

Network Properties

Network Application Diagnostics

B2M32DSAA

Radek Mařík

Czech Technical University
Faculty of Electrical Engineering
Department of Telecommunication Engineering
Prague CZ

October 16, 2025



1 Graph Matrices

- Network Matrices

2 Centrality Measures

- Path Based Centralities
- Spectral Centralities
- Example

Outline

1 Graph Matrices

- Network Matrices

2 Centrality Measures

- Path Based Centralities
- Spectral Centralities
- Example



Adjacency Matrix ^[New10, EK10]

- The **adjacency matrix** $\mathbf{A}$ of a *simple* graph is the $N \times N$ matrix with element A_{ij} such that

$$A_{ij} = \begin{cases} 1 & \text{if there is an edge between vertices } j \text{ and } i, \\ 0 & \text{otherwise} \end{cases}$$

- The adjacency matrix of a *directed* network has matrix elements

$$A_{ij} = \begin{cases} 1 & \text{if there is an edge from } j \text{ to } i, \\ 0 & \text{otherwise} \end{cases}$$



Cocitation Matrix ^[New10]

- Convenient to turn a directed network into an undirected one for the purposes of analysis
- The **cocitation** of two vertices i and j in a directed network is the number of vertices that have outgoing edges pointing to both i and j .
 - The cocitation of two papers is the number of other papers that cite both.
 - $A_{ik}A_{jk} = 1$ if i and j are both cited by k and zero otherwise.
- The cocitations C_{ij} of i and j is

$$C_{ij} = \sum_{k=1}^N A_{ik}A_{jk} = \sum_{k=1}^N A_{ik}A_{kj}^T$$

- The **cocitation matrix** $\mathbf{C}$ is the $N \times N$ matrix with elements C_{ij} , i.e.

$$\mathbf{C} = \mathbf{A}\mathbf{A}^T$$

- $\mathbf{C}$ is a symmetric matrix: $\mathbf{C}^T = (\mathbf{A}\mathbf{A}^T)^T = \mathbf{A}\mathbf{A}^T = \mathbf{C}$



Bibliographic Coupling ^[New10]

- The **bibliographic coupling** of two vertices in a directed network is the number of other vertices to which both point.
 - For instance in a citation network: the bibliographic coupling of two papers i and j is the number of other papers that are cited by both i and j .
 - $A_{ki}A_{kj} = 1$ if i and j both cite k and zero otherwise.
- The bibliographic coupling B_{ij} of i and j is

$$B_{ij} = \sum_{k=1}^N A_{ki}A_{kj} = \sum_{k=1}^N A_{ik}^T A_{kj}$$

- The **bibliographic coupling matrix** $\mathbf{B}$ is the $n \times n$ matrix with elements B_{ij} , i.e.

$$\mathbf{B} = \mathbf{A}^T \mathbf{A}$$

- $\mathbf{B}$ is a symmetric matrix: $\mathbf{B}^T = (\mathbf{A}^T \mathbf{A})^T = \mathbf{A}^T \mathbf{A} = \mathbf{B}$



Bi-adjacency Matrix ^[New10, BJP17]

Bipartite networks

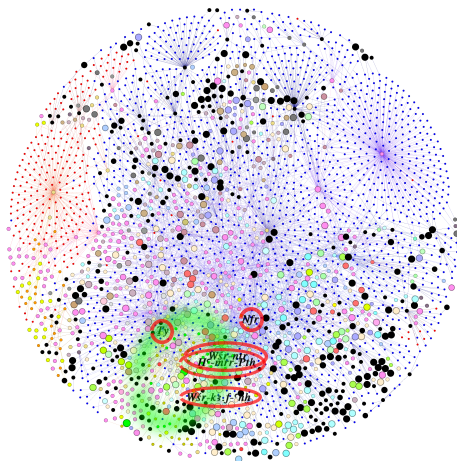
- also called **two-mode** networks in SNA ^[New10]
 - $V = V_1 \cup V_2, V_1 \cap V_2 = \emptyset$
 - movies $\times$ actors
 - articles $\times$ authors
 - timestamps $\times$ active Wifi access points (AP)
 - people $\times$ groups
- Let $N_1 = |V_1|$ and $N_2 = |V_2|$,
then the **bi-adjacency** matrix $\mathbf{B}$ ^[BJP17] is $N_1 \times N_2$ matrix having elements

$$B_{ij} = \begin{cases} 1 & \text{if there is an edge between vertices } n_i \in V_1 \text{ and } n_j \in V_2, \\ 0 & \text{otherwise} \end{cases}$$

- Also called **incidence matrix** ^[New10], **bipartite adjacency matrix** ^[BM08]

Úředníci 5. dynastie a jejich tituly

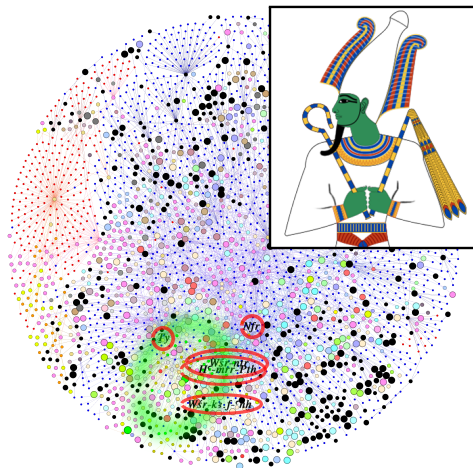
- bipartitní síť:
úředníci a jejich tituly
- modré malé tečky ... muži
- červené malé tečky ... ženy
- zelené malé tečky ... vezíři
- žluté malé tečky ... královští
- azurové kroužky ... soudci
- fialové kroužky ... kněží
- červené kroužky ... pokladnice
- světle zelené kroužky ... sýpka
- tmavě zelené kroužky ... vezíři
- ...
- červené elipsy ... vysokí hodnostáři



Fruchtermanovo-Reingoldovo rozvržení
(pružiny)

Úředníci 5. dynastie a jejich tituly

- bipartitní síť:
úředníci a jejich tituly
- modré malé tečky ... muži
- červené malé tečky ... ženy
- zelené malé tečky ... vezíři
- žluté malé tečky ... královští
- azurové kroužky ... soudci
- fialové kroužky ... kněží
- červené kroužky ... pokladnice
- světle zelené kroužky ... sýpka
- tmavě zelené kroužky ... vezíři
- ...
- červené elipsy ... vysokí hodnostáři



Fruchtermanovo-Reingoldovo rozvržení
(pružiny)

Adjacency and Bi-adjacency Matrix ^[New10, BJP17]

$$\mathbf{A} = \begin{pmatrix} \emptyset_{|V_1|} & \mathbf{B} \\ \mathbf{B}^T & \emptyset_{|V_2|} \end{pmatrix}$$

Bipartite network and its bi-adjacency Matrix

TODO



Incidence Matrix ^[Die05, New10]

- The **incidence matrix** $\mathbf{B}$ by ^[Die05] of a *simple undirected* graph $G(V, E)$ with N vertices $V = \{v_1, \dots, v_N\}$ and M edges $E = \{e_1, \dots, e_M\}$ over the 2-element field $F_2 = \{0, 1\}$ is defined as the $N \times M$ matrix with elements B_{ij} such that

$$B_{ij} = \begin{cases} 1 & \text{if } v_i \in e_j \\ 0 & \text{otherwise} \end{cases}$$

- The **edge incidence matrix** by Newman ^[New10] of a *simple undirected* graph $G(V, E)$ with N vertices and M edges is an $M \times N$ matrix $\mathbf{B}$ with elements B_{ij}

$$B_{ij} = \begin{cases} +1 & \text{if end 1 of edge } i \text{ is attached to vertex } j, \\ -1 & \text{if end 2 of edge } i \text{ is attached to vertex } j, \\ 0 & \text{otherwise} \end{cases}$$

- Each edge has two arbitrarily designated ends, *end 1* and *end 2*.
- Each row of the matrix has exactly one $+1$ and one -1 element.



Projection ^[New10, BJP17]

- A possible way how to analyze bipartite graphs using simple graph methods.
- Significant information on the given network might be lost.

Definition 1 (Based on Definition 3 [BJP17], p.3)

Let $G(V_1, V_2, E)$ be a bipartite graph. The **one-mode projection** of the bipartite graph G for the vertex V_i with respect to the vertex set V_j , $i, j \in \{1, 2\}$, $i \neq j$ is the unipartite (one-mode) network $G'(V_i, E')$ where $V(G') = U$ and $uv \in E(G')$ if $N(u) \cap N(v) \neq \emptyset$.

Projection of a bipartite network - items and groups

TODO



Projection Properties I ^[New10]

- Let $\mathbf{B}$ be a bi-adjacency matrix of $G(V_1, V_2, E)$, then the total number $P_{ij}^{(1)}$ of vertexes $v \in V_2$ to which both $i, j \in V_1$ belong is

$$P_{ij}^{(1)} = \sum_{k=1}^{|V_2|} B_{ik} B_{jk} = \sum_{k=1}^{|V_2|} B_{ik} B_{kj}^T$$

- The product $B_{ik} B_{jk}$ will be 1 if and only if i and j are both linked to the same vertex k from the other vertex set
- Example: relations of items and their groups
- In matrix form

$$\mathbf{P}^{(1)} = \mathbf{B}\mathbf{B}^T$$



Projection Properties II ^[New10]

- $P_{ii}^{(1)}$ is the number of vertexes $j \in V_2$ to which $i \in V_1$ is linked

$$P_{ii}^{(1)} = \sum_{k=1}^{|V_2|} B_{ik}^2 = \sum_{k=1}^{|V_2|} B_{ik}$$

- assuming $B_{ik} \in \{0, 1\}$
- The other one-mode projection onto V_2

$$\mathbf{P}^{(2)} = \mathbf{B}^T \mathbf{B}$$



Undirected Graph - Node Degree ^[New10]

- The **degree** of a vertex in a undirected graph

$$k_i = \sum_{j=1}^N A_{ij}$$

- The number of ends of edges

$$2M = \sum_{i=1}^N k_i$$

- The number of edges

$$M = \frac{1}{2} \sum_{i=1}^N k_i = \frac{1}{2} \sum_{ij} A_{ij}$$



Undirected Graph - Density ^[New10]

- The **mean degree** c of a vertex in a undirected graph

$$c = \frac{1}{N} \sum_{i=1}^N k_i = \frac{2M}{N}$$

- The maximum possible number of edges in a simple graph

$$\binom{N}{2} = \frac{1}{2}N(N-1)$$

- The **connectance** or **density** ρ of a graph is the fraction of edges that are actually present ($0 \leq \rho \leq 1$).

$$\rho = \frac{M}{\binom{N}{2}} = \frac{2M}{N(N-1)} = \frac{c}{N-1}$$



Directed Graph - Vertex Degree ^[New10]

- The **in-degree** k_i^{in} and **out-degree** k_j^{out} of a vertex in a directed graph

$$k_i^{\text{in}} = \sum_{j=1}^N A_{ij}, \quad k_j^{\text{out}} = \sum_{i=1}^N A_{ij}$$

- The number of edges

$$M = \sum_{i=1}^N k_i^{\text{in}} = \sum_{j=1}^N k_j^{\text{out}} = \sum_{ij} A_{ij}$$

- The **mean in-degree** c_{in} and the **mean out-degree** c_{out} of a vertex in a directed graph are equal:

$$c_{\text{in}} = \frac{1}{N} \sum_{i=1}^N k_i^{\text{in}} = \frac{1}{N} \sum_{j=1}^N k_j^{\text{out}} = c_{\text{out}} = c = \frac{M}{N}$$



Paths in Simple Graph ^[New10]

- The element A_{ij} is 1 if there is an edge from i to j , and 0 otherwise in simple graphs.
- The product $A_{ik}A_{kj}$ is 1 if there is a path of length 2 from j to i via k , and 0 otherwise.
- The total number $N_{ij}^{(2)}$ of paths of length two from j to i via any other vertex is

$$N_{ij}^{(2)} = \sum_{k=1}^N A_{ik}A_{kj} = [\mathbf{A}^2]_{ij}$$

- Paths of length three from j to i via l and k in that order

$$N_{ij}^{(3)} = \sum_{k=1}^N A_{ik}A_{kl}A_{lj} = [\mathbf{A}^3]_{ij}$$

- Paths of an arbitrary length r

$$N_{ij}^{(r)} = [\mathbf{A}^r]_{ij}$$



Cycles in Simple Graph ^[New10]

- The number of paths of length r that start and end at the same vertex i is $[\mathbf{A}^r]_{ii}$.
- The total number L_r of cycles (“loops”) of length r anywhere in a network is (the sum over all possible starting vertexes i)

$$L_r = \sum_{i=1}^N [\mathbf{A}^r]_{ii} = \text{Tr} \mathbf{A}^r.$$

- The loop $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$ is considered different from the loop $2 \rightarrow 3 \rightarrow 1 \rightarrow 2$.
- The loops $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$ and $1 \rightarrow 3 \rightarrow 2 \rightarrow 1$ traversed in opposite directions are distinct, too.



Cycles in Simple Graph and Eigenvalues ^[New10]

• Undirected graph

- The adjacency matrix $\mathbf{A}$ is symmetric, i.e. $\mathbf{A} = \mathbf{Q}\mathbf{K}\mathbf{Q}^T$, where $\mathbf{Q}$ is the orthogonal matrix of eigenvectors and $\mathbf{K}$ is the diagonal matrix of eigenvalues κ_i of $\mathbf{A}$.
- $\mathbf{A}^r = (\mathbf{Q}\mathbf{K}\mathbf{Q}^T)^r = \mathbf{Q}\mathbf{K}^r\mathbf{Q}^T$
- $L_r = \text{Tr}\mathbf{A}^r = \text{Tr}(\mathbf{Q}\mathbf{K}^r\mathbf{Q}^T) = \text{Tr}(\mathbf{Q}^T\mathbf{Q}\mathbf{K}^r) = \text{Tr}\mathbf{K}^r = \sum_i \kappa_i^r$

• Directed networks

- Every real matrix can be written in the form $\mathbf{A} = \mathbf{Q}\mathbf{T}\mathbf{Q}^T$, where $\mathbf{Q}$ is an orthogonal matrix and $\mathbf{T}$ is an upper triangular matrix using the *Schur decomposition*.
- Since $\mathbf{T}$ is triangular, its diagonal elements are its eigenvalues.
- The eigenvalues are the same as the eigenvalues of $\mathbf{A}$.

$$\mathbf{A}\mathbf{x} = \mathbf{Q}\mathbf{T}\mathbf{Q}^T\mathbf{x} = \kappa\mathbf{x} \quad \dots \times \mathbf{Q}^T \quad (1)$$

$$\mathbf{T}\mathbf{Q}^T\mathbf{x} = \kappa\mathbf{Q}^T\mathbf{x} \quad (2)$$

$$(3)$$

- $L_r = \text{Tr}\mathbf{A}^r = \text{Tr}(\mathbf{Q}\mathbf{T}^r\mathbf{Q}^T) = \text{Tr}(\mathbf{Q}^T\mathbf{Q}\mathbf{T}^r) = \text{Tr}\mathbf{T}^r = \sum_i \kappa_i^r$



Outline

1 Graph Matrices

- Network Matrices

2 Centrality Measures

- Path Based Centralities
- Spectral Centralities
- Example



Centrality Measures / Ranking ^[BE06, Weh13]

Measuring the importance/prominence of a node within the network

- Degree Centrality (Node Activity)
- Betweenness Centrality (Intermediate Position)
- Closeness Centrality (Distance to other nodes)
- Eigenvector Centrality (Important nodes have important friends)
- Power Centrality (Close to hubs)
- Page Rank

Evaluation of the location actors in the network

- Insight into various roles and groupings in a network
- Connectors, mavens, leaders, bridges, isolates, broker, hubs
- Where are the clusters and who is in them,
- Who is in the core of the network? Who is on the periphery?
- What is a single point of failure?

Degree Centrality [Fre79, BE06, Weh13]

What is the degree of an actor? How active is an actor?

Degree centrality

is a count of the number of edges incident upon a given vertex.

Degree centrality for actor i

$$c_i^d = \sum_j a_{ij} = \mathbf{A}\mathbf{1}$$

- where $\mathbf{A}$ is the adjacency matrix
- $\mathbf{1}$ is a vector of 1 with size N .

Normalized degree centrality for actor i

$$c_i'^d = \frac{\sum_j a_{ij}}{N-1} = \frac{\mathbf{A}\mathbf{1}}{N-1}$$

Degree Centrality [Fre79, BE06, Weh13]

What is the degree of an actor? How active is an actor?

Degree centrality

is a count of the number of edges incident upon a given vertex.

Degree centrality for actor i

$$c_i^d = \sum_j a_{ij} = \mathbf{A}\mathbf{1}$$

- where $\mathbf{A}$ is the adjacency matrix
- $\mathbf{1}$ is a vector of 1 with size N .

Normalized degree centrality for actor i

$$c_i'^d = \frac{\sum_j a_{ij}}{N-1} = \frac{\mathbf{A}\mathbf{1}}{N-1}$$

Degree Centrality [Fre79, BE06, Weh13]

What is the degree of an actor? How active is an actor?

Degree centrality

is a count of the number of edges incident upon a given vertex.

Degree centrality for actor i

$$c_i^d = \sum_j a_{ij} = \mathbf{A}\mathbf{1}$$

- where $\mathbf{A}$ is the adjacency matrix
- $\mathbf{1}$ is a vector of 1 with size N .

Normalized degree centrality for actor i

$$c_i'^d = \frac{\sum_j a_{ij}}{N-1} = \frac{\mathbf{A}\mathbf{1}}{N-1}$$

Degree Centrality [Fre79, BE06, Weh13]

What is the degree of an actor? How active is an actor?

Degree centrality

is a count of the number of edges incident upon a given vertex.

Degree centrality for actor i

$$c_i^d = \sum_j a_{ij} = \mathbf{A}\mathbf{1}$$

- where $\mathbf{A}$ is the adjacency matrix
- $\mathbf{1}$ is a vector of 1 with size N .

Normalized degree centrality for actor i

$$c_i'^d = \frac{\sum_j a_{ij}}{N-1} = \frac{\mathbf{A}\mathbf{1}}{N-1}$$

Examples of degree centrality ^[Weh13]

Examples for degree centrality c_i and normalized degree centrality c'_i :

Star

$$\begin{array}{ll} c_1^d = 4 & c_1'^d = 1 \\ c_2^d = 1 & c_2'^d = 0.25 \\ c_3^d = 1 & c_3'^d = 0.25 \\ c_4^d = 1 & c_4'^d = 0.25 \\ c_5^d = 1 & c_5'^d = 0.25 \end{array}$$

Line

$$\begin{array}{ll} c_1^d = 2 & c_1'^d = 0.5 \\ c_2^d = 2 & c_2'^d = 0.5 \\ c_3^d = 2 & c_3'^d = 0.5 \\ c_4^d = 1 & c_4'^d = 0.25 \\ c_5^d = 1 & c_5'^d = 0.25 \end{array}$$

Circle

$$\begin{array}{ll} c_1^d = 2 & c_1'^d = 0.5 \\ c_2^d = 2 & c_2'^d = 0.5 \\ c_3^d = 2 & c_3'^d = 0.5 \\ c_4^d = 2 & c_4'^d = 0.5 \\ c_5^d = 2 & c_5'^d = 0.5 \end{array}$$

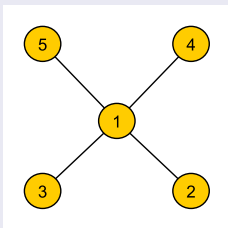
(all actors identical)



Examples of degree centrality ^[Weh13]

Examples for degree centrality c_i and normalized degree centrality c'_i :

Star



$c_1^d = 4$	$c'_1^d = 1$
$c_2^d = 1$	$c'_2^d = 0.25$
$c_3^d = 1$	$c'_3^d = 0.25$
$c_4^d = 1$	$c'_4^d = 0.25$
$c_5^d = 1$	$c'_5^d = 0.25$

Line

$c_1^d = 2$	$c'_1^d = 0.5$
$c_2^d = 2$	$c'_2^d = 0.5$
$c_3^d = 2$	$c'_3^d = 0.5$
$c_4^d = 1$	$c'_4^d = 0.25$
$c_5^d = 1$	$c'_5^d = 0.25$

Circle

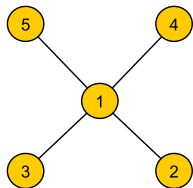
$c_1^d = 2$	$c'_1^d = 0.5$
$c_2^d = 2$	$c'_2^d = 0.5$
$c_3^d = 2$	$c'_3^d = 0.5$
$c_4^d = 2$	$c'_4^d = 0.5$
$c_5^d = 2$	$c'_5^d = 0.5$

(all actors identical)

Examples of degree centrality ^[Weh13]

Examples for degree centrality c_i and normalized degree centrality c'_i :

Star



$c_1^d = 4$	$c'_1{}^d = 1$
$c_2^d = 1$	$c'_2{}^d = 0.25$
$c_3^d = 1$	$c'_3{}^d = 0.25$
$c_4^d = 1$	$c'_4{}^d = 0.25$
$c_5^d = 1$	$c'_5{}^d = 0.25$

Line



$c_1^d = 2$	$c'_1{}^d = 0.5$
$c_2^d = 2$	$c'_2{}^d = 0.5$
$c_3^d = 2$	$c'_3{}^d = 0.5$
$c_4^d = 1$	$c'_4{}^d = 0.25$
$c_5^d = 1$	$c'_5{}^d = 0.25$

Circle

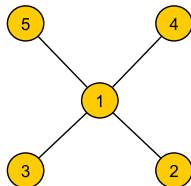
$c_1^d = 2$	$c'_1{}^d = 0.5$
$c_2^d = 2$	$c'_2{}^d = 0.5$
$c_3^d = 2$	$c'_3{}^d = 0.5$
$c_4^d = 2$	$c'_4{}^d = 0.5$
$c_5^d = 2$	$c'_5{}^d = 0.5$

(all actors identical)

Examples of degree centrality [Weh13]

Examples for degree centrality c_i and normalized degree centrality c'_i :

Star



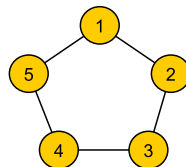
$c_1^d = 4$	$c'_1{}^d = 1$
$c_2^d = 1$	$c'_2{}^d = 0.25$
$c_3^d = 1$	$c'_3{}^d = 0.25$
$c_4^d = 1$	$c'_4{}^d = 0.25$
$c_5^d = 1$	$c'_5{}^d = 0.25$

Line



$c_1^d = 2$	$c'_1{}^d = 0.5$
$c_2^d = 2$	$c'_2{}^d = 0.5$
$c_3^d = 2$	$c'_3{}^d = 0.5$
$c_4^d = 1$	$c'_4{}^d = 0.25$
$c_5^d = 1$	$c'_5{}^d = 0.25$

Circle



$c_1^d = 2$	$c'_1{}^d = 0.5$
$c_2^d = 2$	$c'_2{}^d = 0.5$
$c_3^d = 2$	$c'_3{}^d = 0.5$
$c_4^d = 2$	$c'_4{}^d = 0.5$
$c_5^d = 2$	$c'_5{}^d = 0.5$

(all actors identical)

Closeness centrality ^[Fre79, Dod09]

- **Idea:** Nodes are more central if they can reach other nodes 'easily.'
- Measures average shortest path from a node to all other nodes.
- **Closeness Centrality** for node i as

$$c_i^c = \frac{N - 1}{\sum_{j, j \neq i} (\text{distance from } i \text{ to } j)}$$

- Range is 0 (no friends) to 1 (a single hub).

Meaning

- Unclear what the exact values of this measure tells us because of its ad-hocness.
- General problem with simple centrality measures: what do they exactly mean?
- Perhaps, at least, we obtain an ordering of nodes in terms of 'importance.'

Closeness centrality ^[Fre79, Dod09]

- **Idea:** Nodes are more central if they can reach other nodes 'easily.'
- Measures average shortest path from a node to all other nodes.
- **Closeness Centrality** for node i as

$$c_i^c = \frac{N - 1}{\sum_{j, j \neq i} (\text{distance from } i \text{ to } j)}$$

- Range is 0 (no friends) to 1 (a single hub).

Meaning

- Unclear what the exact values of this measure tells us because of its ad-hocness.
- General problem with simple centrality measures: what do they exactly mean?
- Perhaps, at least, we obtain an ordering of nodes in terms of 'importance.'

Betweenness centrality ^[Dod09]

- **Betweenness centrality** is based on shortest paths in a network.
- **Idea:** If the quickest way between any two nodes on a network disproportionately involves certain nodes, then they are ‘important’ in terms of global cohesion.
- For each node i , **count, over all pairs of nodes x and y , how many shortest paths pass through i .**
- Call frequency of shortest paths passing through node i the **betweenness** of i , B_i .
- Note: Exclude shortest paths between i and other nodes.
- Note: works for weighted and unweighted networks.
- Role played by shortest paths justified by small-world phenomenon (Milgram’s experiment).



Betweenness Centrality - Complexity ^[Dod09]

- Consider a network with N nodes and M edges (possibly weighted).
- *Computational goal*: Find $\binom{N}{2}$ **shortest paths** between all pairs of nodes.
- Traditionally **Floyd-Warshall** algorithm used.
- Computation time grows as $O(N^3)$.
- See also:
 - 1 **Dijkstra's algorithm** for finding the shortest path between two specific nodes, and
 - 2 **Johnson's algorithm** which outperforms Floyd-Warshall for sparse networks:

$$O(MN + N^2 \log N)$$

- Newman (2001) and Brandes (2001) independently derived much faster algorithms.
- Computation times grow as:
 - 1 $O(MN)$ for unweighted graphs, and
 - 2 $O(MN + N^2 \log N)$ for weighted graphs.



Shortest path between node i and all others ^[Dod09]

- Consider unweighted networks.
- Use **breadth-first search**:
 - 1 Start at node i , giving it a distance $d = 0$ from itself.
 - 2 Create a list of all of i 's neighbors and label them being at a distance $d = 1$.
 - 3 Go through list of most recently visited nodes and find all of their neighbors.
 - 4 Exclude any nodes already assigned a distance.
 - 5 Increment distance d by 1.
 - 6 Label newly reached nodes as being at distance d .
 - 7 Repeat steps 3 through 6 until all nodes are visited.
- Record which nodes link to which nodes moving out from i (former are 'predecessors' with respect to i 's shortest path structure).
- Runs in $O(M)$ time and gives N shortest paths.
- Find all shortest paths in $O(MN)$ time
- Much, much better than naive estimate of $O(MN^2)$.



Newman's Betweenness algorithm ^[New01, Dod09]

- 1 Set all nodes to have a value $c_{ij} = 0, j = 1, \dots, N$ (c for count).
- 2 Select one node i .
- 3 Find shortest paths to all other $N - 1$ nodes using breadth-first search.
- 4 Record $\#$ equal shortest paths reaching each node.
- 5 Move through nodes according to their distance from i , starting with the furthest.
- 6 Travel **back towards i from each starting node j** , along shortest path(s), adding 1 to every value of c_{ik} at each node k along the way.
- 7 Whenever more than one possibility exists, a portion **according to total number of short paths** coming through predecessors.
- 8 Exclude starting node j and i from increment.
- 9 Repeat steps 2-8 for every node i and obtain **betweenness** as

$$B_j = \sum_{i=1}^N c_{ij}$$



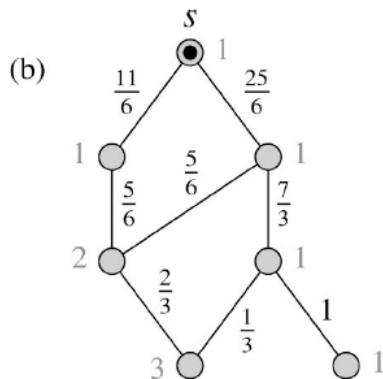
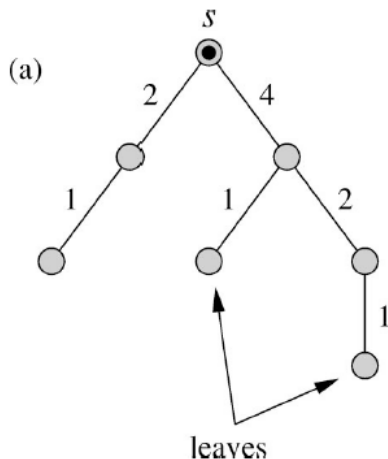
Newman's Betweenness - notes [New01, Dod09]

- For a pure tree network, c_{ij} is the number of nodes beyond j from i 's vantage point.
- For **edge betweenness**, use exact same algorithm but now
 - 1 j indexes edges, and
 - 2 we add one to each edge as we traverse it.
- For both algorithms, computation time grows as $O(MN)$ and space for $O(N + M)$ integers (N nodes, M arcs).
- Both bounds infeasible for large networks, where typically $N \approx 10^9$ and $M \approx 10^{11}$.
- For sparse networks with relatively small average degree, we have a fairly digestible time growth of $O(N^2)$.



Newman's Betweenness - examples

[New01, Dod09]



Outline

1 Graph Matrices

- Network Matrices

2 Centrality Measures

- Path Based Centralities
- Spectral Centralities
- Example



Important nodes have important friends [Dod09, New10]

- Define x_i as the "importance" of node i .
- *Idea*: x_i depends (somehow) on x_j if j is a neighbor of i .
- *Recursive*: importance is transmitted through a network.
- Simplest possibility is a linear combination:

$$x_i \propto \sum_j A_{ij} x_j$$

- Assume further that constant of proportionality, c , is independent of i .
- Above gives $\tilde{\mathbf{x}} = c\mathbf{A}\tilde{\mathbf{x}}$ or $\boxed{\mathbf{A}\tilde{\mathbf{x}} = c^{-1}\tilde{\mathbf{x}} = \lambda\tilde{\mathbf{x}}}$.
- Eigenvalue equation based on adjacency matrix:
 - The greatest eigenvalue and its related eigenvector fulfills only the additional requirement that all the entries in the eigenvector be positive (Perron-Frobenius theorem).
- **Eigenvalue centrality** of the vertex v in the network
 ... The v^{th} component of the related eigenvector



Eigenvalue Centrality - Iterative Approach ^[New10]

- An initial guess about the centrality x_i of each vertex i .
 - e.g. $x_i = 1$ for all i
- One step to calculate a better estimate x'_i

$$x'_i = \sum_j A_{ij} x_j \quad \text{i.e. } \mathbf{x}' = \mathbf{A}\mathbf{x}$$

- Repeat t times: $\mathbf{x}(t) = \mathbf{A}^t \mathbf{x}(0)$
- Express $\mathbf{x}(0)$ as a linear combination of the eigenvectors $\mathbf{v}_i$ of $\mathbf{A}$:
 $\mathbf{x}(0) = \sum_i c_i \mathbf{v}_i$.

$$\mathbf{x}(t) = \mathbf{A}^t \sum_i c_i \mathbf{v}_i = \sum_i c_i \mathbf{A}^t \mathbf{v}_i = \sum_i c_i \kappa_i^t \mathbf{v}_i = \kappa_1^t \sum_i c_i \left[\frac{\kappa_i}{\kappa_1} \right]^t \mathbf{v}_i$$

- κ_i are the eigenvalues of $\mathbf{A}$, κ_1 is the largest of them.
- Since $\kappa_i/\kappa_1 < 1$ for all $i \neq 1$, all terms in the sum other than the first decay exponentially as t becomes large: $\mathbf{x}(t) \rightarrow c_1 \kappa_1 \mathbf{v}_1$ as $t \rightarrow \infty$.



Eigenvalue Centrality - Properties ^[New10]

- **Eigenvalue centrality** by Bonacich in 1987 ^[Bon87]

$$\mathbf{Ax} = \kappa_1 \mathbf{x} \qquad x_i = \kappa_1^{-1} \sum_j A_{ij} x_j$$

- The centrality x_i of vertex i is proportional to the sum of the centralities of i 's neighbors:
 - a vertex has many neighbors,
 - a vertex has important neighbors.
- The eigenvector centralities of all vertices are non-negative.
 - If $x_i(0) \geq 0$ and $A_{ij} \geq 0$ then $x_i(t) \geq 0$.
- Eigenvector centrality works well for *undirected* networks.
- Issues with *directed* networks
 - Asymmetric adjacency matrix has two sets of eigenvectors, left and right, i.e hence two leading eigenvectors.
 - In most cases the right eigenvector should be used
 - to prefer the case in which centralities are driven by vertices pointing to a given vertex (and not to which vertices the given vertex points to)
 - **Zero x_i** are propagated as zero $\implies$ **strong components** taken only.



Katz Centrality ^[Kat53]

- To resolve the issue with zero eigenvalue centralities x_i

Katz Centrality

- Proposed by Katz in 1953

$$\mathbf{C}_{\text{Katz}} = \alpha \mathbf{A} + \alpha^2 \mathbf{A}^2 + \dots + \alpha^k \mathbf{A}^k + \dots \quad (4)$$

$$\mathbf{C}_{\text{Katz}}(i) = \sum_{k=1}^{\infty} \sum_{j=1}^N \alpha^k [\mathbf{A}^k]_{ij} \quad (5)$$

- $\mathbf{C}_{\text{Katz}}(i)$ denotes Katz centrality of a node i .
- The attenuation factor α ... discounted paths (walks)
- A link in the distance k is attenuated by α^k .
- If $\alpha < 1/|\kappa_1|$, where κ_1 is the largest eigenvalue of $\mathbf{A}$, then

$$\vec{\mathbf{c}}_{\text{Katz}} = ((\mathbf{I} - \alpha \mathbf{A}^T)^{-1} - \mathbf{I}) \mathbf{1}$$

Alpha Centrality ^[BL01, New10]

- Proposed by Bonacich in 2001 ^[BL01]
- A generalization of Katz centrality

$$x_i = \alpha \sum_j A_{ij} x_j + \beta \qquad \mathbf{x} = \alpha \mathbf{A} \mathbf{x} + \beta \mathbf{1}$$

where α and β are positive constants.

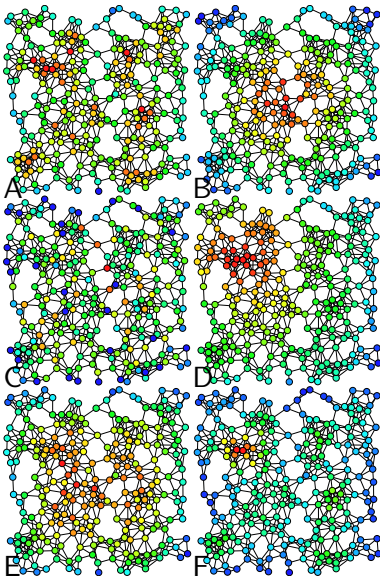
- Each vertex has a non-zero positive centrality because of small $\beta > 0$
- Rearranging for $\mathbf{x}$

$$\mathbf{x} = \beta (\mathbf{I} - \alpha \mathbf{A})^{-1} \cdot \mathbf{1} = (\mathbf{I} - \alpha \mathbf{A})^{-1} \cdot \mathbf{1}$$

- using $\beta = 1$ to care about relative values of centralities only.
- $\mathbf{C}_{\text{Alpha}} = \alpha^0 \mathbf{A}^0 + \mathbf{C}_{\text{Katz}} = \mathbf{I} + \mathbf{C}_{\text{Katz}}$
- Choice of a value of α
 - If $\alpha \rightarrow 0$, then all $x_i \rightarrow \beta = 1$
 - If $\alpha \rightarrow 1/\kappa_1$, then a divergence ... $\det(\mathbf{A} - \alpha^{-1} \mathbf{I}) = 0$



Centrality Measures - Importance of Nodes ^[Roc12]



- Low → middle → high values
- **A** Degree centrality,
 - Node Activity
- **B** Closeness centrality,
 - Distance to other nodes
- **C** Betweenness centrality,
 - Intermediate Position
- **D** Eigenvector centrality,
 - Important nodes have important friends
- **E** Katz centrality,
 - The relative influence of a node within a network
- **F** Alpha centrality
 - Important nodes have important friends for asymmetric relations

PageRank [BP98, BP12, New10]

- In some case, a high-centrality vertex should not distribute its centrality to other vertexes fully,
 - e.g. *Yahoo!* referencing a personal page.
- The centrality of a given vertex is distributed to its neighbors as an amount proportional to its centrality divided by its out-degree.

$$x_i = \alpha \sum_j A_{ij} \frac{x_j}{k_j^{\text{out}}} + \beta \qquad \mathbf{x} = \alpha \mathbf{A} \mathbf{D}^{-1} \mathbf{x} + \beta \mathbf{1}$$

- If $k_j^{\text{out}} = 0$, then $A_{ij} = 0$ for all i .
- In such cases, we set artificially $k_j^{\text{out}} = 1$ to avoid the problem with the term when zero is divided by zero. The result is a zero centrality contribution.
- $\mathbf{D}$ is the diagonal matrix with elements $D_{ii} = \max(k_i^{\text{out}}, 1)$
- By rearranging and setting $\beta = 1$, and $\alpha < 1/|\kappa_1|$, $\kappa_1 = \lambda_{\max}(\mathbf{A})$

$$\mathbf{x} = \beta (\mathbf{I} - \alpha \mathbf{A} \mathbf{D}^{-1})^{-1} \cdot \mathbf{1} = \mathbf{D} (\mathbf{D} - \alpha \mathbf{A})^{-1} \cdot \mathbf{1}$$



Outline

1 Graph Matrices

- Network Matrices

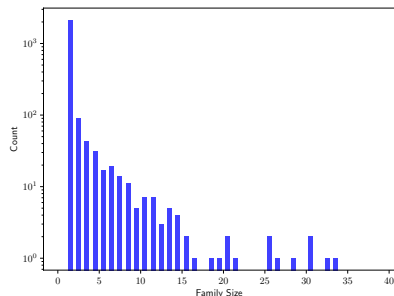
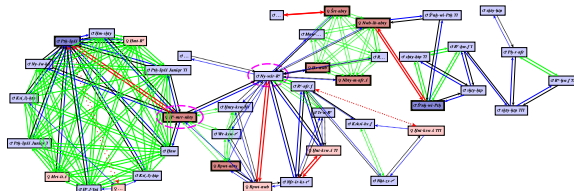
2 Centrality Measures

- Path Based Centralities
- Spectral Centralities
- Example



Egypt Data - Family Formation ^[DM15]

$Ny-wsr-R^c$	0.647
$H^c-mrr-nbty$	0.424
$Nwb-ib-nbty$	0.351
$\acute{S}^c nh-wi-Pth$	0.290
$R^c-hw.f'I$	0.180
$R^c-nfr.f$	0.139
$zhty-htp'III$	0.139
$Pth-\acute{s}p\acute{s}\acute{s}$	0.082
$Pth-r-nfr III$	0.048
$\acute{S}rt-nbty I$	0.048



People with
the top 10 highest betweenness

Extended family size distribution

Summary

- Linear algebra remainder
- Network matrices
- Centrality Measures
 - Path based centralities
 - Spectral centralities



Competencies

- Define adjacency matrix, cocitation matrix, and bibliographic coupling
- Define bi-adjacency matrix, incidence matrix, edge incidence matrix
- Define one-mode projection and its relation to bi-adjacency matrix.
- Show how to compute degree of vertex, the number of edges, the mean degree, and graph density based on the adjacency matrix for undirected and directed graphs.
- Show how to compute number of paths and cycles based on the adjacency matrix.
- Define degree centrality.
- Define closeness centrality.
- Define betweenness centrality.
- Describe an algorithm for betweenness centrality computation.
- Define eigenvalue centrality.
- Define Katz centrality.
- Define PageRank index.

References I

- [BE06] [Stephen P. Borgatti](#) and [Martin G. Everett](#). A graph-theoretic perspective on centrality. *Social Networks*, 28(4):466–484, 2006.
- [BJP17] [S. Banerjee](#), [M. Jenamani](#), and [D. K. Pratihari](#). Properties of a Projected Network of a Bipartite Network. *ArXiv e-prints*, July 2017.
- [BL01] [Phillip Bonacich](#) and [Paulette Lloyd](#). Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*, 23(3):191 – 201, 2001.
- [BM08] [J.A. Bondy](#) and [U.S.R. Murty](#). *Graph Theory*. Springer, 2008.
- [Bon87] [Phillip Bonacich](#). Power and centrality: A family of measures. *American Journal of Sociology*, 92(5):1170–1182, 1987.
- [BP98] [Sergey Brin](#) and [Lawrence Page](#). The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.*, 30(1-7):107–117, April 1998.
- [BP12] [Sergey Brin](#) and [Lawrence Page](#). Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer Networks*, 56(18):3825 – 3833, 2012. The WEB we live in.
- [Die05] [Reinhard Diestel](#). *Graph Theory*. Springer, 2005.
- [DM15] [Veronika Dulíková](#) and [Radek Mařík](#). Social network analysis in the Old Kingdom society: the nepotism case. In [M. Bárta](#), [F. Coppens](#), and [J. Krejčí](#), editors, *Abusir and Saqqara 2015, Czech Institute of Egyptology, Charles University, Prague, CZ*, pages 63–83, 2015.
- [Dod09] [Peter Dodds](#). Measures of centrality, complex networks, course 303a, spring, 2009, lecture notes. <http://www.uvm.edu/~pdodds/teaching/courses/2009-01UVM-303/content/lectures.html>, 2009.
- [EK10] [David Easley](#) and [Jon Kleinberg](#). *Networks, Crowds, and Markets. Reasoning About a Highly Connected World*. Cambridge University Press, July 2010.
- [Fre79] [Linton C. Freeman](#). Centrality in social networks: I. conceptual clarification. *Social Networks*, (1):215–239, 1979.



References II

- [Kat53] Leo Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, Mar 1953.
- [New01] M. E. J. Newman. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E*, 64:016132, 2001.
- [New10] M. Newman. *Networks: an introduction*. Oxford University Press, Inc., 2010.
- [Roc12] Claudio Rocchini. Centrality. <http://en.wikipedia.org/wiki/File:Centrality.svg>, November 2012.
- [Weh13] Stefan Wehrli. Social network analysis, lecture notes, December 2013.

