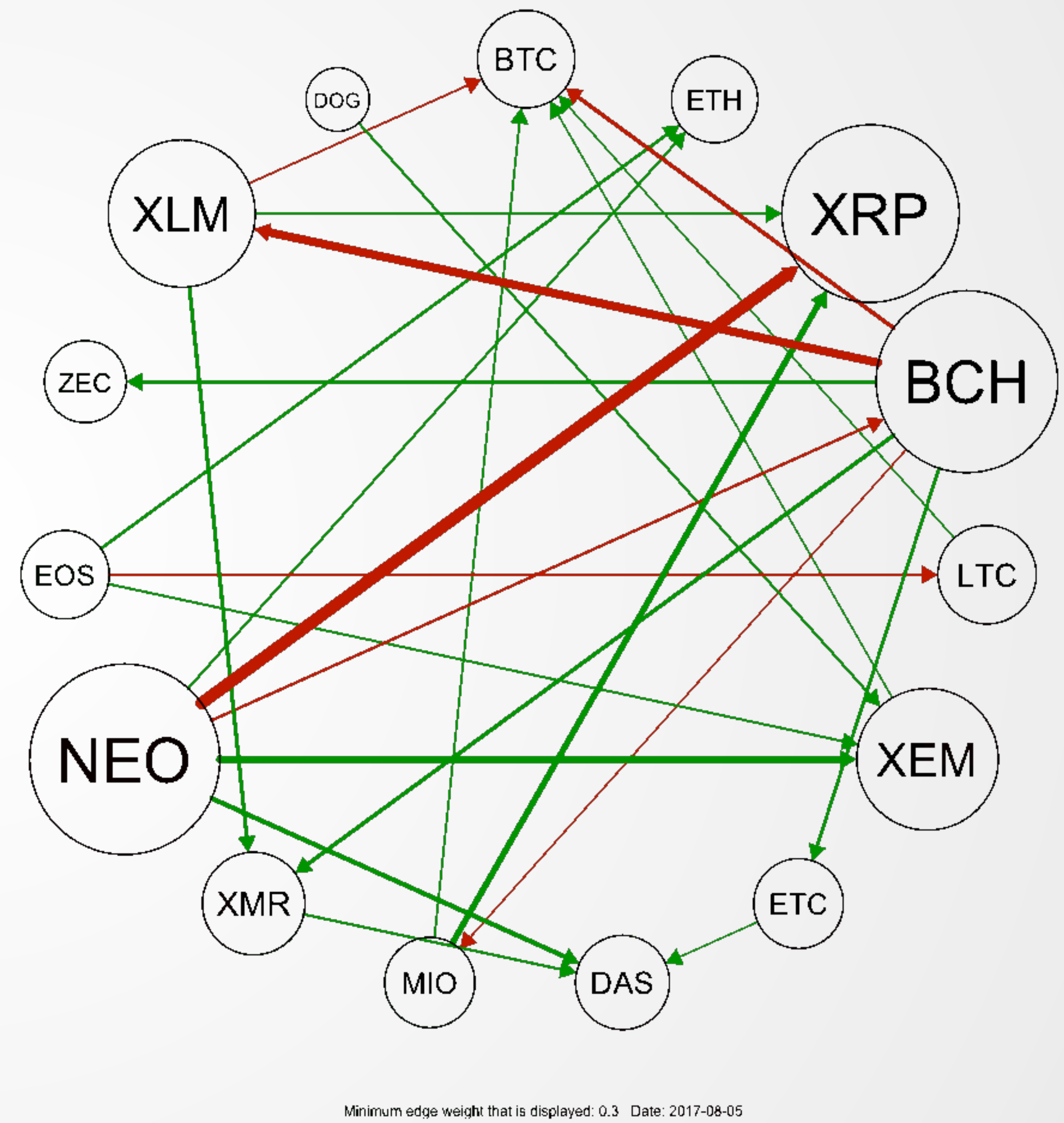


Figure 22: TENET - movie and λ

CRIX Correlation Movie

- Rolling window correlation of Q1 constituents as adjacency matrix
- Edges: correlation intensity
- Color/size: degree centrality

Crix_Network_Plot

CRIX Correlation Movies

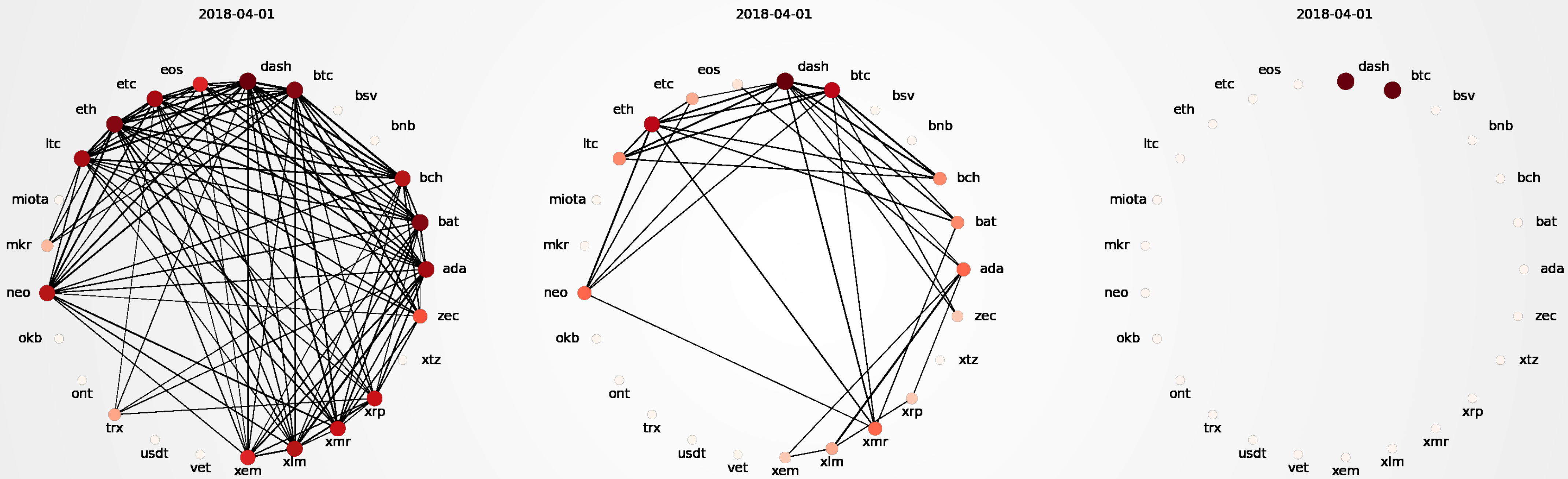


Figure 23: CRIX Movies - window of 90d, threshold: mean (left), mean plus 0.5 std (center), mean plus 1 std (right)

CRIX Correlation Movies

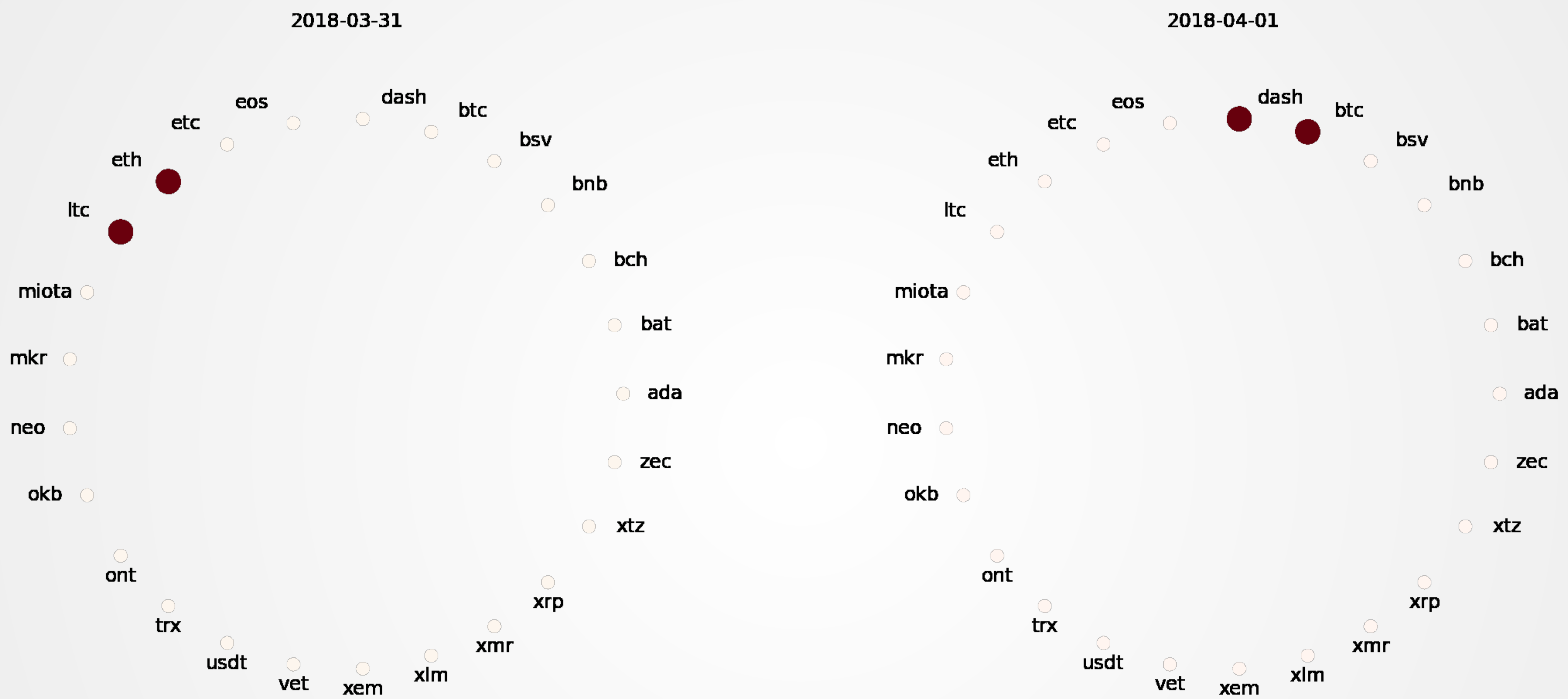


Figure 24: CRIX Movies - threshold mean plus 1 std, window of 30d (left) and 90d (right)

CRIX Correlation Movies

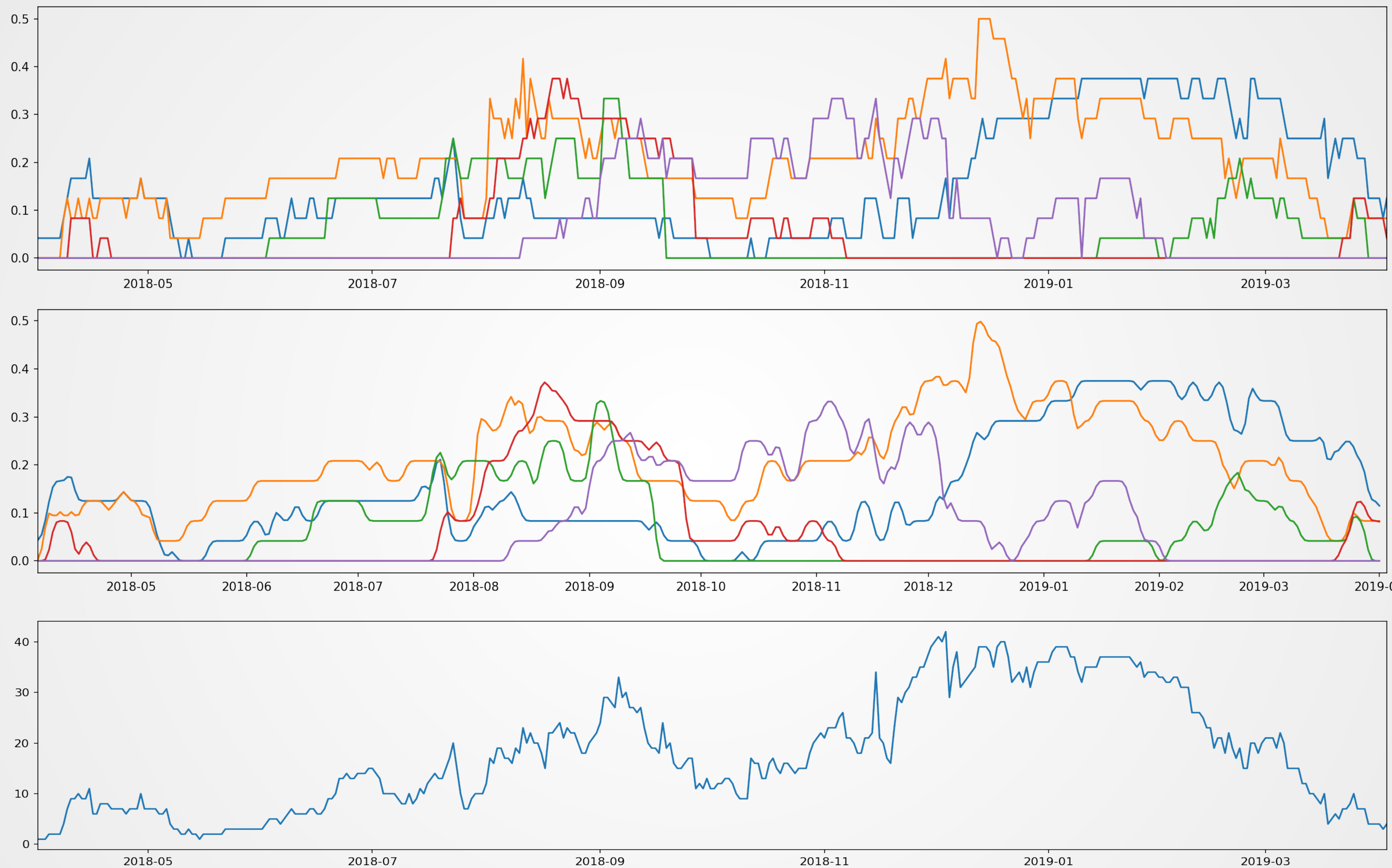


Figure 25: Centrality over time for top 5 CCs, window of 30d, threshold one std over mean - raw (top), smoothed (center), number of edges (bottom)

References

Anthonisse, J.M., 1971. The rush in a directed graph. Stichting Mathematisch Centrum. Mathematische Besliskunde, (BN 9/71)

Bavelas, A., 1950. Communication patterns in task-oriented groups. *The journal of the acoustical society of America*, *22*(6), pp.725-730 DOI: https://doi.org/10.1121/1.1906679

Beauchamp, M.A., 1965. An improved index of centrality. *Behavioral science*, *10*(2), pp.161-163. DOI: https://doi.org/10.1002/bs.3830100205

Freeman, L.C., 1977. A set of measures of centrality based on betweenness. *Sociometry*, pp.35-41. DOI: http://doi.org/10.2307/3033543

Härdle, W.K., Wang, W. and Yu, L., 2016. Tenet: Tail-event driven network risk. *Journal of Econometrics*, *192*(2), pp.499-513. DOI: https://doi.org/10.1016/j.jeconom.2016.02.013

Katz, L., 1953. A new status index derived from sociometric analysis. *Psychometrika*, *18*(1), pp.39-43. https://doi.org/10.1007/BF02289026

References

Keilbar, G., 2018. Modelling Systemic Risk using Neural Network Quantile Regression (Master's thesis, Humboldt-Universität zu Berlin). DOI: https://doi.org/10.18452/19318

Moxley, R.L. and Moxley, N.F., 1974. Determining point-centrality in uncontrived social networks. *Sociometry*, pp.122-130. DOI: http://doi.org/10.2307/2786472

Newman, M.E., 2006. Finding community structure in networks using the eigenvectors of matrices. *Physical review E*, *74*(3), p.036104. DOI: https://doi.org/10.1103/PhysRevE.74.036104

Nieminen, J., 1974. On the centrality in a graph. *Scandinavian journal of psychology*, *15*(1), pp.332-336. DOI: https://doi.org/10.1111/j.1467-9450.1974.tb00598.x

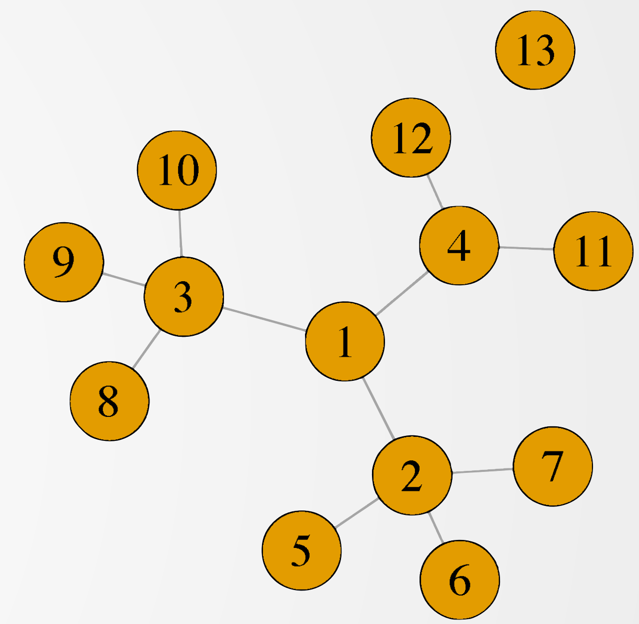
Rogers, D.L., 1974. Sociometric analysis of interorganizational relations: application of theory and measurement. *Rural Sociology*, *39*(4), p.487. DOI: https://search.proquest.com/docview/1290956816?accountid=11531

Shaw, M.E., 1954. Group structure and the behavior of individuals in small groups. *The Journal of psychology*, *38*(1), pp.139-149. DOI: https://doi.org/10.1080/00223980.1954.9712925

Sabidussi, G., 1966. The centrality index of a graph. *Psychometrika*, *31*(4), pp.581-603. DOI: https://doi.org/10.1007/BF02289527

Rodrigues, F.A., 2019. Network centrality: an introduction. In A Mathematical Modeling Approach from Nonlinear Dynamics to Complex Systems (pp. 177-196). Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-78512-7_10

Network Centrality



Ya QIAN
Wolfgang Karl Härdle
Lucas Valentin Umann

Ladislaus von Bortkiewicz Professor of Statistics
Humboldt-Universität zu Berlin
Blockchain Research Center
lvb.wiwi.hu-berlin.de
Charles University, WISE XMU, NCTU 玉山學者