ECON5152 — Network Analysis

Week 3: Centrality and Importance

Santiago Montoya-Blandón

January 29, 2026

Outline

- Measuring importance in a network
- Centrality measures
  - Degree
  - Distance-based (closeness and betweenness)
  - Eigenvector
  - Katz-Bonacich
  - PageRank
- Interpretation and comparison
- Applications
  - Sectorial importance
  - Metro systems
  - Key players
  - Global corporate control

Measuring importance in a network

- Graphs have two sets of information: nodes $\mathcal{N}$ and edges $\mathcal{E}$
- How to use edges to understand importance of nodes?
- Intuitively: who matters most in a network?
  - influential individuals (e.g., celebrities, influencers or billionaires)
  - key firms or sectors
  - critical infrastructure (e.g., bridges or energy plans)
  - criminal bosses
  - disease super-spreaders

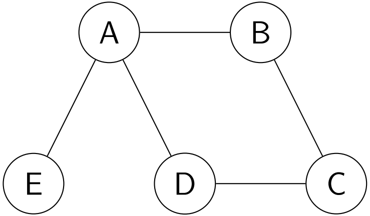
Measuring importance in a network

- Graphs have two sets of information: nodes $\mathcal{N}$ and edges $\mathcal{E}$
- How to use edges to understand importance of nodes?
- Intuitively: who matters most in a network?
  - influential individuals (e.g., celebrities, influencers or billionaires)
  - key firms or sectors
  - critical infrastructure (e.g., bridges or energy plans)
  - criminal bosses
  - disease super-spreaders
- How to define importance?
  - Many connections?
  - Connections to distant parts of the network?
  - Connected to other important nodes?
- Different notions of importance lead to different *centrality measures*

Centrality measures

- Centrality measures capture distinct ideas of importance
- Some are simple; e.g., number of friends
- Others can arise from complex processes, e.g., internet surfing
- Centrality measures formalize different economic and social mechanisms
- Today, focus on calculation and interpretation
- Introduce standard measures, plus application-specific ones

Running example network



$$\mathbf{A} = \begin{array}{c c} & \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} & \begin{pmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix} \end{array}$$

Example network used throughout this lecture.

Degree centrality

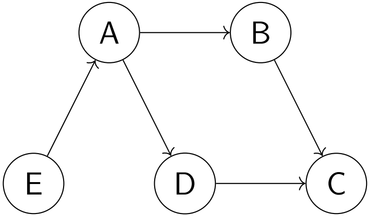
Definition

Degree centrality d_i of node $i \in \mathcal{N}$ equals its (weighted) degree.

- Minimum degree centrality is 0, max is $n - 1$
- Binary vs weighted:
  - Binary: degree measures number of friends
  - Weighted: total connection intensity to neighbours
- Undirected vs directed:
  - Undirected: single definition of degree
  - Directed: two definitions based on either incoming (in-degree) or outgoing (out-degree)

Degree centrality: Running example

Node	Degree	In-degree	Out-degree
A	3	1	2
B	2	1	1
C	2	2	0
D	2	1	1
E	1	0	1



Directed version

- Node A has the highest number of friends
- In- and out-degree capture directed importance
- Node C has the most incoming links
- Node A has the most outgoing links

Closeness centrality

- Distance (length of shortest path) between nodes $i, j \in \mathcal{N}$ denoted $d(i,j)$
- Total distance from i: $T_i := \sum_{j \in \mathcal{N}} d(i,j)$ where $d(i,i) = 0$

Definition

Closeness centrality c_i of node $i \in \mathcal{N}$ is the inverse of the total distance to i:

$$c_i := \frac{1}{T_i} = \left[\sum_{j \in \mathcal{N}} d(i,j)\right]^{-1}$$

- Widely used average closeness $\bar{c}_i := c_i \cdot (n-1)$
- Central nodes are those with lower average distances to others

Betweenness centrality

For nodes $i, j, k \in \mathcal{N}$:

- σ_{jk} counts the *number* of shortest paths between j and k
- $\sigma_{jk}(i)$ counts only the number of those shortest paths *passing through* i

Definition

Betweenness centrality b_i of node $i \in \mathcal{N}$ measures how often i lies on shortest paths between other nodes:

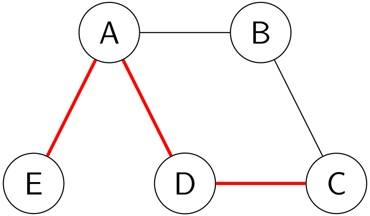
$$b_i := \sum_{\substack{j,k \in \mathcal{N} \\ i \neq j \neq k}} \frac{\sigma_{jk}(i)}{\sigma_{jk}} \text{ or scaled as } \bar{b}_i := \frac{b_i}{n_{\mathrm{p}}},\ n_{\mathrm{p}} := \begin{cases} (n-1)(n-2) & \text{if directed} \\ \frac{(n-1)(n-2)}{2} & \text{if undirected} \end{cases}$$

- Central nodes are intermediaries of many connections
- Captures brokerage in finance or gatekeeping in social networks

Digression: Number of shortest paths

- To obtain betweenness centrality, we need to count the number of shortest paths between many nodes
- Doing this for large networks is not an easy task
- We can still easily count the number of *walks* (allowing repetitions and cycles)
- Use powers of adjacency matrix $\mathbf{A}$
- $\mathbf{A}^2 = \mathbf{A} \cdot \mathbf{A}$ counts all walks of length 2
- In general, $\mathbf{A}^k$ counts all walks of length k

Running example



Node A is close to several nodes and lies on many shortest paths.

Running example: Details

- We can calculate all pairwise shortest-path distance between nodes

$$\mathbf{D} = \begin{array}{c c} & \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} & \begin{pmatrix} 0 & 1 & 2 & 1 & 1 \\ 1 & 0 & 1 & 2 & 2 \\ 2 & 1 & 0 & 1 & 3 \\ 1 & 2 & 1 & 0 & 2 \\ 1 & 2 & 3 & 2 & 0 \end{pmatrix} \end{array}$$

- Total distances: $T_A = 5,\ T_B = 6,\ T_C = 7,\ T_D = 6,\ T_E = 8$
- Closeness centralities: $c_A = 1/5,\ c_B = 1/6,\ c_C = 1/7,\ c_D = 1/6,\ c_E = 1/8$
- Average closeness: $c_A = 4/5,\ c_B = 2/3,\ c_C = 4/7,\ c_D = 2/3,\ c_E = 1/2$

Running example: Details

- Walks of order 2

$$\mathbf{A}^2 = \mathbf{A}\cdot\mathbf{A} = \begin{array}{c c} & \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} & \begin{pmatrix} 3 & 0 & 2 & 0 & 0 \\ 0 & 2 & 0 & 2 & 1 \\ 2 & 0 & 2 & 0 & 0 \\ 0 & 2 & 0 & 2 & 1 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix} \end{array}$$

- Number of shortest paths (by hand)

$$\mathbf{S} = \begin{array}{c c} & \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} & \begin{pmatrix} 0 & 1 & 2 & 1 & 1 \\ 1 & 0 & 1 & 2 & 1 \\ 2 & 1 & 0 & 1 & 2 \\ 1 & 2 & 1 & 0 & 1 \\ 1 & 1 & 2 & 1 & 0 \end{pmatrix} \end{array}$$

Running example: Details

- Walks of order 2

$$\mathbf{A}^2 = \mathbf{A}\cdot\mathbf{A} = \begin{array}{c c} & \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} & \begin{pmatrix} 3 & 0 & 2 & 0 & 0 \\ 0 & 2 & 0 & 2 & 1 \\ 2 & 0 & 2 & 0 & 0 \\ 0 & 2 & 0 & 2 & 1 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix} \end{array}$$

- Number of shortest paths (by hand)

$$\mathbf{S} = \begin{array}{c c} & \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} & \begin{pmatrix} 0 & 1 & 2 & 1 & 1 \\ 1 & 0 & 1 & 2 & 1 \\ 2 & 1 & 0 & 1 & 2 \\ 1 & 2 & 1 & 0 & 1 \\ 1 & 1 & 2 & 1 & 0 \end{pmatrix} \end{array}$$

- Shortest paths not involving A: $\{B, C\} = 1$, $\{B, D\} = 2$, $\{B, E\} = 1$, $\{C, D\} = 1$, $\{C, E\} = 2$, $\{D, E\} = 1$
- Those passing through A: $\{B, C\} = 0$, $\{B, D\} = 1$, $\{B, E\} = 1$, $\{C, D\} = 0$, $\{C, E\} = 2$, $\{D, E\} = 1$
- Finally:

$$b_A = \frac{1}{2} + \frac{1}{1} + \frac{2}{2} + \frac{1}{1} = 3.5$$

- Others: $b_B = 1$, $b_C = 1/2$, $b_D = 1$, $b_E = 0$

Eigenvector centrality

Definition

The eigenvector centrality e_i of node $i \in \mathcal{N}$ is proportional to the centrality of its neighbours $\mathcal{N}_i := \{j \in \mathcal{N} \mid i \text{ is connected to } j\} = \{j \in \mathcal{N} \mid A_{i,j} = 1\}$:

$$e_i := \frac{1}{\lambda} \sum_{j \in \mathcal{N}_i} e_j = \frac{1}{\lambda} \sum_{j=1}^{n} A_{i,j} \cdot e_j$$

- Intuitively: a node is central if it is connected to other central nodes

Eigenvector centrality

Definition

The eigenvector centrality e_i of node $i \in \mathcal{N}$ is proportional to the centrality of its neighbours $\mathcal{N}_i := \{j \in \mathcal{N} \mid i \text{ is connected to } j\} = \{j \in \mathcal{N} \mid A_{i,j} = 1\}$:

$$e_i := \frac{1}{\lambda} \sum_{j \in \mathcal{N}_i} e_j = \frac{1}{\lambda} \sum_{j=1}^{n} A_{i,j} \cdot e_j$$

- Intuitively: a node is central if it is connected to other central nodes
- Write in matrix notation: $\lambda \cdot e_i = \sum_{j=1}^{n} A_{i,j} \cdot e_j \implies \mathbf{Ae} = \lambda \mathbf{e} \implies \mathbf{e}$
- In linear algebra jargon: centrality vector $\mathbf{e} = (e_1, \ldots, e_n)$ is the eigenvector of $\mathbf{A}$ with (largest) eigenvalue λ

Eigenvector centrality

Interpretation

At each step, nodes pass a "packet of influence" of size $1/\lambda$ to their connections. Eigenvector centrality is the long-run average of influence for each node.

Pros

- Intuitive network-based measure
- Uses degree: more friends $\implies$ more eigencentrality
- Additionally uses *quality* of connections (more links to popular nodes)
- λ controls weight of influences

Cons

- Sensitive to network feedback and cycles
- All nodes in smaller disconnected components get no centrality
- Any node connected to a popular node is deemed popular

Running example: Eigenvector centrality

- Largest eigenvalue of $\mathbf{A}$ is $\lambda \approx 2.1358$ and all nodes start with influence 1
- First step for node A:

$$e_A = \frac{1}{\lambda}(e_B + e_D + e_E) = \frac{3}{\lambda} \approx 1.4047$$

- First step for others: $e_B = e_C = e_D \approx 0.9364$ and $e_E \approx 0.4682$
- Long-run average (number of steps grows to infinity): $\mathbf{e} = (\lambda \cdot \mathbf{I}_n - \mathbf{A})^{-1}$
- Resulting eigenvector centralities $e_A = 1$, $e_B \approx 0.8338$, $e_C \approx 0.7808$, $e_D \approx 0.8338$, and $e_E \approx 0.4682$

Katz-Bonacich centrality

Definition

The Katz-Bonacich centrality k_i of node $i \in \mathcal{N}$ is proportional to the centrality of its neighbours $\mathcal{N}_i$ plus a constant:

$$k_i := \alpha + \beta \sum_{j=1}^{n} A_{i,j} \cdot k_j$$

- Similar intuition with less caveats
- In matrix notation: $\mathbf{k} = \alpha \cdot (\mathbf{I}_n - \beta \cdot \mathbf{A})^{-1} \mathbf{1}_n$
- $\mathbf{I}_n$ is an $n \times n$ identity matrix and $\mathbf{1}_n$ is an n-dimensional vector of ones

Katz-Bonacich centrality

Interpretation

At each step, nodes receive a baseline influence α and pass β to their connections. Katz-Bonacich centrality is the long-run average of influence.

Pros

- Well-defined even when eigenvector centrality is not
- Counts all walks and discounts longer paths
- All pros of eigenvector centrality
- Allows for disconnected components

Cons

- Less but still sensitive to network feedback and cycles
- Any node connected to a popular node is deemed popular

PageRank

Definition

PageRank centrality p_i of node $i \in \mathcal{N}$ is a normalized version of Katz centrality for directed networks:

$$p_i = \alpha + \beta \sum_{j=1}^{n} A_{j,i} \cdot \frac{p_j}{\max\{d_j^{\text{out}}, 1\}}$$

- Used by Google to rank webpages by importance
- Note use of $A_{j,i}$: for directed networks, node is assumed more important if nodes *point to it*
- In matrix notation: $\mathbf{p} = (p_1, \ldots, p_n) = \alpha \cdot (\mathbf{I}_n - \beta \cdot \mathbf{D}^{-1}\mathbf{A})^{-1}\mathbf{1}_n$
- $\mathbf{D}$ is diagonal with $D_{i,i} = \max\{d_i^{\text{out}}, 1\}$ for all $i \in \mathcal{N}$

PageRank centrality: random surfer interpretation

- Each website starts with influence $1/n$ (where $n =$ total number of websites)
- An internet surfer clicks on webpages at each step, sending influence β (normalized by number of outgoing links) with each click
- Surfer randomly jumps to another part of the internet with probability α/n
- *PageRank* centrality is the long-run average of webpages' influence

PageRank centrality: random surfer interpretation

- Each website starts with influence $1/n$ (where $n =$ total number of websites)
- An internet surfer clicks on webpages at each step, sending influence β (normalized by number of outgoing links) with each click
- Surfer randomly jumps to another part of the internet with probability α/n
- *PageRank* centrality is the long-run average of webpages' influence
- Network $A_{i,j}$ represents existence of a hyperlink from pages i to j
- When you use Google, PageRank estimates importance across millions of webpages
- Usual to set $\alpha = 1 - \beta$ with $\beta \approx 0.85$
- More complicated surfer behaviour can be included

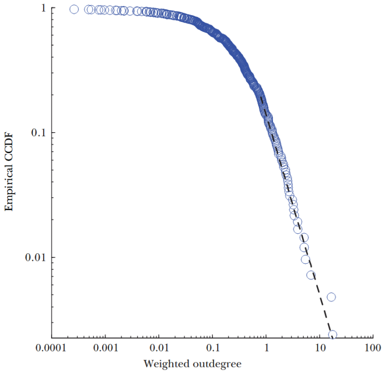
Comparing centrality Measures

Measure	Captures	Typical Use
Degree	Direct activity	Degree distributions and hub detection
Closeness	Speed of access	Distance and diffusion
Betweenness	Control	Brokerage or intermediation
Eigenvector	Influence	Systemic importance
Katz / PageRank	Global reach	High-dimensional ranking

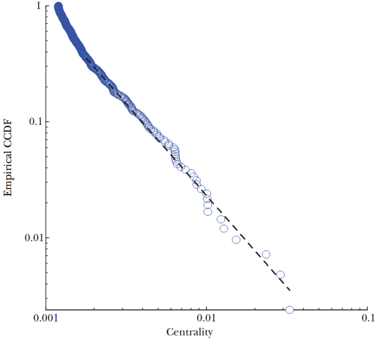
Application: Systemically important sectors

- Example: Carvalho, V. M. (2014). "From micro to macro via production networks". *Journal of Economic Perspectives*, 28(4), 23–48.
- Input–output networks (weighted and directed)
- Rows are buyers, columns are suppliers
- Central sectors defined as those that supply many sectors
- Measured by weighted out-degree
- Systemic downstream importance: eigenvector centrality
- Summarizes propagation potential for economic shocks

Application: Systemically important sectors



Source: Bureau of Economic Analysis, detailed input-output table for 2002. For more details, see Data Appendix available with this paper at http://e-jep.org.
Notes: The x-axis gives the weighted outdegree for each sector, presented on a log scale. The y-axis, also in log scale, gives the probability of finding a sector with weighted outdegree larger than or equal to x, that is the empirical counter-cumulative distribution (CCDF).



Source: Bureau of Economic Analysis, detailed input-output tables for 2002. For more details, see the Data Appendix available with this paper at http://e-jep.org.
Notes: On the x-axis is the Bonacich centrality score of the different sectors in the 2002 input-output data, where I have imposed a baseline centrality measure of $\eta = (1 - 0.5)/417$ and a parameter for weighting downstream sectors of $\lambda = 0.5$. The y-axis gives the probability of finding a sector with a centrality score larger than or equal to x, that is, the empirical counter-cumulative distribution (CCDF).

Application: Transportation networks

- Example: Derrible, S. (2012). "Network centrality of metro systems". *PloS One*, 7(7), e40575.
- Transportation networks defined as links between train stations
- Information across several cities in the world
- Betweenness measures stations that act as heavy transit points (i.e., crossed by many lines)
- Paper identifies a relationship of betweenness with size: larger train systems create more bottlenecks
- Size relationship is consistent across cities and magnitudes

Application: Transportation networks

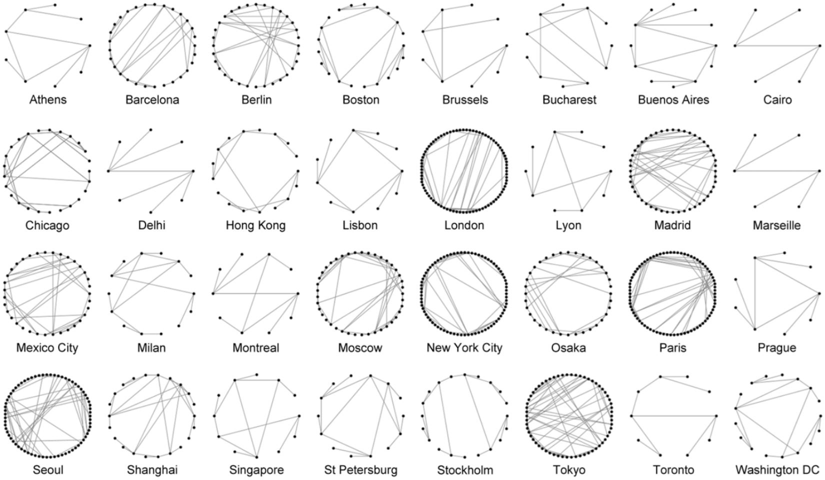
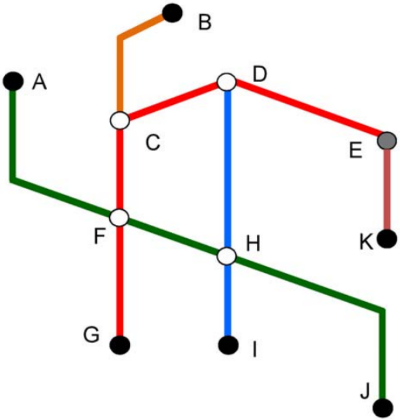


Figure 1. Circle representation of 32 metro networks in the world (using NodeXL [52]).
doi:10.1371/journal.pone.0040575.g001

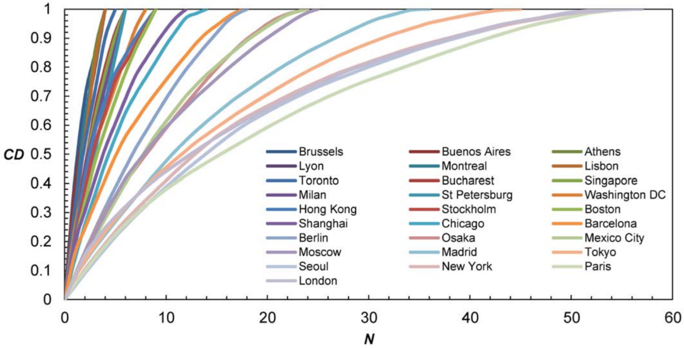
Application: Transportation networks



	A	B	C	D	E	F	G	H	I	J	K
A	0	0	0	0	0	1	0	0	0	0	0
B	0	0	1	0	0	0	0	0	0	0	0
C	0	1	0	1	0	1	0	0	0	0	0
D	0	0	1	0	1	0	0	1	0	0	0
E	0	0	0	1	0	0	0	0	0	0	1
F	1	0	1	0	0	0	1	1	0	0	0
G	0	0	0	0	0	1	0	0	0	0	0
H	0	0	0	1	0	1	0	0	1	1	0
I	0	0	0	0	0	0	0	1	0	0	0
J	0	0	0	0	0	0	0	1	0	0	0
K	0	0	0	0	1	0	0	0	0	0	0

Figure: Train network of Lyon and its adjacency matrix representation

Application: Transportation networks



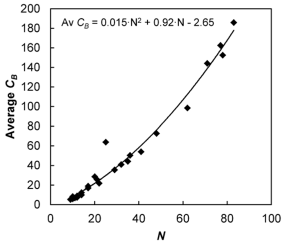


Figure 3. Evolution of average betweenness centrality C_B with network size.

Figure 4. Cumulative distributions (*CD*) of normalized betweenness centrality for 28 metros.

Application: Identifying key players

- Example: Ballester, C., Calvó-Armengol, A., & Zenou, Y. (2006). "Who's who in networks. Wanted: The key player". *Econometrica*, 74(5), 1403–1417.
- In game theory, can link equilibrium (optimal) actions of individuals to their Bonacich centrality
- Provides a direct link from network position to individual outcomes

Application: Identifying key players

- Example: Ballester, C., Calvó-Armengol, A., & Zenou, Y. (2006). "Who's who in networks. Wanted: The key player". *Econometrica*, 74(5), 1403–1417.
- In game theory, can link equilibrium (optimal) actions of individuals to their Bonacich centrality
- Provides a direct link from network position to individual outcomes
- The *key player problem* from sociology asks: which node, when removed, leads to the largest change in *aggregate* equilibrium activity?
- Key player is fully characterized by their *intercentrality*:
  - captures node's own centrality, plus
  - its contribution to the centrality of others
- Two distinct centralities connected to key economic theory

Application: Identifying key players

Intercentrality

Fix a given centrality measure c. The intercentrality I_i of node $i \in \mathcal{N}$ can be written as:

$$I_i := \sum_{j \in \mathcal{N}} \left[c_j - c_j^{(-i)} \right]$$

where $c_j^{(-i)}$ is the centrality of node j after removing node i from the network.

- Nodes are central if they are themselves central according to an existing centrality measure
- Nodes are also central if they highly influence the centrality of others
- Paper provides solutions starting with eigenvector-like centralities c

Application: Global control networks

- Application: Vitali, S., Glattfelder, J. B., & Battiston, S. (2011). "The network of global corporate control". *PloS One*, 6(10), e25995.
- Define an ownership network $\mathbf{W}$ from entities $\mathcal{S}$ to firms $\mathcal{F}$
- Weighted directed network: $W_{i,j} \in [0,1]$ is the percentage of ownership that the shareholder $i \in \mathcal{S}$ holds in firm $j \in \mathcal{F}$
- Simplest control network $\mathbf{C}$: $C_{i,j} = 1$ when $W_{i,j} > 0.5$ (majority control)

Application: Global control networks

- Application: Vitali, S., Glattfelder, J. B., & Battiston, S. (2011). "The network of global corporate control". *PloS One*, 6(10), e25995.
- Define an ownership network $\mathbf{W}$ from entities $\mathcal{S}$ to firms $\mathcal{F}$
- Weighted directed network: $W_{i,j} \in [0, 1]$ is the percentage of ownership that the shareholder $i \in \mathcal{S}$ holds in firm $j \in \mathcal{F}$
- Simplest control network $\mathbf{C}$: $C_{i,j} = 1$ when $W_{i,j} > 0.5$ (majority control)
- Map global network of control for many trans-national corporations (TNCs)
- Can use structure of network to understand global control over capital
- Let $\mathbf{v}$ be value (measured as revenue) of all firms. Define network control centrality $\mathbf{c}^{\text{net}}$ for all shareholders as

$$\mathbf{c}^{\text{net}} := \mathbf{C}\mathbf{v} + \mathbf{C}\mathbf{c}^{\text{net}} = (\mathbf{I} - \mathbf{C})^{-1}\mathbf{C}\mathbf{v}$$

Application: Global control networks

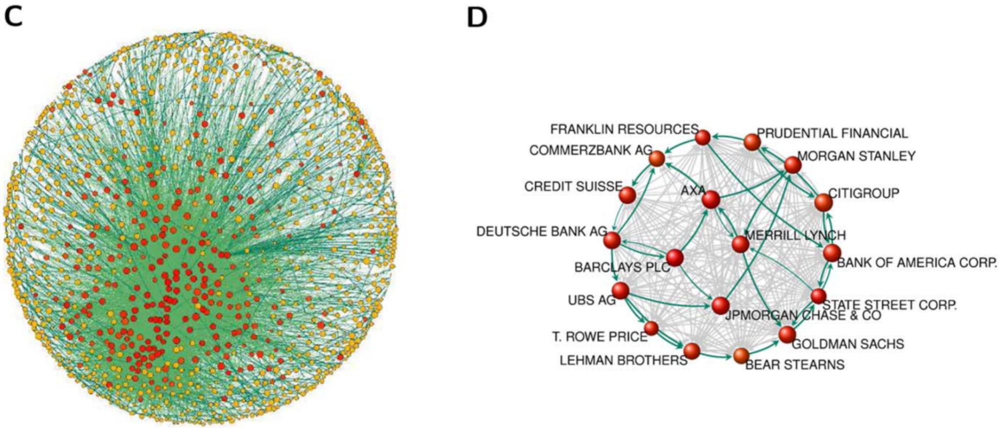


Figure: Graph representation of global control network

Application: Global control networks

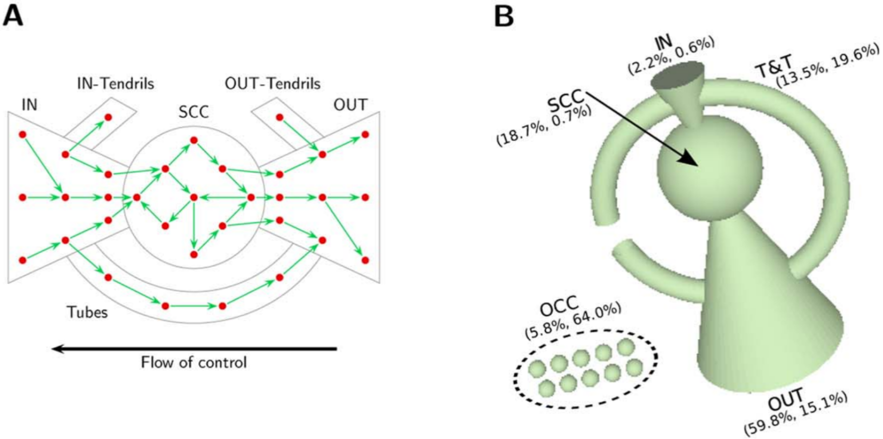


Figure: Network topology of global control

Application: Global control networks

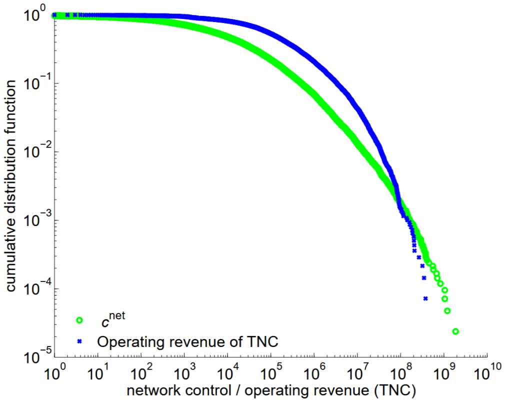


Figure: Control and firm revenue centralities

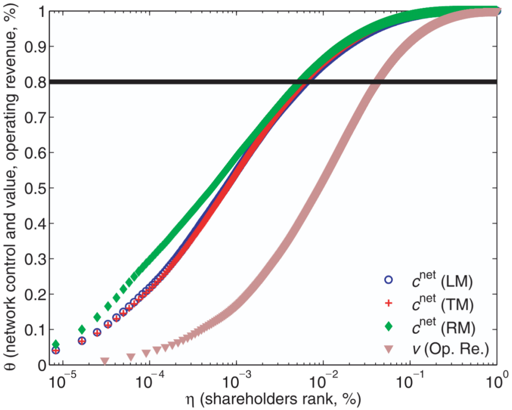


Figure: Control and firm revenue vs. concentration of shareholders

Takeaways

- Centrality formalizes importance of a node
- Different measures answer different questions
- Choice depends on context and definition of importance
- Centrality operates through many mechanisms
- Read more details:
  - Chapter 2, *Social and Economic Networks* (Jackson)
  - Chapter 7, *Networks* (Newman)