V Economic Networks and Finance Conference

Áureo de Paula
𝕏 : @PaulaAureo

UCL, IFS, Cemmap and CEPR

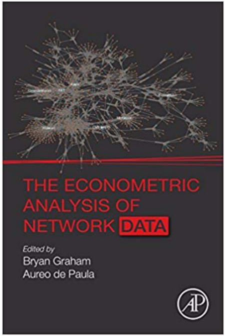
Basic concepts

Networks in economics; Descriptives.

- Background material:
  - Jackson [book, 2010];
  - de Paula [ESWC, 2017];
  - Graham-de Paula [book, 2020];
  - Graham [book chapter, 2020];

  - . . . but also a few additional elements.

(This presentation borrows from related slides by B.Graham.)



- There are many “fun” resources out there:
  - e.g., KONECT, Stanford Network Analysis Project and its data collection (data, analysis, etc.)
  - e.g., Networkx, R (for instance, here, here, here and here) (analysis, visualisation, etc.)
  - e.g., Gephi, Graphviz, Cytoscape (visualisation)
  - . . .

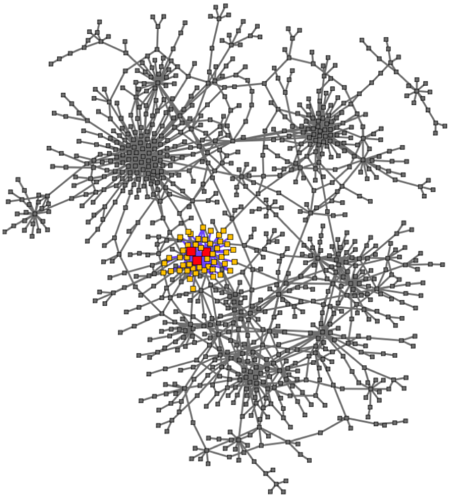
Networks in Economics

- Social and economic networks mediate many aspects of individual choice and outcomes.
  - Individuals, Households.
    ...tech adoption, risk sharing, learning, crime, consumption ...
  - Firms.
    ...buyer-supplier networks, contagion ...
  - Other (countries, states, etc.)
    ...gravity equations, yardstick competition ...

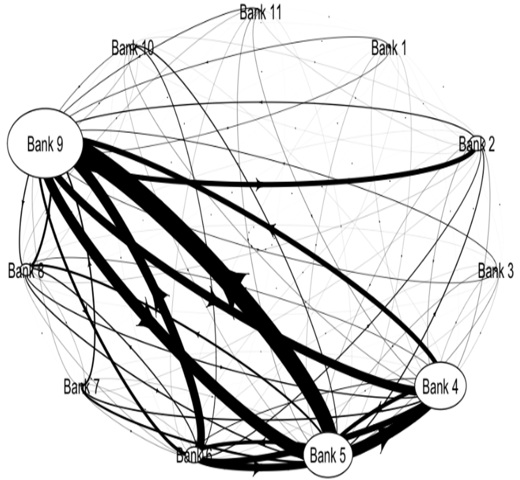
- "Connections" (direct and indirect) define (and are possibly defined by) how information, prices and quantities reverberate.

  Network formation models $\Rightarrow$ correlates and determinants of such relationships.

Look around!



Source: Atalay, Hortacsu, Roberts and Syverdson (2011)



Source: Denbee, Julliard, Li and Yuan (2021)

Source: Paul Butler

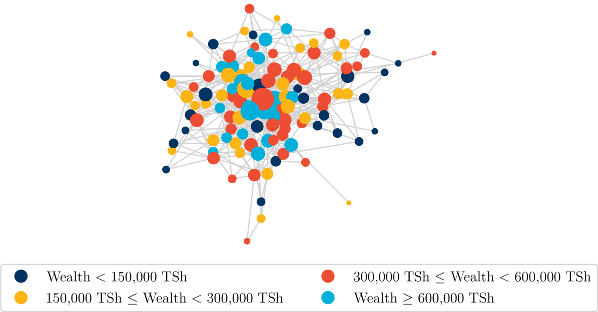
Some Basic Terminology

- Networks $\equiv$ graphs: $g = (\mathcal{N}_g, \mathcal{E}_g)$.
  ($\mathcal{N}_g$: *nodes*, *vertices*); ($\mathcal{E}_g$: *edges*, *links*, *ties*)

- $\mathcal{E}_g$ = unordered (ordered) node pairs $\Rightarrow$ undirected (directed) network.
  (e.g., Fafchamps-Lund [2003]) (e.g., Atalay et al. [2011])

- Connections can also be "weighted."
  (e.g., Diebold-Yilmaz [2015]) (e.g., Attanasio-Krutikova [2020])

- Consider, for instance, the (undirected) Nyakatoke risk-sharing network collected by De Weerdt [2004]:



- $|\mathcal{N}_g| = 119$ and $|\mathcal{E}_g| = 490 \ll \binom{|\mathcal{N}_g|}{2} = 7,021$.

  You can download the data here.

A few ways to represent networks

- Adjancency matrix: $W_{N\times N}$. ($N \equiv |\mathcal{N}_g|$)
  (W_{ij} represents ij edge)

- In an undirected and unweighted network,

$$W_{ij} = \mathbf{1}(\{i,j\} \in \mathcal{E}_g).$$

- No self-ties (loops) and unordered edges (with no more than one edge per pair) (i.e., ‘simple’ graph) $\Rightarrow$ W is symmetric with zero diagonal.

- The adjacency matrix for a directed network (or di-graph) is not necessarily symmetric.

- A weighted network will yield a non-binary (weighted) adjacency matrix.

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

- Agent 1 is connected with agents 2 and 5
- Agent 2 is connected with agent 1
- Agent 3 is isolated, etc.

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

- Agent 5 is connected to agents 1 and 4.
- Agents 2 and 5 are indirectly connected through agent 1 (i.e., share her as a common friend)

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

- 3 out of 10 possible ties are present in the network.

- There are other registries for a network:
  - Incidence matrix: $|\mathcal{N}_g| \times |\mathcal{E}_g|$ binary matrix.
  - Adjacency list: list of neighbours for every vertex.
  - ...
- These may matter computationally.
  - A sparse network may be more efficiently stored as an adjacency list than matrix.
  - Number of neighbours in adjacency list $=$ length of the list; adjacency matrix: needs to scan a whole row ($O(|\mathcal{N}_g|)$).
  - ...
- We will nonetheless focus here on the adjacency matrix as is commonly done in the literature.

Agents, Dyads, Triads and Tetrads

- However we register it, a network consists of

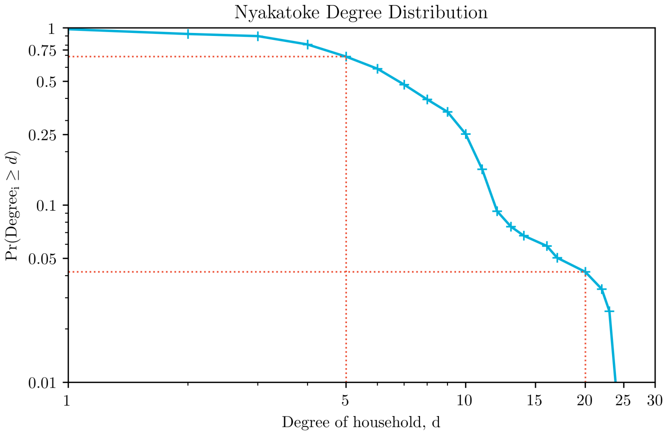
- $|\mathcal{N}_g| = N$ agents;
- $\binom{N}{2} = \frac{1}{2}N(N-1) = O(N^2)$ pairs of agents or dyads;
- $\binom{N}{3} = \frac{1}{6}N(N-1)(N-2) = O(N^3)$ triples of agents or triads;
- $\binom{N}{4} = \frac{1}{24}N(N-1)(N-2)(N-3) = O(N^4)$ quadruples of agents or tetrads ...

- When summarising a network adjacency matrix, it is convenient in fact to conceptualise statistics in terms of

  1. agent;
  2. dyad;
  3. triad; or
  4. p-subgraph-level attributes.

Agent-level Statistics: Degree

- $N_i(g)$: set of neighbours incident with node i in g.
- Degree of node $i \equiv D_i = |N_i(g)|$ the degree sequence of a network is $\mathbf{D}_{N\times 1} = [D_i]_{i=1}^{N})$
- The degree distribution gives the frequency of each possible agent-level degree count $\{0, 1, \ldots, N-1\}$ in the network.
- Some datasets might report agent degrees without much further network information.
  (For example, Aggregate Relational Data registers "How many of your social connections have trait k?")


Nyakatoke Degree Distribution
1
0.75
0.5
0.25
0.1
0.05
0.01
$\Pr(\text{Degree}_i \geq d)$
1
5
10
15
20
25
30
Degree of household, d

Dyad-level Statistics: Density

- Dyads are either linked or unlinked.
- The count of linked dyads in the (undirected) network is $\sum_{i=1}^{N}\sum_{j<i} W_{ij} = \frac{\sum_{i=1}^{N} D_i}{2}$.
- The density of a network equals the frequency with which a randomly drawn dyad is linked:

$$\rho_N \equiv \binom{N}{2}^{-1} \sum_{i=1}^{N}\sum_{j<i} W_{ij} = \frac{2}{N(N-1)} \sum_{i=1}^{N}\sum_{j<1} W_{ij}.$$

- For the Nyakatoke network, $\sum_{i=1}^{N}\sum_{j<i} W_{ij} = 490$, $N = 119 \Rightarrow \binom{119}{2} = 7,201$.
- $\rho_N = 490/7,021 = 0.0698$.

- Note also its relation to the average degree λ_N:

$$\lambda_N \equiv \sum_{i=1}^{N} D_i/N \Rightarrow \lambda_N = \rho_N(N-1).$$

- For the Nyakatoke network, $\lambda_N = 8.23$.

- Low density and skewed degree distributions (with fat tails) are common features of real world social and economic networks.

Walks and Paths

- A walk is a sequence of edges that joins a sequence of nodes or vertices (i.e., $(e_1, \ldots, e_{n-1})$ for which $(v_1, \ldots, v_n)$ such that $e_i = (v_i, v_{i+1})$.
- A trail is a walk with no repeated edges.
- A path is a trail with no repeated nodes or vertices.
  (In graphs allowing for multiple edges between dyads, there can be trails that are not paths.)
- Oriented walks, trails and paths are analogously defined as one would naturally imagine in directed networks.

$$
W^2 = \begin{bmatrix} D_1 & \sum_i W_{1i}W_{2i} & \dots & \sum_i W_{1i}W_{Ni} \\ \sum_i W_{1i}W_{2i} & D_2 & \dots & \sum_i W_{2i}W_{Ni} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_i W_{1i}W_{Ni} & \sum_i W_{2i}W_{Ni} & \dots & D_N \end{bmatrix}
$$

- The i^{th} diagonal element in W^2 equals the number of agent i's links or her degree.
- The $(i,j)^{th}$ element of W^2 gives the number of links agent i has in common with agent j (i.e., the number of "friends in common").

- Graph Theory: the $(i,j)^{th}$ element of W^2 gives the number of walks of length two from agent i to agent j.
- If i and j share the common friend k, then a length two walk from i to j is given by $i \rightarrow k \rightarrow j$. (This is actually a path!)
- In our previous example,

$$W^2 = \begin{bmatrix} 2 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 2 \end{bmatrix}$$

$$W^3 = \begin{bmatrix} \sum_{i,j} W_{1i}W_{ij}W_{j1} & \dots & \sum_{i,j} W_{1i}W_{ij}W_{jN} \\ \vdots & \ddots & \vdots \\ \sum_{i,j} W_{1i}W_{ij}W_{jN} & \dots & \sum_{i,j} W_{Ni}W_{ij}W_{jN} \end{bmatrix}$$

- The $(i,j)^{th}$ element of W^3 gives the number of walks of length 3 between i and j.
- If both i and j are connected to k as well as to each other, then the $\{i,j,k\}$ triad is transitive (i.e., “the friend of my friend is also my friend”).

- The i^{th} diagonal element in W^3 counts the number of transitive triads or triangles to which i belongs (with $i - j - k$ and $i - k - j$ counted separately).
  - If $\{i, j, k\}$ is a closed triad it is counted twice each in the i^{th}, j^{th} and k^{th} diagonal elements in W^3.
  - $Tr(W^3)/6$ is the number of *unique* triangles in the network.

- In our previous example,

$$W^3 = \begin{bmatrix} 0 & 2 & 0 & 0 & 3 \\ 2 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 2 \\ 3 & 0 & 0 & 2 & 0 \end{bmatrix}$$

- There are three (length-3) walks betwen 1 and 5: $1 \to 5 \to 1 \to 5$, $1 \to 2 \to 1 \to 5$ and $1 \to 5 \to 4 \to 5$. (None of which is a path in this case.)

- There are no transitive triads in the network.

K-Length Walks

- In general, the $(i, j)^{th}$ element of W^K gives the number of walks of length K from i to j.

- Let $(W^K)_{ij}$ denote the $(i, j)^{th}$ element of W^K.

- $W^0 = I_N$ and the only zero length walks in the network are from each agent to herself.

- Under the maintained hypothesis, $(W^K)_{ij}$ equals the number of K-length walks from i to j. The number of $K+1$ length walks from i to j then equals

$$\sum_{k=1}^{N} (W^K)_{ik} W_{kj}$$

which is the $(i,j)^{th}$ element of W^{K+1}.

- The claim follows by induction.

Distance

- The distance between i and j equals the minimum length path connecting them.
- If there is no path connecting i and j, then the distance between them is infinite.
- Agents separated by a finite distance are *connected*, otherwise they are *unconnected*.

- We can use powers of the adjacency matrix to calculate these distances:

$$M_{ij} = \min_k \{k : (W^k)_{ij} > 0\}$$

- If the network consists of a single connected component, we can compute average path length as

$$\overline{M} = \binom{N}{2}^{-1} \sum_{i=1}^{N} \sum_{j<i} M_{ij}.$$

- Common protocols to find shortest paths between two nodes build on Dijkstra's algorithm [1956, published in 1959] for directed or undirected networks (with non-negative edge weights).
- Other algorithms exist and their relative performance depends on features of the network (e.g., sparsity) and its storage (e.g., adjacency list or other).
- Many of those have connections with (approximate) dynamic programming computational methods (see, e.g., Sniedovich [2006]).

- Other alternatives exist (see, e.g., here.)

A Small Detour

- If direct computation of network features is costly, one can alternatively resort to sampling (see, e.g., here for average path length).

"[W]e point out that sampling and estimation are also being used in a proactive manner in the context of large network graphs, as a way of producing computationally efficient 'approximations' to quantities that, if computed for the full network graph, would be prohibitively expensive. Examples include the estimation of centrality measures (. . .) and the detection of so-called 'network motifs' (. . .)" (Kolaczyk [2009])

- There are different ways to sample from a network:
  - induced subgraph sampling: random sampling of vertices (and edges between those);
  - incident subgraph sampling: random sampling of edges (and incident vertices);
  - star (and snowball) sampling: random sampling of vertices and all their direct neighbours (and indirect, for “snowball” as in a ‘spider’ programme);
  - ...

- And these matter!
  - Different sampling schemes can be used to rationalise, for example, the friendship paradox.
  - For average degree, star sampling produces good estimates while incident subgraph sampling tends to produce lower estimates (see Fig.5.1 in Kolaczyk [2009]).

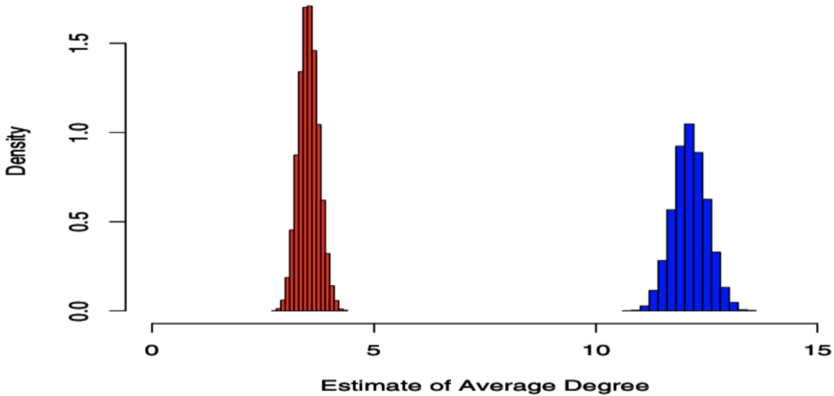


Fig. 5.1 Histograms of estimated average degree in the yeast protein interaction network, based on sampling under Design 1 (blue) and Design 2 (red), over 10,000 trials.

Small Worlds

Table: Frequency of minimum path lengths in Nyakatoke network

	1	2	3	4	5
Count	490	2,666	3,298	557	10
Frequency	0.0698	0.3797	0.4697	0.0793	0.0014

- Less than 7% of all pairs of households are directly connected.
- ...but over 40% of dyads are no more than two degrees apart.
- ...and over 90% are separated by three or fewer degrees.

- Diameter: largest distance between two agents.
- The diameter of the Nyakatoke network is 5.
- Small worlds: sparsity and low diameter together (Milgram [1967]).

- Goyal, van der Leij and Moraga-Gonzalez [2006] (updated in Rose [2022]):
  1. $|\mathcal{N}_g| = N \gg |\mathcal{E}_g|$;
  2. Diameter is small ($O(\ln N)$);
  3. (Clustering is high: $CI_g \gg \lambda_N/N \approx \rho_N$;) CI
  4. Large share of $\mathcal{N}_g$ is connected.

((1)-(3) = Watts [1999])

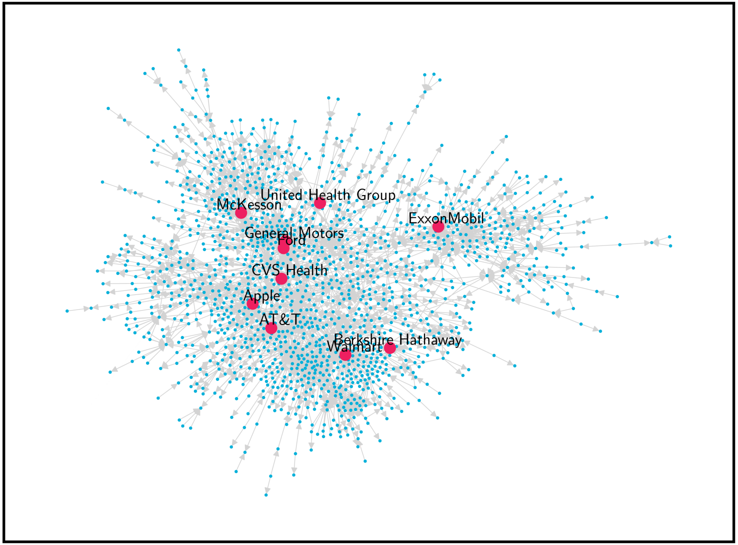
TABLE 1
Network Statistics for the Coauthor Networks

	1970s	1980s	1990s
Total authors	33,770	48,608	81,217
Degree:			
Average	.894	1.244	1.672
Standard deviation	1.358	1.765	2.303
Giant component:			
Size	5,253	13,808	33,027
Percentage	15.6%	28.4%	40.7%
Second-largest component	122	30	30
Isolated authors:			
Number	16,735	19,315	24,578
Percentage	49.6%	39.7%	30.3%
Clustering coefficient	.193	.182	.157
Distance in giant component:			
Average	12.86	11.07	9.47
Standard deviation	4.03	3.03	2.23

Directed Networks (Digraphs)

- In some settings ties are naturally directed:
  - Buyer-Supplier networks
  - International trade flows
  - Financial networks

- (In these cases the ties are also naturally weighted. Several of the measures discussed here can be adapted to that context (see, e.g., Barrat et al. [2004], Newman [2004] or Horvath [2011])


United Health Group
McKesson
ExxonMobil
General Motors
Ford
CVS Health
Apple
AT&T
Berkshire Hathaway
Walmart

- If a firm supplies inputs to another firm, then there is an oriented edge ($\bullet \rightarrow \bullet$) from the supplier to the buyer.
- The supplying firm (left node) is called the *tail* of the edge and the buying firm (right node) is its *head*.
- $\mathcal{E}_g$ is a set of ordered pairs. (In undirected networks, $\mathcal{E}_g$ is a set of unordered pairs.)

Walks and Paths in Directed Networks

- Walks and paths in directed networks have an orientation (i.e., like a one-way road).
- It may be possible to travel from i to j via series of directed paths, but not the reverse direction.
- If a path runs from i to j, but not from j to i, we say i and j are *weakly connected*.
- If a path runs in both directions, the two agents are *strongly connected*.

- In directed networks $W_{ij} = 1$ if i directs a link to j.
- If j also directs a tie to i, then $W_{ji} = 1$ and we say that the link is *reciprocated* $(\bullet \leftrightarrow \bullet)$
- The adjacency matrix for a directed network need not be symmetric.
- The $(i, j)^{th}$ entry of W^K still gives the number of K length walks from i to j.

- The indegree of agent i equals the number of arcs directed toward her, while her outdegree equals the number of links she directs toward other agents.

Indegree: $W_{+i} = \sum_j W_{ji}$ (column sums of W)

Outdegree: $W_{i+} = \sum_j W_{ij}$ (row sums of W)

Table: Top Buying Firms by Indegree, 2015

Firm	Number of Suppliers
Walmart Stores Inc.	115
Royal Dutch Shell pls	48
McKesson Corp.	41
Cardinal Health Inc.	40
Home Depot Inc.	37
AmerisourceBergen Cop.	35
Ford Motor Co.	28
Target Corp.	26
AT&T Inc.	22

Reciprocity Index

- The frequency of asymmetric dyad configurations in g equals

$$\hat{P}(\bullet \to \bullet) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j<i} [W_{ij}(1 - W_{ji}) + W_{ji}(1 - W_{ij})].$$

- The frequency of reciprocated dyad conficurations in g equals

$$\hat{P}(\bullet \leftrightarrow \bullet) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j<i} W_{ij} W_{ji}.$$

- A standard measure of reciprocity (see, e.g., Newman [2010]) is

$$R_N = \frac{2\hat{P}(\bullet \leftrightarrow \bullet)}{2\hat{P}(\bullet \leftrightarrow \bullet) + \hat{P}(\bullet \rightarrow \bullet)}.$$

- If edges form completely at random with probability ρ_N, then

$$R_N = \frac{2\rho_N^2}{2\rho_N^2 + 2(1-\rho_N)\rho_N} = \rho_N$$

- In practice, R_N is far from ρ_N.
- For example, reciprocity is
  - common in social networks (i.e., $R_N \gg \rho_N$)
  - rare in supply-chains (i.e., $R_N \ll \rho_N$).

Centrality

- Will removal of a particular agent reduce crime more than the withdrawal of another one in a criminal network?
- 'Where' should a policy-maker introduce new technologies or innovations?
- How do agent-specific shocks percolate through a network?
- Merger analysis?
- A measure of agent "centrality" may be useful for many policy questions.

- We can start with (in- or out-) degree centrality.

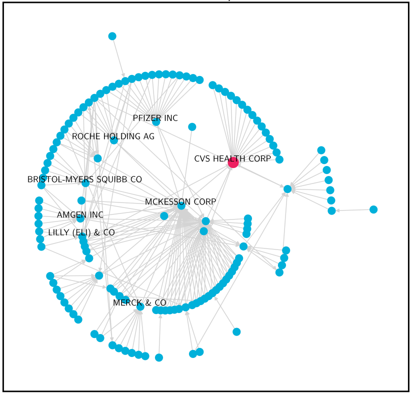
Table: Top Buying Firms by Indegree, 2015

Firm	Number of Suppliers
Walmart Stores Inc.	115
Royal Dutch Shell pls	48
McKesson Corp.	41
Cardinal Health Inc.	40
Home Depot Inc.	37
AmerisourceBergen Cop.	35
Ford Motor Co.	28
Target Corp.	26
AT&T Inc.	22

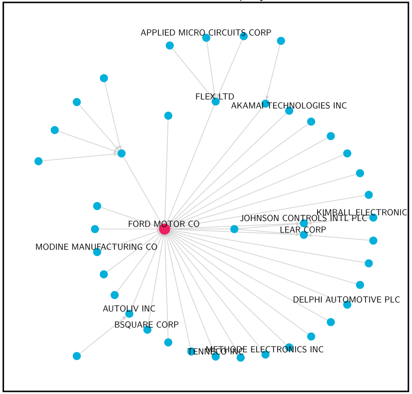
Indegree: Limitations

- Imagine two firms, both with ten suppliers.
- For the first, each of its suppliers has only one upstream supplier each.
- Firm 1 has ten direct, and ten indirect suppliers.
- For the second, each of its suppliers has ten upstream suppliers each.
- Firm 2 has ten direct and one hundred indirect suppliers.
- Which firm is more central?

CVS Health Corporation



Ford Motor Company



- Many generalisations of indegree and outdegree centrality designed to address above limitation.
- Let me focus on indegree extensions.
- The generalisation to outdegree-type measures follows easily. (Just replace W with $W^\top$.)

Eigenvector Centrality

- Bonacich (1972), building on Katz (1953), recursively defined an agent's centrality, power, or importance within a network, $c_i^{EC}(W, \phi)$, to be proportional to the sum of her links to other agents, weighted by their own centralities (see also Gould [1967]).
- Letting $\mathbf{c}^{EC}(W, \phi)$ be the N vector of centrality measures, this gives:

$$
\begin{aligned}
c_i^{EC}(W, \phi) &= \phi \sum_j c_j^{EC}(W, \phi) W_{ji} \Leftrightarrow \\
\mathbf{c}^{EC}(W, \phi)^\top &= \phi \mathbf{c}^{EC}(W, \phi)^\top W
\end{aligned}
$$

- Typically $\phi = 1/\lambda_{max}$, with λ_{max} the largest eigenvalue of W, is used for normalisation.

- This choice ensures a solution with positive values when the network is strongly connected (Perron-Frobenious Theorem).

- Since $\mathbf{c}^{EC}(W, \phi)$ is the solution to

$$\mathbf{c}^{EC}(W, \phi)[\frac{1}{\phi} I_N - W] = 0,$$

  it corresponds to the left eigenvector associated with the largest eigenvalue of W.

Row Normalisation

- Katz (1953) suggested an alternative approach to normalisation.
- The row normalised adjacency matrix is

$$G = diag\{\max(1, W_{1+}), \ldots, \max(1, W_{N+})\}^{-1} \times W$$

- The i^{th} row of G sums to either zero (if agent i has an outdegree of zero) or one (if agent i has a positive outdegree).
- If all agents have positive outdegree, then G will be a row stochastic matrix.

- Katz (1953) suggested the centrality measure

$$\begin{aligned} c_i^K(W) &= \sum_j c_j^K(W) G_{ji} \Leftrightarrow \\ \mathbf{c}^K(W)^\top &= \mathbf{c}^K(W)^\top G \end{aligned}$$

- Row normalisation ensures that the largest eigenvalue of G is one and hence that $\mathbf{c}^K(W)$ is well defined.

UCL

Markov Chain Interpretation

- If G is row stochastic, then $\mathbf{c}^K(W)$ corresponds to a stationary vector a Markov chain with transition matrix G.
- If the matrix G is irreducible, then this is stationary vector is unique (Perron-Frobenius Theorem).
- Irreducibility holds if, and only if, the network is strongly connected.
- Few real world digraphs are strongly connected.

- Assume strong connectivity.
- Traveling saleswoman process:
  1. Saleswoman begins at any node.
  2. She chooses a buyer at random from the set of buyers of her current supplier/node and moves downstream to the selected buyer/node.
  3. Repeat Step 2 many times...

- In the long run the elements of $\mathbf{c}^K(W)$ equal the proportions of time our saleswoman will spend at each node.
- Our saleswoman will spend more time at important 'buyer' nodes.
- Such nodes will be chosen more frequently at Step 2 of the traveling saleswoman process.

Dangling Nodes

- Few real world social and economic (directed) networks are strongly connected.
- “Buckets”: a strongly connected component of the digraph without outgoing links to the rest of the graph.
- Not only does strong connectivity typically fails, but many directed networks have “dangling nodes” (agents with zero outdegree).
- Traveling saleswoman will get stuck at such nodes $\Rightarrow$ problems with finding $\mathbf{c}^K(W)$.

PageRank

- The problem of dangling nodes, as well as the failure of strong connectivity, motivated Sergey Brin and Lawrence Page, then graduate students in computer science at Stanford University, to develop the PageRank centrality measure, which was the basis for Google to rank web-search results (see Gleich [2015] for a recent survey).

- Brin and Page made two changes to the Katz (1953) measure (see Franceschet [2010]):

  1. Regularise the (row normalised) adjacency matrix so that all rows, including those associated with dangling nodes, sum to one.

  2. As in Bonacich (1987), endow each agent with a small amount of exogenous centrality.

Modification # 1

- Brin and Page defined the Google Matrix $H = [H_{ij}]$ with elements

$$H_{ij} = \begin{cases} \phi G_{ij} + \frac{(1-\phi)}{N} & \text{if} W_{i+} > 0 \\ \frac{1}{N} & \text{otherwise.} \end{cases}$$

- Observe that H is both row stochastic and irreducible.

UCL

Modification # 2

- Each agent has a small amount of exogenous centrality:

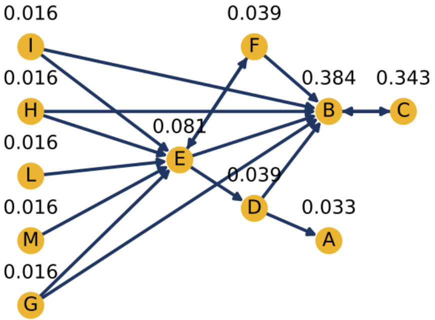
$$\mathbf{c}^{PR}(W,\phi)^\top = \phi\mathbf{c}^{PR}(W,\phi)^\top H + \frac{(1-\phi)}{N}\mathbf{1}_N$$

- A typical value for ϕ, at least in web search, is 0.85.

- For $|\phi| < 1$, we can solve for the PageRank vector as

$$\mathbf{c}^{PR}(W, \phi) = \frac{(1-\phi)}{N} \mathbf{1}_N^\top (I_N - \phi H)^{-1}$$

Figure: Example: Franceschet [2010])



- Modified traveling saleswoman process:
  1. Saleswoman begins at any node.
  2. She chooses a buyer at random
     a ...with probability ϕ from the set of buyers of her current supplier/node
     b ...with probability $1-\phi$ from the set of all firms.
  3. She moves downstream to the node selected in Step 2.
  4. Repeat Steps 2 and 3 many times...

At Step 2, if the firm has zero customers, then the saleswoman just moves to a firm, from the set of all firms, at random.

Table: Top Buyers PageRank, 2015

Firm	Buyer's PageRank
Walmart Stores Inc.	0.0272
CVS Heath Corp.	0.0198
Royal Dutch Shell pls	0.0124
AmerisourceBergen Cop.	0.0094
McKesson Corp.	0.0086
Cardinal Health Inc.	0.0060
Walgreen's Boots Alliance Inc.	0.0060
HP Inc.	0.0056
Express Scripts Holding Co.	0.0050

Social Multiplier Centrality

- Quadratic complementarity game (e.g., Jackson and Zenou [2015])
- Let Y_i be a continuously-valued action chosen by agent $i = 1, \ldots, N$.
- Let $\mathbf{Y}$ be the $N \times 1$ vector of all agents' actions.
- Let G be the row-normalised adjacency matrix.

- Observe that

$$\mathbf{G}_{i\cdot}\mathbf{Y} = \sum_{j \neq i} G_{ij} Y_j \equiv \overline{Y}_{N(i)}$$

  equals the average actional of player *i*'s direct peers.

- Assume that the network strongly connected (perhaps some open questions here?)

- The utility agent *i* receives from action profile **Y** given the network structure is

$$u_i(\mathbf{Y}, W) = (\alpha_0 + \epsilon_i) Y_i - \frac{1}{2} Y_i^2 + \beta_0 \overline{Y}_{N(i)} Y_i$$
$$= (\alpha_0 + \epsilon_i) Y_i - \frac{1}{2} Y_i^2 + \beta_0 \mathbf{G}_{i.} \mathbf{Y} Y_i$$

with $0 < \beta_0 < 1$ and $\mathbb{E}[\epsilon_i] = 0$.

- Here ϵ_i captures heterogeneity in agents' preferences for action.

- Holding peers' actions fixed, there are diminishing returns to additional action.

- The marginal utility associated with an increase in Y_i is increasing in the average action of one's peers, $\overline{Y}_{N(i)}$:

$$\frac{\partial^2 u_i(\mathbf{Y}, W)}{\partial Y_i \partial \overline{Y}_{N(i)}} = \beta_0.$$

- Own and peer action are complements.

- The magnitude of β_0 indexes the strength of any endogenous social interactions (Manski [1993]).

- The observed action **Y** corresponds to a Nash equilibrium.
- Agents observe W, the network structure, and ϵ the vector of individual level heterogeneity terms.
- The best response function is:

$$Y_i = \alpha_0 + \beta_0 \overline{Y}_{N(i)} + \epsilon_i$$

for $i = 1, \ldots, N$.

- Special case of *linear-in-means* model of social interactions.

- The best response functions define a system of simultaneous equations.
- Writing the system in matrix form gives:

$$\mathbf{Y} = \alpha_0 \mathbf{1} + \beta_0 \mathbf{G}\mathbf{Y} + \epsilon$$

- For $|\beta_0| < 1$, solving for the equilibrium action vector, $\mathbf{Y}$, as a function of W and ϵ alone, yields the reduced form:

$$\mathbf{Y} = \alpha_0 (I_N - \beta_0 \mathbf{G})^{-1} \mathbf{1} + (I_N - \beta_0 \mathbf{G})^{-1} \epsilon.$$

- Using a series representation:

$$\mathbf{Y} = \frac{\alpha_0}{1 - \beta_0} \mathbf{1} + \left[\sum_{k=0}^{\infty} \beta_0^k \mathbf{G}^k \right] \epsilon.$$

- The infinite series representation provides insight into the social multiplier.
- Consider a policy which increases the i^{th} agent's value of ϵ_i by Δ.
- The full effect of this increase on the network's distribution of outcomes occurs in "waves".
- In the initial wave only agent i's outcome increases. The change in the entire action vector is therefore

$$\Delta \mathbf{e}_i,$$

where $\mathbf{e}_i$ is an N-vector with a one in its i^{th} element and zeroes elsewhere.

- In the second wave all of agent i's peers experience outcome increases.
- Their best reply actions change in response to the increase in agent i's action in the initial wave.
- The action vector in wave two therefore changes by

$$\Delta\beta_0 \mathbf{G}\mathbf{e}_i.$$

- In the k^{th} wave we have a change in the action vector of

$$\Delta \beta_0^{k-1} \mathbf{G}^{k-1} \mathbf{e}_i.$$

- Observing the pattern of geometric decay, the "long-run" effect of a Δ change in ϵ_i on the entire distribution of outcomes is given by

$$\Delta (I_N - \beta_0 \mathbf{G})^{-1} \mathbf{e}_i.$$

- The effect of perturbing ϵ_i by Δ on the equilibrium action vector coincides with the i^{th} column of the matrix $\Delta(I_N - \beta_0 \mathbf{G})^{-1}$.
- Hence the row vector

$$\mathbf{c}^{SM}(W, \beta) = (I_N - \beta_0 \mathbf{G})^{-1}\mathbf{1}$$

equals the social multiplier centrality.

- In the presence of non-trivial network structure, the full effect of an intervention will vary heterogeneously across agents.
- Shocks to central agents will have larger aggregate effects than equally-sized shocks to less central agents.
- If we multiply the elements of $\mathbf{c}^{SM}(W, \beta)$ by $(1 - \beta)/N$ we recover PageRank (without regularisation).

Katz-Bonacich Centrality

- This measure is increasing in the number of direct friends and indirect friends, with weights discounted according to the degree of separation.
- The vector of centrality measures for each agent is:

$$
\begin{aligned}
\mathbf{c}^{KB}(W,\phi) &= \phi \mathbf{1}^\top W + \phi^2 \mathbf{1}^\top W^2 + \dots \\
&= (\phi \mathbf{1}^\top W)(I_N + \phi W + \phi^2 W^2 + \dots) \\
&= (\phi \mathbf{1}^\top W)\left[\sum_{k=0}^{\infty} \phi^k W^k\right].
\end{aligned}
$$

- For $\phi < 1/\lambda_{max}$, the sequence converges so that

$$\mathbf{c}^{KB}(W,\phi) = (\phi \mathbf{1}^\top W)(I_N - \phi W)^{-1}.$$

- For $\phi \to 1/\lambda_{max}$ from below $\mathbf{c}^{KB}(W,\phi) \to \mathbf{c}^{EC}(W,\phi)$.

- Related to equilibrium effort in quadratic complementarity games on networks (e.g., Jackson and Zenou [2015]).

- See Calvó-Armengol, Patacchini and Zenou [2009] for an early example and Denbee, Julliard, Li and Yuan [2021] for a recent one.

Laplacian and Some Properties

- The Laplacian matrix for a graph is given by $L = D - W$.

- For an undirected, unweighted network the Laplacian is symmetric with node degrees on the diagonal and 0 or -1 in the off-diagonals.
- $L\mathbf{1} = 0 \Rightarrow$ L is singular with 0 as an eigenvalue.
- $\mathbf{x}^\top L\mathbf{x} = \sum_{ij \in \mathcal{E}_g} (x_i - x_j)^2 \Rightarrow L$ is positive semi-definite so 0 is the smallest eigenvalue.
- The multiplicity of the eigenvalue 0 corresponds to the number of components in the network.

- A related matrix is the normalised Laplacian:

$$\mathcal{L} = D^{-1/2} L D^{-1/2} - W = \mathbf{I} - D^{-1/2} W D^{-1/2}$$

(If $d_i = 0$, let $(D^{-1/2})_{ii} = 0$.)

- It has the same properties as L above and in addition:

Let g be connected, and let $\lambda_{\max}$ *be the largest eigenvalue of* $\mathcal{L}$*. Then* $\lambda_{\max} \leqslant 2$*, and equality holds if and only if g is bipartite.*

- Features of the (normalised) Laplacian are informative about the network and have been used in different contexts (eg., Jochmans and Weidner [2019], Leung [2023]). ◂ Example 1 ◂ Example 2

- Consider, for example, the *conductance* of a particular cut:

$$\phi(S) \equiv \frac{\sum_{i \in S, j \notin S} W_{ij}}{\min(\sum_{i \in S, j \in \mathcal{N}_g} W_{ij}, \sum_{i \in S^c, j \in \mathcal{N}_g} W_{ij})}$$

where $S \subset \mathcal{N}_g$ (see, eg., Kannan, Vempala and Vetta [2004]).

- The partition of $\mathcal{N}_g$ into S and S^c is a *cut* and $\sum_{i \in S, j \notin S} W_{ij}$ is the *size* or *weight* of the cut.
  - Cuts and related quantities appear in various domains of interest (eg., clustering, graphons, etc.)

- The conductance of a network or graph (a.k.a, Cheeger constant or isoperimetric constant when $|S| < \frac{1}{2}|\mathcal{N}_g|$) is the minimum conductance taken over all possible (non-trivial) cuts: $\phi(g) = \min_{S\subset\mathcal{N}_g, S\neq\emptyset, S\neq\mathcal{N}_g} \phi(S)$
- It encodes how interwoven a graph is and it can be shown that:

  (Cheeger's Inequality for Undirected Graphs) Let G be any undirected graph, and let $0 = \lambda_{\min} \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_{\max} \leqslant 2$ *be the eigenvalues of* $\mathcal{L}$*.Then*

$$2\phi(g) \geqslant \lambda_2 \geqslant \phi(g)^2/2$$

- So ... λ_2 can be seen as a measure of how easy it is to split the network. (It is known as Fiedler value or algebraic connectivity of the graph.)

Network Formation

- In some cases, peer structure plausibly (econometrically) exogenous or predetermined . . .

  . . . but many times network formed in articulation with outcomes or incentives determined on those very networks.

- Models for network formation are of interest *per se* and for their articulation with the determination of outcomes.

- Useful (though possibly imperfect) categorization:
  - Statistical Models
  - Strategic Models

Statistical Models

- Statistical model: $(\mathcal{G}, \sigma(\mathcal{G}), \mathcal{P})$, where $\mathcal{P}$ is a class of probability distributions on $(\mathcal{G}, \sigma(\mathcal{G}))$.
- Data is one or more networks.

- Example: Erdös-Rényi. $\mathcal{G}$ is the set of $2^{N(N-1)/2}$ graphs on N nodes, $\mathcal{P}$ is indexed by p. (Zheng, Salganik and Gelman [2006] study a heterogeneous version, see also Hong and Xu [2019])

- Example: A generalization is given by the ERGM:

$$\mathbb{P}(G = g) = \exp\left(\sum_{k=1}^{p} \alpha_k S_k(g) - A(\alpha_1, \ldots, \alpha_p)\right),$$

where $S_k(g), k = 1, \ldots, p$ enumerate features of the graph g (eg., edges, triangles) and $A(\alpha_1, \ldots, \alpha_p)$ ensures that probabilities integrate to one.

- ERGM $\in$ exponential family.
  - $(S_k(g))_{k=1}^p$ is a sufficient statistic for $(\alpha_k)_{k=1}^p$ (natural parameter);
  - $A(\alpha_1, \ldots, \alpha_p) = \ln \left[\sum_{g \in \mathcal{G}} \exp \left(\sum_{k=1}^p \alpha_k S_k(g) \right) \right]$ is its cumulant or log-partition function;
  - ...

  (See Schweinberger et al. [2020] for a recent survey.)
- In principle, we can use MLE ...but $A(\alpha_1, \ldots, \alpha_p)$ involves a sum over $2^{N(N-1)/2}$ graphs.
- $N = 24 \Rightarrow |\mathcal{G}| > \#$ atoms in universe!

- One strategy: (log) pseudo-likelihood $\sum_{\{i,j\}} \ln \mathbb{P}(W_{ij} = 1 | W_{-ij} = w_{-ij}; \alpha)$ (Besag [1975], Strauss and Ikeda [1990]). Unreliable if not close to indep links.

- Two alternative avenues:
  > Variational principles (Wainwright and Jordan [2008], Blei et al. [2017]);
  > MCMC (Kolaczyk [2009]).

- Variational methods $\Rightarrow$ cumulant function as solution to an optimisation problem.
- Take an Erdos-Rényi graph on two nodes: $\{i, j\} \Rightarrow W_{ij}$ is a Bernoulli RV.

$$\mathbb{P}(W_{ij} = w_{ij}) = \exp(\alpha w_{ij} - A(\alpha)),$$

where $w_{ij} = 0, 1$ and $A(\alpha) = \ln(1 + \exp(\alpha))$.
- Since $A''(\alpha) = \exp(\alpha)/(1 + \exp(\alpha))^2 > 0$, we obtain that

$$A(\alpha) = \sup_{\mu \in [0,1]} \{\alpha\mu - A^*(\mu)\},$$

where $A^*(\mu)$ is the *convex conjugate* or Legendre-Fenchel transformation of $A(\alpha)$:

$$A^*(\mu) \equiv \sup_{\alpha \in \mathbb{R}} \{\mu\alpha - A(\alpha)\} = \mu \ln \mu + (1 - \mu) \ln(1 - \mu).$$

- How do we obtain $A^*(\mu)$ without $A(\alpha)$?

- It turns out that

$$A^*(\mu) = -\max_{p} H(p) \quad \text{s.t.} \quad \mathbb{E}_p(W_{ij}) = \mu,$$

  where $H(p) \equiv -p \ln p - (1-p)\ln(1-p)$ is the Shannon entropy (for the Bernoulli distribution).

- More generally: to obtain $A^*(\mu)$, compute H and domain of optimisation problem (not always easily characterised) $\Rightarrow$ various approximations are employed to estimate $A(\alpha)$.
  (Jordan [2004], Wainwright and Jordan [2008], Blei et al. [2017]; Braun and McAuliffe [2010], Athey et al. [2018], Ruiz et al. [2020], Mele and Zhu [2023])

- MCMC: various protocols (see, e.g., Kolaczyk [2009]).

- e.g., following Geyer and Thompson [1992]: optimise

$$\mathcal{L}(\alpha) - \mathcal{L}(\tilde{\alpha}) = \sum_{k=1}^{p} (\alpha_k - \tilde{\alpha}_k) S_k(g) - [A(\alpha_1, \ldots, \alpha_p) - A(\tilde{\alpha}_1, \ldots, \tilde{\alpha}_p)]$$

  for a fixed $\tilde{\alpha}$ where $\mathcal{L}(\cdot)$ is the (log-)likelihood function for the ERGM.

> Note that

$$\exp[A(\alpha_1, \ldots, \alpha_p) - A(\tilde{\alpha}_1, \ldots, \tilde{\alpha}_p)] = \mathbb{E}_{\tilde{\alpha}} \left[\exp \left(\sum_{k=1}^{p} (\alpha_k - \tilde{\alpha}_k) S_k(G) \right) \right].$$

  Then, estimate this by simulation under $\tilde{\alpha}$ and obtain the SMLE.

> The simulation can be done by Gibbs sampling (Glauber dynamics), Metropolis-Hastings, or other methods (e.g., inversion); one edge per iteration, or possibly more (e.g., triads) (see Snjiders [2002], Kolaczyk [2009], Mele [2017]).

- MCMC: various protocols (see, e.g., Kolaczyk [2009]).

- e.g., following Robbins and Monro [1951] (Snjiders [2002]): MLE solves the moment equations

$$\mathbb{E}_{\hat{\alpha}}\left[S(G)\right] = S(g)$$

(see, e.g., Lehmann and Casella [1998]).

> Update estimate according to

$$\hat{\alpha}_{(t+1)} = \hat{\alpha}_t - a_t D_t^{-1}(S_t - S(g))$$

where $a_t \to 0$, D_t plays the role of the Hessian (ie., an estimate for $\partial\mathbb{E}_\alpha\left[S(G)\right]/\partial\alpha$) and S_t is generated according to $\hat{\alpha}_t$.

> "The Robbins-Monro algorithm may be considered to be a Monte Carlo variant of the Newton-Raphson algorithm." (It is a precursor to stochastic gradient descent methods used in ML.)

> As before, one also needs a simulation scheme for S.

- For recent related developments see, e.g., Zhang and Liang [2023]

- Beware!

> Degeneracy or near degeneracy: abrupt changes in probable graphs as parameters change (see Snjiders [2002]). Rinaldo et al. [2009], Geyer [2009]: general in discrete exponential families. When observed sufficient statistics at or near support boundary, MLE does not exist and, when it does, MC-ML badly behaved.

> "Whenever the observed graph statistics fall on the convex hull of the sample space of graph statistics, then the MLE does not exist (Barndorff-Nielsen [1978]; Handcock [2003]) (. . .) this problem is virtually guaranteed to occur, since typically at least one element of $S(g)$ is zero for any realistic network." (Handcock and Hunter [2006])

> For parameter regions where distribution is multimodal, mixing time is slow (see discussion in Mele [2017] and the formalization in Bhamidi et al. [2011] for Glauber dynamics).

> For parameter regions where distribution is unimodal, Bhamidi et al. [2011], Chatterjee and Diaconis [2013] show that graph draws $\approx$ Erdös-Rényi model with indep link formation.

- $\mathbb{P}(W_{ij} = 1 | W_{-ij} = w_{-ij}; \alpha) = \mathbb{P}(W_{ij} = 1; \alpha) \Rightarrow$ focus on dyads.

  - Example: Holland and Leinhardt [1981] (directed network).

$$\mathbb{P}(W_{ij} = W_{ji} = 1) \propto \exp(\alpha^{\text{rec}} + 2\alpha + \alpha_i^{\text{out}} + \alpha_i^{\text{in}} + \alpha_j^{\text{out}} + \alpha_j^{\text{in}})$$

    and

$$\mathbb{P}(W_{ij} = 1, W_{ji} = 0) \propto \exp(\alpha + \alpha_i^{\text{out}} + \alpha_j^{\text{in}}).$$

    Dzemski [2019] takes αs to be "fixed effects."

  - Example: Chatterjee et al. [2011], Yan and Xu [2013] (undirected network, β-model). Graham [2017] characterizes MLE (with covar) and studies a conditional ML (using sufficient stats for α_i).

  - (A similar conditional MLE for the directed case is studied in Charbonneau [2017], Jochmans [2018].)

  - These are special cases of ERGMs (see Schweinberger et al. [2020]).

- In the models above, exchangeability plays a salient role. In particular, a result due to Aldous and Hoover for *infinite* random graphs (Kallenberg [2005], Theorem 7.22):

  The simple (infinite) random graph W is jointly exchangeable if and only if

  $$W_{ij} \overset{d}{\sim} \tilde{h}\left(\xi_0, \xi_i, \xi_j, \zeta_{ij}\right) \forall (i,j) \in \mathbb{N}^2 i \neq j$$

  for some i.i.d. random variables $\left(\xi_0, (\xi_i)_{\forall i \in \mathbb{N}}, (\zeta_{ij})_{\forall (i,j) \in \mathbb{N}^2, i \neq j}\right)$ *all uniformly distributed in* $[0,1]$*, with* $\zeta_{ij} \equiv \zeta_{ji}$*, and for some Borel measurable function* $h : [0,1]^4 \to \{0,1\}$*, symmetric in* ξ_i, ξ_j*.*

- If our data is on a single network, it is customary to condition on the realised $\tilde{h}(\xi_0, \cdot, \cdot, \cdot)$ and express the kernel function as $\tilde{h}(\xi_0, \cdot, \cdot, \cdot) \equiv h(\cdot, \cdot, \cdot)$

- Let $\pi : \mathbb{N} \rightarrow \mathbb{N}$ be a one-to-one mapping and $W \circ \pi$ be a simple random graph defined as
$$W \circ \pi \equiv \left((W \circ \pi)_{ij} \; \forall (i,j) \in \mathbb{N}^2, i \neq j \right)$$
with the random variable $(W \circ \pi)_{ij} \equiv W_{\pi(i)\pi(j)} \; \forall (i,j) \in \mathbb{N}^2$ with $i \neq j$.

- W is jointly exchangeable if $W \circ \pi \overset{d}{\sim} W \; \forall$ permutations π of $\mathbb{N}$ permuting a finite number of elements in $\mathbb{N}^2$

- This is a generalisation of the celebrated representation theorem by De Finetti ([1930],[1937], Hewitt and Savage [1955]):

  An infinite sequence $\{W_i\}_{i=1}^{\infty}$ *is exchangeable if and only if there exists a random variable* ξ_0 *with probability distribution* $F(\xi_0)$ *such that:*

$$p(W_1, \ldots, W_n) = \int \Pi_{i=1}^{n} p(W_i|\xi_0) dF(\xi_0)$$

- This is seen as a foundational result in Bayesian statistics as it says that if the data are exchangeable, then (i) a parameter ξ_0 must exist; (ii) a likelihood must exist; (iii) a prior distribution on ξ_0 must exist (Schervish [1995], Ch. 1).

- In general, ξ_0 will turn out to be related to the limit of the empirical distributions for $W_1, \ldots, W_n$.

- (Schervish [1995], Ex.1.45) Let $\{W_i\}_{i=1}^{\infty}$ be Bernoulli random variables. Then, $\overline{W}_n$ converges (a.s.) to ξ_0 and W_i are iid Bernoulli conditional on $\xi_0 = \mathbb{P}(W_i)$. ξ_0 is itself a random variable and its distribution is unique.

- It is important to recognise that De Finetti (or Aldous-Hoover) will **not** hold for finite sequences (or arrays) (see, eg., Diaconis and Freedman [1980]) though a similar representation holds with signed measures (see Konstatopoulos and Yuan [2019], Thrm 1).

- That said, AH allows one to represent the probability of a link as a mixture of conditionally independent dyadic (CID) models. A CID model is one where

$$\mathbb{P}(W_{ij} = 1 | \xi_i, \xi_j) = \overline{h}(\xi_i, \xi_j),$$

and $W_{ij} \perp\!\!\!\perp W_{kl}$ if $ij \cap kl = \varnothing$.

- According to Aldous-Hoover,

$$\mathbb{P}(W_{ij} = 1|\xi_i, \xi_j) = \int \overline{h}(\xi_0, \xi_i, \xi_j) d\xi_0,$$

with $\overline{h}(\xi_0, \xi_i, \xi_j) \equiv \int h(\xi_0, \xi_i, \xi_j, \zeta_{ij}) d\zeta_{ij}$.

- So, when $|\mathcal{N}_g| = \infty$, joint exchangeability $\Rightarrow$ mixture of CID models (AH). It can also be shown that (even when $|\mathcal{N}_g| < \infty$) a mixture of CID models is jointly exchangeable.

- If $|\mathcal{N}_g| < \infty$, exchangeability does not necessarily imply that links are formed according to a mixture of CID models (see Graham [2020]).

- Graham [2020] discusses the inclusion of covariates where permutations are taken conditional on realisations of the covariates (see Crane and Towsner [2018] and Crane [2018]). Additional work on this includes Yan et al. [2019] and Chandna et al. [2022] and references therein.

- If the sampling framework is one where the network is the one induced by randomly drawn nodes from a large (i.e., infinite) population, exchangeability would nonetheless allow one to resort to AH.
- In this case,

$$W_{ij} = \mathbf{1}\left(\overline{h}(\xi_0, \xi_i, \xi_j) \geqslant \overline{\zeta}_{ij}\right) = \mathbf{1}\left(h_0(\xi_i, \xi_j) \geqslant \overline{\zeta}_{ij}\right),$$

  where $h_0(\cdot, \cdot)$ corresponds to the realised $\overline{h}(\xi_0, \cdot\cdot)$ and is symmetric in its arguments.
- Such a measurable, symmetric function mapping $[0, 1] \times [0, 1]$ into $[0, 1]$ is usually referred to as a *graphon*. Back
- The "sampling distribution" for particular statistics is thus the one induced by repeated random sampling from the underlying infinite population and there is an active, related literature on graphons and graph limits. (A recent set of lectures on this topic can be found here: Lecture 1, Lecture 2 and Lecture 3.)

Strategic Formation

- Statistical framework "indexed" by economic models.
  (Payoff structure and equilibrium notion)

- A common $u_i(g)$ (in undirected network) is a variation of

$$\sum_{j \neq i} W_{ij} \times \left(u + \epsilon_{ij} \right) + \left| \cup_{j:W_{ij}=1} N_j(g) - N_i(g) - \{i\} \right| \nu + \sum_j \sum_{k>j} W_{ij} W_{ik} W_{jk} \omega$$

- $W_{ij} \neq 0$ if there is a link between i and j.

- u : direct utility from a link; ν : utility from indirect links (friends of friends); ω : utility from common links (friends who are friends).

- Similar specifications for directed networks.

- Transferable or non-transferable utility.

  "The issue, here, is whether a technology exists that would allow one to transfer utility between agents participating to a matching process. (...) [W]hen available, they allow agents to *bid* for their preferred mate by accepting the reduction of own gain from the match in order to increase the partner's. The exact nature of these bids depends on the context and may not take the form of monetary transfers; in family economics, for instance, they typically affect the allocation of time between paid work, domestic work, and leisure; the choice between current and future consumption; or the structure of expenditures for private or public goods." (Matching with Transfers, Chiappori, pp.5-6)

- NTU: no technology enabling agents to decrease their utility to the benefit of a potential partner;
- TU: allows transfer of utility at a constant “exchange rate” and the total gain from the matching (surplus) is what matters for stability;
- (ITU: allows for transfers but recognizes that the exchange rate between individual utilities is not constant and endogenous to the economic environment; surplus maximisation $\neq$ stability.)
- Network formation:
  - iterative;
    (Blume [1993], Watts [2001], Jackson and Watts [2002])
  - static.
    (Jackson and Wolinsky [1996], Bala and Goyal [2000])

UCL

Iterative Network Formation

- Iterative network formation: sequential meeting protocol and individuals add or subtract links at each iteration

- Example: Christakis, Fowler, Imbens and Kalyanaraman [2020], Ch.6, undirected.
  (formation $\approx$ stochastic stability analysis in Jackson and Watts [2002])

- Example: Mele [2017], Badev [2021], directed. Details
  (Potential function $\Rightarrow$ NE or k-Nash stable equilibria w/o unobservables)
  (Meeting protocol + myopic updating $\Rightarrow$ unique invariant distr on graphs)

> i.i.d. EV unobservables $\Rightarrow$ ERGM...
> (Mele [2017] suggests MC scheme to improve on performance)

- Models are fitted to AddHealth data on friendships (Mele [2020]) and outcomes (smoking, Badev [2021]) using Bayesian methods or ML.

Static Network Formation

- "Static" network formation: e.g., pairwise stability (Jackson and Wolinsky [1996]).

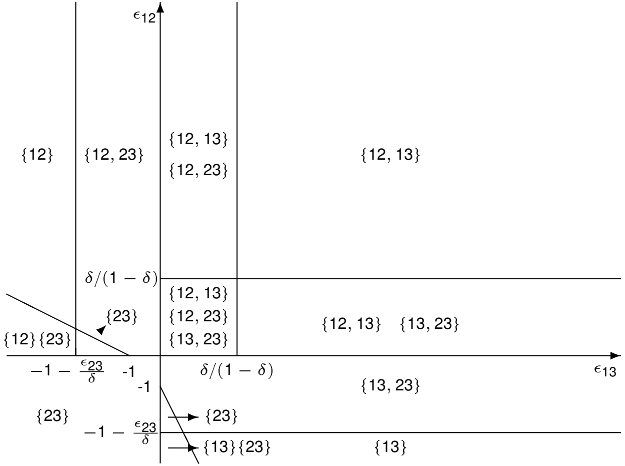
- For undirected, NTU case:

$$\forall ij \in w,\ u_i(w) \geqslant u_i(w - ij) \text{ and } u_j(w) \geqslant u_j(w - ij)$$
$$\forall ij \notin w,\ u_i(w) > u_i(w + ij) \text{ or } u_j(w) > u_j(w + ij)$$

- Any link is beneficial to both parties; and
- Non-existing links are detrimental to at least one of the parties.

- Not pairwise stability as in Gale and Shapley [1962]!
- Other versions (e.g., for TU) and alternative solution concepts (e.g., Nash for directed) are also possibilities.

- Example: $N = 3$ with payoffs $\sum_{j \in 1,\ldots,n, j \neq i} \delta^{d(i,j;w)-1}(1+\epsilon_{ij}) - |N_i(w)|$. For $\epsilon_{ij} = \epsilon_{ji}$, $0 < \epsilon_{23} < \delta/(1-\delta)$:



- Usual approach (e.g., Berry and Tamer [2006]) $\Rightarrow$ bounds on δ.

> In this graph above, for example,

$$\mathbb{P}(\epsilon_{12}, \epsilon_{13} \geqslant 0) \geqslant \mathbb{P}(\{12, 13\}) \geqslant \mathbb{P}(\epsilon_{12}, \epsilon_{13} \geqslant \delta/(1-\delta))$$

...and one could form similar bounds for all (= 8) possible networks (exploring the whole space of unobservables).

- Issue: explore equilibrium networks in the space of unobservables for different δ, but $N = 24 \Rightarrow |\mathcal{G}| > \#$ atoms in the observable universe!

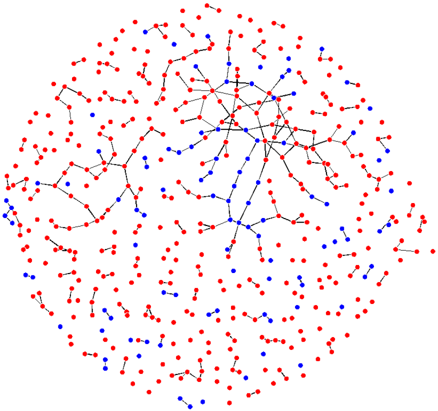
- de Paula, Richards-Shubik and Tamer [2018]: pairwise stability in (non-transferable utility) *large* networks.
  - *Large* networks: N is continuous (see Lovasz [2012] on cont graphs).
  - Payoffs: depend on characteristics (not identity), finite links and finite depth $\Rightarrow$ sparse, bounded degree graph (*graphing*).

> Focus on *network types*: characteristics of local payoff-relevant networks. Covariates with finite support $\Rightarrow$ # network types is finite.

Given parameters, proportion of network types in possible *equilibria* can be matched to data.

⋆ Verifying whether parameter is consistent with (necessary, sometimes sufficient) conditions for pairwise stability is a quadratic programme!

$N = 500 \Rightarrow$ 30secs. per parameter (on average).



- Application to co-authorship networks: Anderson and Richards-Shubik [2022].

- Sheng [2020]. Use small size subnetworks consistent with PS + additional payoff structures ⇒ bounds. (In the article, Sheng imposes (exchangeability) restrictions (on eqm sel and payoff primitives) that guarantee that these bounds are nontrivial and (if estimable), sample versions converge.) (Exchangeability ⇒ dense network: total number of links = $O_p(N^2)$.))

- Miyauchi [2016]. Payoff restrictions ⇒ complementarity (supermodularity). Use lattice structure of equilibrium set to improve computation.

- Other examples: Boucher and Mourifié [2013], Leung [2015], Gualdani [2021] . . .

- Dynamic (farsighted) network formation: e.g., Lee and Fong [2013] (bipartite), Johnson [2012] ... a few more recent developments.
  (≈ empirical dynamic games)

- Network formation and outcomes: Gilleskie and Zheng [2009], Badev [2021],. Goldsmith-Pinkham and Imbens [2013] (dyadic formation + linear-in-means), Hsieh and Lee [2016] (ERGM + linear-in-means).
  (Partial identification in formation model ⇒ partial identification in outcome model parameters. E.g., Ciliberto, Murry and Tamer [2021].)

Outcomes on Networks

- Many interdependent outcomes are mediated by connections ("networks").
- A popular representation follows the linear-in-means specification suggested in Manski [1993]. For example,

$$y_i = \alpha + \beta \sum_{j=1}^{N} W_{ij} y_j + \eta x_i + \gamma \sum_{j=1}^{N} W_{ij} x_j + \epsilon_i,$$

  with $\mathbb{E}(\epsilon_i | \mathbf{x}, W) = 0$.
- In matrix form, we have

$$\mathbf{y}_{N\times 1} = \alpha \mathbf{1}_{N\times 1} + \beta W_{N\times N} \mathbf{y}_{N\times 1} + \eta \mathbf{x}_{N\times 1} + \gamma W_{N\times N} \mathbf{x}_{N\times 1} + \epsilon_{N\times 1}$$
$$\Leftrightarrow$$
$$\mathbf{y} = \alpha(\mathbf{I} - \beta W)^{-1}\mathbf{1} + (\mathbf{I} - \beta W)^{-1}(\eta \mathbf{I} + \gamma W)\mathbf{x} + (\mathbf{I} - \beta W)^{-1}\epsilon$$

- uncover<1-3>This system can be obtained from interaction models with maximizing agents with quadratic payoffs.
  - Example: Blume, Brock, Durlauf and Jayaraman [2015]. Bayes-Nash equilibrium with

$$U_i(\mathbf{y}; W) = \left(\alpha + \eta x_i + \gamma \sum_{j \neq i} W_{ij} x_j + z_i\right) y_i + \beta \sum_{j \neq i} W_{ij} y_i y_j - \frac{1}{2} y_i^2.$$

  - Example: Calvó-Armengol, Patacchini and Zenou [2009]. Nash equilibrium with $y_i = e_i + \epsilon_i$ and

$$U_i(e_i, \epsilon; W) = \left(\eta x_i + \gamma \sum_{j \neq i} W_{ij} x_j\right) e_i - \frac{1}{2} e_i^2 + (\alpha W_i \mathbf{1} + \nu_i)\epsilon_i - \frac{1}{2}\epsilon_i^2 + \tilde{\beta} \sum_{j=1}^{N} W_{ij} \epsilon_i \epsilon_j$$

$$\Rightarrow \mathbf{y} = \frac{\alpha}{\tilde{\beta}} (\mathbf{I} - \tilde{\beta} W)^{-1} \tilde{\beta} W \mathbf{1} + (\eta \mathbf{I} + \gamma W)\mathbf{x} + (\mathbf{I} - \tilde{\beta} W)^{-1} \nu.$$

  (e.g., Denbee, Julliard, Li and Yuan [2021] and other studies.)

- See also Besley and Case [1995], De Giorgi, Frederiksen and Pistaferri [2020].

- Manski [1993] categorises "social effects" as:
  - Endogenous effect: group outcomes on individual outcome;
  - Exogenous or contextual effect: group characteristics on individual outcome;
  - Correlated effects.

. . . and the "reflection problem".

If $|\beta| < 1$, $\eta\beta + \gamma \neq 0$, $W_{ij} = (N-1)^{-1}$ if $i \neq j$ and $W_{ii} = 0$, $(\alpha, \beta, \eta, \gamma)$ is not point-identified.

Corollary to Proposition 1 in Bramoullé et al. [2009], also in Manski [1993], Kelejian et al. [2006] and others.

- Outlook improves with further restrictions on the model and/or data.
  - Example. Take the related representation originally considered in Manski [1993]:

$$y_i = \alpha + \beta\mathbb{E}(y_j|w) + \eta x_i + \gamma\mathbb{E}(x_j|w) + \epsilon_i, \quad \mathbb{E}(\epsilon_i|\mathbf{x}, w) = \delta w.$$

Manski [1993] (Prop 2) $\Rightarrow (\alpha, \beta, \eta)$ are point-identified when $\delta = \gamma = 0$ and $1, \mathbb{E}(x_j|w), x_i$ are "linearly independent in the population".
(A similar result appears in Angrist [2014].)

- This identification argument uses between-group variation in $\mathbb{E}(x_j|w)$, not used in the proposition.

- Alternative strategies explore restrictions to higher moments.

If $|\beta| < 1$, $W_{ij} = (N-1)^{-1}$ if $i \neq j$, $W_{ii} = 0$, and $\mathbb{V}(\epsilon|\mathbf{x}) = \sigma^2\mathbf{I}$ then $(\alpha, \beta, \eta, \gamma)$ is point-identified.

Moffitt [2001] ($N = 2$) and reminiscent of results like Fisher [1966].

- The cov restriction also leads to testable implications!

Proposition. If $|\beta| < 1$, $W_{ij} = (N-1)^{-1}$ if $i \neq j$, $W_{ii} = 0$, and $\mathbb{V}(\epsilon|\mathbf{x}) = \sigma^2\mathbf{I}$ then

$$\frac{\mathbb{C}(y_i, y_j|\mathbf{x})}{\mathbb{V}(y_i|\mathbf{x})} > \frac{4-3N}{4N^2-11N+8}.$$

$N \geqslant 3 \Rightarrow$ lower bound on $Corr(y_i, y_j|\mathbf{x})$, e.g.: $N = 3 \Rightarrow$ lower bound > -0.5.

- Additive group effect or shock $\Rightarrow$ identification with cov restrictions and at least two groups of different size. (Davezies, d'Haultfoeuille and Fougére [2009])

- Graham [2008] also uses higher moments to identify

$$\mathbf{y}_{l\ N_l \times 1} = \tilde{\gamma} W_{l\ N_l \times N_l} \epsilon_{l\ N_l \times 1} + \alpha_l \mathbf{1}_{N_l \times 1} + \epsilon_{l\ N_l \times 1},$$

  (see also Glaeser, Sacerdote and Scheinkman [2003]).

  $\tilde{\gamma}$ is identified if there are two groups under random assignment and additional distributional restrictions.

- Blume, Brock, Durlauf and Jayaraman [2015] explore similar ideas for the more general model.

- One setting that bears some reseblance and also uses higher-order moment restrictions is Gabaix and Koijn [2023]:

> To explain the intuition, we specialize our analysis to the case where there is only a single factor and all entities have the same loading on the factor, $\lambda_i = 1_{1\times r}$, and there are no other controls, $C_t^p = C_t^y = 0$. The single factor is then absorbed by a time fixed effect. It allows us to develop the main intuition in a transparent way. The system is
>
> $$p_t = \psi y_{St} + \varepsilon_t, \tag{38}$$
>
> $$y_{it} = \phi^d p_t + \eta_t + u_{it}. \tag{39}$$
>
> To take advantage of the great analytical simplicity of that example, we retrace the derivation steps in an elementary manner. We cannot estimate ψ and ϕ^d by OLS as ε_t and η_t are typically correlated, implying that y_{St} is correlated with ε_t in (38), and p_t with η_t in (39).

where $y_{St} = \sum_{i=1}^{N} S_i y_{it}$.

- Let $y_{0t} \equiv p_t$, $\mathbf{y}_t \equiv [y_{0t} \ \ldots \ y_{Nt}]^\top$, $\mathbf{S} \equiv [0 \ S_1 \ \ldots \ S_N]^\top$ and $\mathbf{e}_1^\top \equiv [1, 0, \ldots, 0]$ and notice that the system before can be written as

$$\mathbf{y}_t = \underbrace{\begin{bmatrix} \psi \mathbf{S}^\top \\ \phi^d \mathbf{e}_1^\top \\ \ldots \\ \phi^d \mathbf{e}_1^\top \end{bmatrix}}_{\equiv W(\psi, \phi^d, \mathbf{X})} \mathbf{y}_t + \underbrace{\begin{bmatrix} \epsilon_t \\ \eta_t + u_{1t} \\ \ldots \\ \eta_t + u_{Nt} \end{bmatrix}}_{\equiv e_t}.$$

  Then, $var(\mathbf{Y}_t) = (I - W(\psi, \phi^d, \mathbf{S})^\top)^{-1} var(e_t)(I - W(\psi, \phi^d, \mathbf{S}))^{-1}$.
- The article assumes restrictions on $var(e_t)$ (i.e., $\mathbf{u}_t \perp \eta_t, \epsilon_t$ and homoscedasticity for $\mathbf{u}_t$ or *known* heteroscedasticity). Under these conditions, one can see the above as an equation system on the parameters of interest.

- Another avenue: "exclusion restrictions" in W.

  If $\eta\beta + \gamma \neq 0$ and $\mathbf{I}$, W, W^2 are linearly independent, $(\alpha, \beta, \eta, \gamma)$ is point-identified.
  (Bramoullé, Djebbari and Fortin [2009])

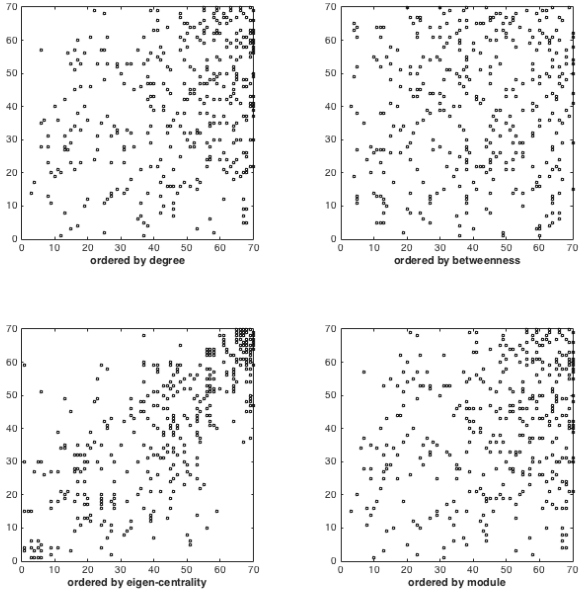
  - $W_{ij} = (N-1)^{-1}, i \neq j; W_{ii} = 0 \Rightarrow W^2 = (N-1)^{-1}\mathbf{I} + (N-2)/(N-1)W$
  - W block diagonal and two blocks of different sizes $\Rightarrow$

$$y_i = \frac{\alpha}{1-\beta} + \left[\eta + \frac{\beta(\eta\beta+\gamma)}{(1-\beta)(N_l - 1 + \beta)}\right] x_i + \frac{\eta\beta + \gamma}{(1-\beta)(1+\frac{\beta}{N_l - 1})}\overline{x}_i + \nu_i.$$

  (Lee [2007], Davezies, d'Haultfoeuille and Fougére [2009])

- Linear independence valid more generally. In fact,

  $\sum_{j=1}^{N} W_{ij} = 1$ and $\mathbf{I}$, W, W^2 linearly dependent $\Rightarrow$ W block diagonal with blocks of the same size and nonzero entries are $(N_l - 1)^{-1}$.
  (Blume, Brock, Durlauf and Jayaraman [2015])


ordered by degree
ordered by betweenness
ordered by eigen-centrality
ordered by module

de Paula, Rasul and Souza [2023]

- What if W is unknown?
- "If researchers do not know how individuals form reference groups and perceive reference-group outcomes, then it is reasonable to ask whether observed behavior can be used to infer these unknowns" (Manski [1993])

- Suppose one has panel data on outcomes and covariates:

$$y_{it} = \rho_0 \sum_{j=1}^{N} W_{0,ij} y_{jt} + \beta_0 x_{it} + \gamma_0 \sum_{j=1}^{N} W_{0,ij} x_{jt} + \alpha_t + \alpha_i + \epsilon_{it}$$

$$\Leftrightarrow$$

$$\mathbf{y}_{t,N\times 1} = \rho_0 W_{0,N\times N} \mathbf{y}_{t,N\times 1} + \beta_0 \mathbf{x}_{t,N\times 1} + \gamma_0 W_0 \mathbf{x}_t + \alpha_t \mathbf{1}_{N\times 1} + \alpha^* + \epsilon_{t,N\times 1}$$

with $\mathbb{E}(\epsilon_{it} | \mathbf{x}_t, \alpha_t, \alpha^*) = 0$.

Identification

- The model has reduced-form (assuming, for simplicity that $\alpha_t = 0$)

$$\mathbf{y}_t = \Pi_0 \mathbf{x}_t + \mathbf{v}_t$$

  where

$$\Pi_0 = (\mathbf{I} - \rho_0 W_0)^{-1} (\beta_0 \mathbf{I} + \gamma_0 W_0)$$

- If $(\rho_0, \beta_0, \gamma_0)$ were known, W_0 would be identified:

$$W_0 = (\Pi_0 - \beta_0 \mathbf{I})(\rho_0 \Pi_0 + \gamma_0 \mathbf{I})^{-1}$$

- In practice, $(\rho_0, \beta_0, \gamma_0)$ is not known.

Identification

- Further assumptions are necessary to identify $\theta_0 = (\rho_0, \beta_0, \gamma_0, W_0)$.
- Take, for example, θ_0 and θ such that $\beta_0 = \beta = 1$, $\rho_0 = 0.5$, $\rho = 1.5$, $\gamma_0 = 0.5$, $\gamma = -2.5$,

$$W_0 = \begin{bmatrix} 0 & 0.5 & 0 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0 & 0.5 & 0 \end{bmatrix} \; W = \begin{bmatrix} 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0 & 0 & 0.5 \\ 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 \end{bmatrix}.$$

- Then $(I - \rho_0 W_0)^{-1}(\beta_0 I + \gamma_0 W_0) = (I - \rho W)^{-1}(\beta I + \gamma W)$.
- (Notice that I, W_0 and W_0^2 are LI and so are I, W and W^2!)

Local Identification

- Can the model identify $\theta_0 = (\rho_0, \beta_0, \gamma_0, W_0)$?
- Assume:

(A1) $(W_0)_{ii} = 0, i = 1, \ldots, N$ (no self-links);

(A2) $\sum_{j=1}^{N} |\rho_0 (W_0)_{ij}| \leqslant 1$ for every $i = 1, \ldots, N$, $\|W_0\| < C$ for some positive $C \in \mathbb{R}$ and $|\rho_0| < 1$;

(A3) There is i such that $\sum_{j=1}^{N} (W_0)_{ij} = 1$ (normalization);

(A4) There are l and k such that $(W_0^2)_{ll} \neq (W_0^2)_{kk}$ ($\Rightarrow \mathbf{I}, W_0, W_0^2$ LI as in Bramoullé, Djebbari and Fortin [2009]);

(A5) $\beta_0 \rho_0 + \gamma_0 \neq 0$ (social effects do not cancel).

- *Under (A1)-(A5)* $(\rho_0, \beta_0, \gamma_0, W_0)$ *is locally identified.*

(Application of Rothenberg [1971].)

Global Identification

- It is nevertheless possible to strengthen local identification conclusions obtained previously.
- *Assume (A1)-(A5).* $\{\theta : \Pi(\theta) = \Pi(\theta_0)\}$ *is finite.*
  (This obtains as $\Pi(\theta)$ is a proper mapping.)
- Let $\Theta_+ = \{\theta \in \Theta : \rho\beta + \gamma > 0\}$. Then we can state that:

  Assume (A1)-(A5), then for every $\theta \in \Theta_+$ *we have that* $\Pi(\theta) = \Pi(\theta_0) \Rightarrow \theta = \theta_0$. *That is,* θ_0 *is globally identified with respect to the set* Θ_+.

Global Identification

- This uses the following result:

  Suppose the function $\Pi(\cdot)$ is continuous, proper and locally invertible with a connected image. Then the cardinality of $\Pi^{-1}(\{\overline{\Pi}\})$ is constant for any $\overline{\Pi}$ in the image of $\Pi(\cdot)$.

  (see, e.g., Ambrosetti and Prodi [1995], p.46)

- We show that the mapping $\Pi : \Theta_{+} \to \mathbb{R}^{N \times N}$ is proper with connected image, and non-singular Jacobian at any point.
- This implies that the cardinality of the pre-image of $\{\Pi(\theta)\}$ is finite and constant.
- Take $\theta \in \Theta_{+}$ such that $\gamma = 0$, $W_{1,2} = W_{2,1} = 1$ and $W_{i,j} = 0$, otherwise. The cardinality of $\Pi^{-1}(\{\Pi(\theta)\})$ is one for such θ and the result follows.

Global Identification

- Since an analogous result holds for $\Theta_- = \{\theta \in \Theta$ such that $\rho\beta + \gamma < 0\}$, we can state that:

  Assume (A1)-(A5). The identified set contains at most two elements.

- Furthermore, if $\rho_0 > 0$ and $(W_0)_{ij} \geqslant 0$ one is able to sign $\rho_0\beta_0 + \gamma_0$ and obtain that:

  Assume (A1)-(A5), $\rho_0 > 0$ and $(W_0)_{ij} \geqslant 0$. Then θ_0 is globally identified.

- Finally, if W_0 is non-negative and irreducible, one is also able to sign $\rho_0\beta_0 + \gamma_0$!

  Assume (A1)-(A5). $(W_0)_{ij} \geqslant 0$ and W_0 irreducible. Then θ_0 is globally identified if W_0 has at least two real eigenvalues or $|\rho_0| \leqslant \sqrt{2}/2$.

A Few Remarks

- One can also allow for β to vary by $i = 1, \ldots, N$:
  - ... with multivariate $\mathbf{x}_{i,t}$ as long as one of the covariates has homogeneous β; or
  - ... if $\gamma = 0$ as long as $\beta_i \neq \beta_j$ for every $i \neq j$.

- Time-varying $(\rho_t, \beta_t, \gamma_t, W_t)$ can be identified from Π_t. Estimation can be adapted from strategies available in the current literature (e.g., kernels, STAR, etc.).

- Further extensions in the paper!

Estimation Strategies

- Identification results hold for any protocol delivering an estimator for Π_0.
- Π_0 has N^2 parameters, and possibly $NT \ll N^2$ ($N = 48 \Rightarrow N^2 = 2,304$ parameters).
- Feasible if W_0 (or Π_0) are sparse.
  (e.g., Atalay et al. [2011] $< 1\%$; Carvalho [2014] $\approx 3\%$; AddHealth $\approx 2\%$; US state neighbors $\approx 7\%$; Manresa [2016] ($\beta = 0$, LASSO), Bonaldi, Hortacsu and Kastl [2014] ($\beta = 0$, elastic net)).

- **Penalization in the structural form** (e.g., Adaptive Elastic Net GMM of Caner and Zhang [2014]:

  - $\mathbf{x}_t \perp \epsilon_t \Rightarrow$ moment conditions.

$$\tilde{\theta} = (1 + \lambda_2/T) \cdot \arg\min_{\theta \in \mathbb{R}^p} \left\{ g(\theta)^\top M_T g(\theta) + \lambda_1 \sum_{i,j=1}^{n} |w_{i,j}| + \lambda_2 \sum_{i,j=1}^{n} |w_{i,j}|^2 \right\}$$

  and

$$\hat{\theta} = (1 + \lambda_2/T) \cdot \arg\min_{\theta \in \mathbb{R}^p} \left\{ g(\theta)^\top M_T g(\theta) + \lambda_1^* \sum_{\tilde{w}_{i,j} \neq 0} \frac{|w_{i,j}|}{|\tilde{w}_{i,j}|^\gamma} + \lambda_2 \sum_{i,j=1}^{n} |w_{i,j}|^2 \right\}$$

  where $\theta = (\text{vec}(W)^\top, \rho, \beta, \gamma)^\top$ and λ_1^*, λ_1 and λ_2 chosen by BIC.

  - The convergence rate for this estimator is shown to be $\sqrt{T/(dN)}$, where d is the density of the network.

- Nonlinearities:
  - "social effects might be transmitted by distributional features other than the mean" Manski [1993], and/or
  - in the "link" function (i.e., $y_i = f\left(\sum_{j=1}^{N} W_{ij} y_j, x_i, \sum_{j=1}^{N} W_{ij} x_j, \epsilon_i\right)$).
- Example: Tao and Lee [2014], Tincani [2018].
- Example: Brock and Durlauf [2001, 2007], Xu and Lee [2015] ← Bramoullé et al. [2014]; Blume, Brock, Durlauf and Ioannides [2011].

- Multiplicity. (de Paula [2013])

- Manski [2013]: potential outcomes with social interactions.

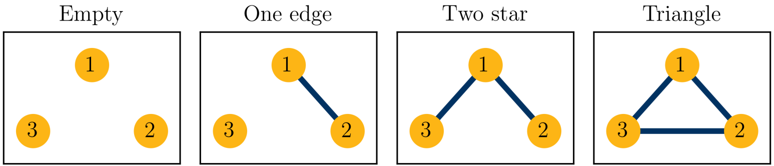
$$y_i(\mathbf{d}) = f\left(W_i, \mathbf{y}_{-i}(\mathbf{d}), \mathbf{d}, \epsilon_i\right)$$

(Consumption in PROGRESA, Angelucci and De Giorgi [2009]; spillovers in scholarship program, Dieye et al. [2014]; epidemiology)

- W also possibly affected by the treatment (Comola and Prina [2021]).

Appendix

Triad Census



- Triads come in four types (isomorphisms):
  - no connections or empties;
  - one connection or one-edges;
  - two connections or two-stars;
  - three connections or triangles.

- A complete enumeration of them into their four possible types constitutes a triad census.

Triad Census: Triangles

- Each agent can belong to as many as $\binom{N-1}{2} = \frac{(N-1)(N-2)}{2}$ triangles.
- The N diagonal elements of W^3 count those triangles as we have seen before, but we need to adjust for 'double counting':

$$T_T = \frac{Tr(W^3)}{6}.$$

- In our toy example, this is zero.

(There might be more efficient ways of counting triangles: see here.)

UCL

Triad Census: Two-Stars

- Each dyad can share up to $N-2$ links in common.
- These counts are contained in the lower (or upper) off-diagonal elements of W^2.
- Each triad appears three times in these counts: as $\{i,j,k\}$, $\{i,k,j\}$ and $\{j,k,i\}$. If it is a
  - two-star, only one of $W_{ji}W_{ki}$, $W_{ij}W_{kj}$ or $W_{ik}W_{jk}$ quantities will equal one.
  - triangle, then all three will equal one.

- This means that $vech(W^2)^\top \mathbf{1}$ gives the network count of *three times* the number of triangles plus the number of two-stars.

- Therefore
$$T_{TS} = vech(W^2)^\top \mathbf{1} - \frac{Tr(W^3)}{2}$$
equals the number of two-star triads in the network.

- In our toy example, this is two and corresponds to $1-2-4$ and $2-4-5$.

Triad Census: One-Edges

- If all triads are empty or have only one edge, then there will be $(N-2)vech(W)^\top \mathbf{1}$ one edge triads.
- If some triads are two-stars or triangles, this count will be incorrect.
- Subtracting twice the number of two stars and three times the number of triangles gives the correct answer:

$$T_{OE} = (N-2)vech(W)^\top \mathbf{1} - 2vech(W^2)^\top \mathbf{1} + \frac{Tr(W^3)}{2}$$

- In our toy example, this is $(5-2) \times 3 - 2 \times 2 = 5$.

UCL

Triad Census: Empties

- The number of empty triads, T_E, equals $\binom{N}{3}$ minus the total number of other triad types.
- In our toy example this is equal to $\binom{5}{3} - 0 - 2 - 5 = 10 - 7 = 3$. These are the triads $\{1,3,4\}, \{2,3,4\}$ and $\{2,3,5\}$.

Table: Triad Census: Nyakatoke Network

	Empty	One-Edge	Two-Star	Triangle
Count	221,189	48,245	4,070	315
Proportion	0.8078	0.1762	0.0149	0.0012
Random	0.8049	0.1812	0.0136	0.0003

Clustering

- The clustering index (a.k.a. transitivity index) is

$$CI = \frac{3T_T}{T_{TS} + 3T_T}$$

- In random graphs, the CI should be close to network density.
- For the Nyakatoke network $CI = 0.1884$ and $\rho_N = 0.0698$.
- In the economics co-authors network, $CI_{1990s} = 0.157$ and $\rho_{N,1990s} = 0.0000206$.

- Let $Pr(W_{ij} = 1) = \rho_N$ with all edges forming independently.
- Probability that a randomly drawn triad is a triangle is ρ_N^3.
- Probability that a randomly drawn triad is a two-star is $3 \times \rho_N^2(1 - \rho_N)$.
- In a random graph,

$$CI \approx \frac{3\binom{N}{3}\rho_N^3}{3\binom{N}{3}\rho_N^3 + \binom{N}{3}3\rho_N^2(1-\rho_N)} = \rho_N.$$

Back

Degree Distribution Redux

- Average degree equals $\lambda_N = \frac{2T_{OE}+4T_{TS}+6T_T}{N(N-2)}$
- Degree Variance equals

$$S_N^2 = \frac{2}{N}(T_{TS} + 3T_T) - \lambda_N(1 - \lambda_N)$$

- Knowledge of mean degree, variance and number of triangles is equivalent to knowledge of triad census.
- Degree distribution constrains other (local) features of the network.

Example 1: Jochmans and Weidner [2019]

- This paper studies how network structure (in particular, algebraic connectivity) affects the accuracy of fixed effects estimates in linear models on bipartite networks (eg., worker-firm, teacher-student,...).
- Other papers dealing with related aspects include:
  - Andrews et al. [2008] (downward bias on worker-firm effect correlation);
  - Rockoff [2004] (upward bias on teacher effect variance in teacher-student panel).

- Consider an undirected network g on $|\mathcal{N}_g| = n$ vertices and $|\mathcal{E}_g| = m$ edges with (possibly weighted) adjacency matrix W as before.
- For $e \in \mathcal{E}_g$, let $\varepsilon_e \in \{1, \ldots, m\}$ be an enumeration of its edges. (The paper allows g to be a multigraph, but I will abstract from that.)
- Its $m \times n$ (oriented) incidence matrix $\boldsymbol{B}$ is given by

$$(\boldsymbol{B})_{\varepsilon_e i} := \begin{cases} \sqrt{W_e} & \text{if } e = \{i, j\} \text{ for some } j \in \mathcal{N}_g \text{ and } i < j, \\ -\sqrt{W_e} & \text{if } e = \{i, j\} \text{ for some } j \in \mathcal{N}_g \text{ and } i > j, \\ 0 & \text{otherwise.} \end{cases}$$

  (The choice of orientation is immaterial for their analysis.)
- The (*oriented*) incidence, adjacency and Laplacian matrices are related as $L := B'B = D - W$. (If $\tilde{B}$ is the *un*oriented incidence matrix, $\tilde{B}'\tilde{B} = D + W = L + 2W$.)

- Given the graph g, for each edge $e \in \mathcal{E}_g$ we observe an outcome y_{ε_e} and a p-vector of covariates $\boldsymbol{x}_{\varepsilon_e}$. (A multi-graph would accomodate a panel!)

- Let $\boldsymbol{\alpha} := (\alpha_1, \ldots, \alpha_n)' \in \mathbb{R}^n$ be a vector of vertex-specific parameters.

- Stacking the observations one gets:

$$y_{|\mathcal{E}_g| \times 1} = B_{|\mathcal{E}_g| \times |\mathcal{N}_g|} \boldsymbol{\alpha}_{|\mathcal{N}_g| \times 1} + X_{|\mathcal{E}_g| \times p} \boldsymbol{\beta}_{p \times 1} + u_{|\mathcal{E}_g| \times 1}$$

- The outcomes for a given pair (i, j) depend on the individual effects through $\alpha_i - \alpha_j$ which remains the same if we switch to $\tilde{\alpha}_i = \alpha_i + c, c \in \mathbb{R}$. (In other words, $\mathbf{1} \in \mathcal{N}(B)$.)

- Let $d := (d_1, \ldots, d_n)'$ and impose the normalisation:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} (W)_{ij} \left(\alpha_i + \alpha_j\right) = 0 \Leftrightarrow \boldsymbol{d}'\boldsymbol{\alpha} = 0$$

The standard estimator of α is the constrained least-squares estimator

$$\check{\boldsymbol{\alpha}} := (\check{\alpha}_1, \ldots, \check{\alpha}_n)' = \underset{a \in \{a \in \mathbb{R}^n : d'a = 0\}}{\arg\min} \left\| \boldsymbol{M}_X \boldsymbol{y} - \boldsymbol{M}_X \boldsymbol{B} \boldsymbol{a} \right\|^2,$$

where $\| \cdot \|$ is the Euclidean norm, $\boldsymbol{M}_{\boldsymbol{X}} := \boldsymbol{I}_m - \boldsymbol{X} \left(\boldsymbol{X}' \boldsymbol{X} \right)^{-1} \boldsymbol{X}'$, and $\boldsymbol{I}_m$ is the identity matrix of dimension $m \times m$.

- For any matrix $\boldsymbol{C}_{n\times n}$, let $\boldsymbol{C}^{\dagger}$ be its Moore-Penrose pseudoinverse. Define

$$\boldsymbol{C}^{\star} := \boldsymbol{D}^{-1/2}\left(\boldsymbol{D}^{-1/2}\boldsymbol{C}\boldsymbol{D}^{-1/2}\right)^{\dagger}\boldsymbol{D}^{-1/2}.$$

($\boldsymbol{C}^{\star}$ is itself a pseudoinverse of $\boldsymbol{C}$.)

Let g be connected, $\mathrm{rank}(\boldsymbol{X}) = p$*, and* $\mathrm{rank}((\boldsymbol{X}, \boldsymbol{B})) = p + n - 1$*. Then*

$$\check{\alpha} = \left(B' M_X B\right)^{\star} B' M_X y$$

and is unique.

- Omit $\boldsymbol{X}$ for simplicity so that

$$\hat{\alpha} := \left(\boldsymbol{B}'\boldsymbol{B}\right)^{\star} \boldsymbol{B}'\boldsymbol{y}$$

- To study how the structure of g affects the estimation problem, assume first that $\boldsymbol{u} \sim \left(\mathbf{0}, \sigma^2 \boldsymbol{I}_m\right)$. Then

$$\text{var}(\hat{\alpha}) = \sigma^2 \left(\boldsymbol{B}'\boldsymbol{B}\right)^{\star} = \sigma^2 \boldsymbol{L}^*$$

where $\boldsymbol{L}^* = \boldsymbol{D}^{-1/2}\left(\boldsymbol{D}^{-1/2}\boldsymbol{L}\boldsymbol{D}^{-1/2}\right)^{\dagger}\boldsymbol{D}^{-1/2} = \boldsymbol{D}^{-1/2}\left(\mathcal{L}\right)^{\dagger}\boldsymbol{D}^{-1/2}$ and $\mathcal{L}$ is the normalised Laplacian.

- This implies that

$$\text{var}(\hat{\alpha}_i) = \sigma^2 \frac{\left(\mathcal{L}^{\dagger}\right)_{ii}}{d_i}$$

- The estimator precision will depend on sample size through d_i, but even as it grows with the sample the variance is still dependent on $\mathcal{L}$, which may change as the network grows.

- Let

$$h_i := \left(\frac{1}{d_i} \sum_{j \in N(i)} \frac{(\boldsymbol{W})_{ij}^2}{d_j} \right)^{-1}.$$

This is a (weighted) harmonic mean which is increasing in the degree of i's direct neighbours.

Let g be connected and suppose that $\boldsymbol{u} \sim (\boldsymbol{0}, \sigma^2 \boldsymbol{I}_m)$. *Then*[1]

$$\sigma^2 \left(\frac{1}{d_i} - \frac{1}{m} \right) \leqslant \text{var}(\hat{\alpha}_i) \leqslant \sigma^2 \left(\frac{1}{d_i} \left(1 + \frac{1}{\lambda_2 h_i} \right) - \frac{1}{m} \right).$$

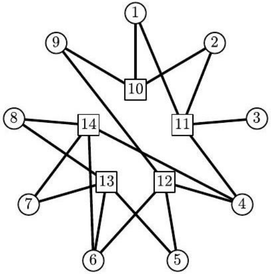
- This indicates that $h_i \lambda_2 \to \infty \Rightarrow \hat{\alpha}_i$ converges at parametric rates (ie., $d_i^{-1/2}$).

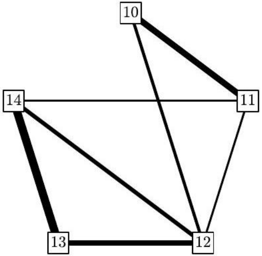
[1] In the paper, the result has $2/m$ instead of $1/m$. This is a typo.

- The paper connects the above model to two-way fixed effects in bipartite graphs representing workers and firms, teachers and students, etc. where

$$y_{|\mathcal{E}_g|\times 1} = B_{1,|\mathcal{E}_g|\times v_1}\boldsymbol{\mu}_{v_1\times 1} + B_{2,|\mathcal{E}_g|\times v_2}\boldsymbol{\eta}_{v_2\times 1} + X_{|\mathcal{E}_g|\times p}\boldsymbol{\beta}_{p\times 1} + u_{|\mathcal{E}_g|\times 1},$$

  where $\boldsymbol{\alpha} = (\boldsymbol{\mu}', -\boldsymbol{\eta}')'$, and $B = (B_1, -B_2)$.
- There, one can also look at the projection on one side of the graph (eg, firms or teachers).
- This corresponds to a graph g' where two teachers are connected by an edge if there is at least one student who was taught by both. The edge weight will be larger the more students in common there are.


1
2
3
4
5
6
7
8
9
10
11
12
13
14


10
11
12
13
14

The device of a one-mode projection highlights the importance of having movers in panel data. In matched worker-firm data sets, workers do not frequently switch employers over the course of the sampling period. This lack of mobility is one cause of the substantial bias that is observed in the correlation coefficient between (estimated) worker and firm effects (...). While this is now well recognized, limited mobility has broader consequences. (...) Therefore, the induced graph may be only weakly connected (and λ_2 will be close to zero) and the variance of the estimator of the firm effects may be large. This is not only detrimental for identifying sorting between workers and firms, but, indeed, complicates estimation and inference of the firm effects as well as all their moments, such as their variance. (Jochmans and Weidner [2019], p.1552)

Back

Example 2: Leung [2023]

- This paper studies clustering for econometric inference for settings where observations pertain to a network.
- Here, the parameter of interest is $\theta_0 \in \mathbb{R}^{d_\theta}$ which relates to the following moment condition:

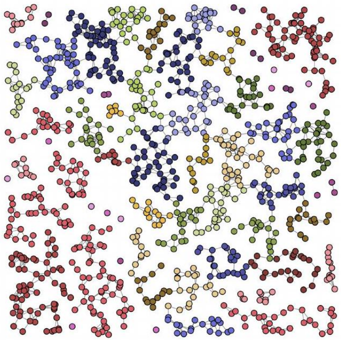
$$\mathbf{E}\left[g\left(W_i, \theta_0\right)\right] = \mathbf{0} \quad \forall i \in \mathcal{N}_g.$$

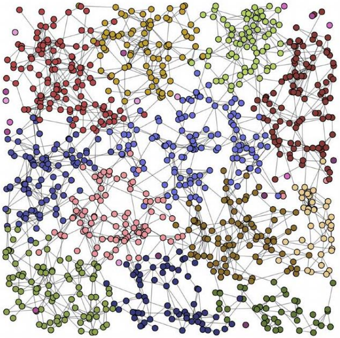
- Estimates can be obtained using the sample analogues $\hat{G}(\theta) = n^{-1} \sum_{i=1}^{n} g\left(W_i, \theta\right)$ through GMM:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \hat{G}(\theta)' \Psi_n \hat{G}(\theta),$$

where Ψ_n is a weighting matrix.

- A clustering scheme partitions the observations into L observation clusters $\{\mathcal{C}_\ell\}_{\ell=1}^{L}$ with $n_\ell = |\mathcal{C}_\ell|$.
- For this to work, one needs the observations to be unrelated or weakly related across clusters.
- Notice that a particular cluster $\mathcal{C}_\ell$ defines a cut of the graph!
- Consequently, one can compute $\phi(\mathcal{C}_\ell)$ and an overall conductance measure for the clustering scheme: $\max_{\ell \in 1,\ldots,L} \phi(\mathcal{C}_\ell)$.





- Let $\hat{\theta}_\ell$ the *GMM* estimator and $\hat{G}_\ell(\theta) = n_\ell^{-1} \sum_{i \in \mathcal{C}_\ell} g\left(W_i, \theta\right)$ be the sample moment vector constructed using only observations in $\mathcal{C}_\ell$.
- Small-L cluster-robust methods use estimates $\left(\hat{\theta}_\ell\right)_{\ell=1,\ldots,L}$ or moments $\left(\hat{G}_\ell\left(\hat{\theta}_\ell\right)\right)_{\ell=1,\ldots,L}$ to construct tests and confidence sets for the parameters of interest.

- Under weak network dependence, the article argues that

$$\frac{1}{\sqrt{n}}\begin{pmatrix} n_1\hat{G}_1(\theta_0) \\ \vdots \\ n_L\hat{G}_L(\theta_0) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}^*\right).$$

- This is an intermediate result for the vector of GMM estimates $\left(\sqrt{n}\left(\hat{\theta}_\ell - \theta_0\right)\right)_{\ell=1}^{L}$ to be asymptotically normal.
- Theorem 1 then establishes that if maximal conductance across clusters (times average degree) goes to zero, the off-diagonal elements in $\mathbf{\Sigma}^*$ go to zero.
- This provides (asymptotic) guarantees for the deployment of a clustering scheme.

- The analysis relies on a generalisation of the Cheeger constant we examined before.
- For any integer $L > 1$, the Lth-order Cheeger constant of g is given by:

$$\phi_L(g) = \min \left\{ \max_{1 \leqslant \ell \leqslant L} \phi(\mathcal{C}_\ell) : \{\mathcal{C}_\ell\}_{\ell=1}^{L} \text{ partitions } \mathcal{N}_g \right\}$$

- This is the lowest possible maximal conductance over all possible partitions of size L. As before,

$$\frac{\lambda_L}{2} \leqslant \phi_L(g) \leqslant C\lambda_L^{1/2}$$

where $0 = \lambda_{\min} \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_L \leqslant \cdots \leqslant \lambda_{\max} \leqslant 2$ are the eigenvalues of the (normalised) Laplacian matrix.

- Based on this, the article offers a few recommendations:
  1. (Conductance): Given a candidate set of clusters $\{\mathcal{C}_\ell\}_{\ell=1}^{L}$,, compute its maximal conductance. The asymptotic results suggest this should be small compared to *n*. (In simulations, it should be no larger than 0.1 to ensure adequate size control.)
  2. (Laplacian): Based on simulations, select the largest L such that λ_L is at most 0.05 to ensure that clusters have sufficiently small conductance. ($L > 5$ appears to ensure sufficient power.)
  3. (Computing clusters): Given L, ideally select the partition that minimizes conductance. An exact solution is computationally infeasible, alternatives such as spectral clustering can be employed.

Spectral clustering algorithms work like this (see von Luxburg [2007] and von Luxburg et al. [2008]):

1. Given a graph g and desired number of clusters L, compute the (unnormalised) Laplacian and its eigenvalues $0 = \lambda_{\min} \leqslant \ldots \leqslant \lambda_{\max}$.
2. Let V_ℓ be the eigenvector associated with λ_ℓ and $V_{\ell i}$ its i th component. Embed the n units in $\mathbb{R}^L$ by associating each unit i with a position

$$X_i = \left(\frac{V_{1i}}{\left(\sum_{\ell=1}^{L} V_{\ell i}^2\right)^{1/2}}, \ldots, \frac{V_{Li}}{\left(\sum_{\ell=1}^{L} V_{\ell i}^2\right)^{1/2}} \right)$$

3. Cluster the positions $\{X_i\}_{i=1}^n$ using k-means with $k = L$ to obtain $\mathcal{C}_1, \ldots, \mathcal{C}_L$.

Variations using the normalised Laplacian, are also discussed in the above reference. Back

Graph Limits

- We can characterise the moments of a graph by the frequency at which configurations or subgraphs appear in the sampled graph.
- In doing this, it is important to distinguish partial and induced subgraphs and to note potentially isomorphic graphs (see Graham [2020]):

  The graphs R and S are isomorphic if there exists a structure-maintaining bijection $\varphi : \mathcal{N}(R) \to \mathcal{N}(S)$. (Structure is maintained if the edges and non-edges in R correspond to edges and non-edges in S.)

- Example: $S = \triangle$.

- One can then define the frequency at which a graph S (on p nodes) appears as an induced subgraph in G (on $n > p$ nodes).

- This can be expressed as

$$t_{\text{ind}}(S, G) \equiv = \frac{1}{\binom{n}{p} \mid \text{iso}(S) \mid} \sum_{\mathbf{v}_p \in \mathcal{C}_{p,n})} \mathbf{1}(S \cong G[\mathbf{v}_p]),$$

where $\mathbf{v}_p$ are p different nodes in $\mathcal{N}_G$, $G[\mathbf{v}_p]$ is the induced subgraph on those nodes, $\cong$ indicates isomorphism, $\mathcal{C}_{p,n}$ denotes the (unordered) set of possible such p nodes and $|\text{iso}(S)|$ is the number of isomorphisms of S.

- For a generic graph S, one can then obtain that

$$\begin{aligned}\mathbb{E}[t_{\text{ind}}(S,G)] &= \mathbb{E}\left[\prod_{\{i,j\}\in\mathcal{E}(S)} h_0(\xi_i,\xi_j) \prod_{\{i,j\}\in\mathcal{E}(\bar{S})} [1-h_0(\xi_i,\xi_j)]\right] \\ &\equiv t_{\text{ind}}(S,h_0).\end{aligned}$$

- A related notion of subgraph density that often appears in the literature is injective homomorphism density.
- A homomorphism of a graph S into G is an edge preserving map $\mathcal{N}_S \to \mathcal{N}_G$.[2]

[2]Note that non-adjecencies are not necessarily preserved.

- Formally,

$$t_{\text{inj}}(S,G) = \frac{1}{\binom{n}{p} \mid \text{iso}(S) \mid} \sum_{R \subseteq K_n, R \cong S} \mathbf{1}(R \subset G),$$

where K_n is the complete graph on n nodes.
- Example: $S = \wedge$ in $G = \boxslash$.
- As before, we have that under AH:

$$\begin{aligned} \mathbb{E}[t_{\text{inj}}(S,G)] &= \mathbb{E}\left[\prod_{\{i,j\} \in \mathcal{E}(S)} h_0(\xi_i, \xi_j)\right] \\ &\equiv t_{\text{inj}}(S, h_0). \end{aligned}$$

- Related concepts are discussed in Lovasz [2012].

- If G_n is a random graph on n nodes, the frequency measures above (with respect to a particular S) are random variables.

- Diaconis and Janson [2008] study random graph limits through the (probability) limits of t_{ind}, t_{inj} (or variations) across all subgraphs S as n grows larger.

- They also note that limits obtained by such criteria are (infinite) exchangeable random graphs (see their Theorem 5.2) and thus amenable to AH.

- In this case the limiting graph can be expressed using the graphon related to the kernel function in AH. (This limit is unique up to measure-preserving mappings, ie, relabeling. See Section 7 in Diaconis and Janson [2008].)

- In related work, Borgs et al. [2008] analyse related notions of convergence and rely on a particular (pseudo-)metric on graphons known as the cut metric.
- Intuitively, it is defined in terms of the cut distance:

$$d_{\square}(G, G') = \max_{S,T \subset V} \frac{1}{|V|^2} |\mathcal{E}_G(S, T) - \mathcal{E}_{G'}(S, T)|$$

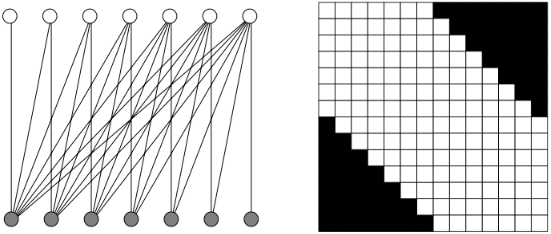
where $V = \mathcal{V}_G = \mathcal{V}_{G'}$ and S, T form a partition for V. $\mathcal{E}_G(S, T)$ is the count of edges between S and T. The cut distance minimises the above across all isomorphic graphs:

$$\delta_{\square}(G, G') = \min_{\tilde{G} \cong G} d_{\square}(G, G').$$

- A recent set of lectures on this topic can be found here: Lecture 1, Lecture 2 and Lecture 3.

- In fact, one can embed a finite graph G_n in $[0,1] \times [0,1]$ by defining:

$$h_{G_n}(u,v) = \begin{cases} 1 & \text{if } (\lceil un \rceil, \lceil vn \rceil) \in \mathcal{E}(G_n) \\ 0 & \text{otherwise} \end{cases}.$$



- With this in hand, the cut metric can be used to compare any simple graph.

- While one can evaluate the limit for finite graphs using the framework above, it is important to keep in mind that sparse random graphs will converge to a trivial graphons (ie., the zero graphon).

- Intuitively, take the average degree, λ, among n agents to be a small positive constant (indep on n). The prob of an edge between the two indep random draws from this population is

$$\Pr\left(W_{12} = 1\right) = \frac{\frac{1}{2}\lambda n}{\begin{pmatrix} n \\ 2 \end{pmatrix}} \approx \frac{\lambda}{n}.$$

"Exchangeable graphs are not sparse. If a random graph is exchangeable, it is either dense or empty." (Orbanz and Roy [2015], Fact 7.2).

- One possible way of accommodating sparsity is to allow for drifting parameters. For example,

$$\Pr\left(W_{ij} = 1 \mid \xi_i = u, \xi_j = v\right) = \rho_n h(u, v),$$

(see Bickel and Chen [2009], Bickel, Chen and Levina [2011], Olhede and Wolfe [2013]). The rate at which $\rho_n \to 0$ then controls the rate of the average degree growth as n grows large.

- This relates to earlier work by Bollobas and Riordan [2009] who rescale the graph metrics but in doing so assume that there are “no dense spots”.
- Other recent works relax those conditions (see, eg, Borgs et al. [2019]) but still retain other features (eg, unbounded degrees).

Back

Mele [2017, 2020]

- Directed network: $W_{ij} = 1$ if $i \to j$ and $= 0$, otherwise.

- The utility function for individual i is given by:

$$\sum_{j \neq i} W_{ij} u_{ij}^{\theta} + \sum_{j \neq i} W_{ij} W_{ji} m_{ij}^{\theta} + \sum_{j \neq i} W_{ij} \sum_{k \neq i,j} W_{jk} \nu_{ik}^{\theta} + \sum_{j \neq i} W_{ij} \sum_{k \neq i,j} W_{ki} p_{kj}^{\theta}$$

$u_{ij}^{\theta} \equiv u(X_i, X_j; \theta)$: direct utility

$m_{ij}^{\theta} \equiv m(X_i, X_j; \theta)$: mutual link

$\nu_{ij}^{\theta} \equiv \nu(X_i, X_j; \theta)$: friends of friends

$p_{ij}^{\theta} \equiv p(X_i, X_j; \theta)$: popularity.

When an agent forms a link, he/she automatically creates an indirect link for other agents that are connected to him/her, thus generating externalities and impacting his/her 'popularity.'

- Assumption 1: Preferences are such that $m_{ij}^{\theta} = m_{ji}^{\theta}$ and $\nu_{ij}^{\theta} = p_{ij}^{\theta}$.
  "... i internalizes the externality he creates ..."

⇒ The deterministic components of utitility are summarised by a potential function:

$$Q(W, X; \theta) = \sum_{(i,j)} W_{ij} u_{ij}^{\theta} + \sum_{(i,j)} W_{ij} W_{ji} m_{ij}^{\theta} + \sum_{(i,j,k)} W_{ij} W_{jk} \nu_{ik}^{\theta}$$

- Maxima for this function correspond to Nash equilibria of the game with payoffs as defined previously.

- Network formation process: stochastic best-response dynamics (Blume [1993]). There is a meeting sequence $m = \{m^t\}_{t=1}^{\infty}$ where $m^t = (i, j)$ (i plays) and
$$\mathbb{P}(m^t = ij | W^{t-1}, X) = \rho(W^{t-1}, X_i, X_j).$$
- Assumption 2: $\rho(W^{t-1}, X_i, X_j) = \rho(W^{t-1}_{-ij}, X_i, X_j) > 0, \forall ij$.
The meeting probability does not depend on the existence of a link between them and each meeting has positive probabiility of occuring.

$\Rightarrow$ likelihood does not depend on ρ.

- Conditional on the meeting ij, i receives an idiosyncratic shock $\epsilon \sim F_\epsilon$ and W^t_{-ij} iff

$$U_i(W^t_{ij} = 1, W^t_{-ij}, X; \theta) + \epsilon_{1t} \geqslant U_i(W^t_{ij} = 0, W^t_{-ij}, X; \theta) + \epsilon_{0t}$$

$\Rightarrow$ Markov chain of networks.

$\Rightarrow$ Absent shocks (i.e., $F_\epsilon = \mathbf{1}(0 \leqslant \cdot)$), chain converges to one of the NE with probability one.

- Assumption 3: F_ϵ is EV Type I, i.i.d. among links and across time.

- Under Assumptions 1-3, the Markov chain above converges to a unique stationary distribution

$$\pi(w, X; \theta) = \frac{\exp[Q(w, X; \theta)]}{\sum_{\omega \in \mathcal{G}} \exp[Q(\omega, X; \theta)]}$$

  where Q is the potential function previously defined.
- Appears to bypass issues of multiplicity, but in the long-run the chain spends more time around networks with high potential (NE of the game without shocks).
- The model generates dense networks: as $n \to \infty$, the unconditional probability of a link does not decrease.

- If the utility functions are linear in parameters ($Q(w, X; \theta) = \theta^\top \mathbf{t}(w, X)$), the stationary distribution $\pi(w, X; \theta)$ describes an exponential random graph (ERGM) $\in$ exponential family.
- The normalizing constant $\sum_{\omega \in \mathcal{G}} \exp[Q(\omega, X; \theta)] \equiv \exp(A(\theta))$ is an important computational obstacle.
- $n = 10 \Rightarrow 2^{90} \approx 10^{27}$ network configurations.

  "A supercomputer that can compute 10^{12} potential functions in one second would take almost 40 million years to compute the constant."
- Mele [2017, 2020] relies on a Metropolis-Hastings algorithm.

- Metropolis-Hastings for Network Simulations.

> Fix a parameter θ. At iteration r, with current network w_r:

1. Propose a network w' from a proposal distribution $w' \sim q_w(w'|w_r)$.
2. Accept network w' with probability

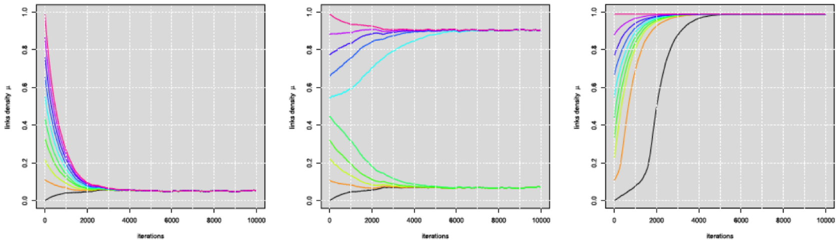
$$\alpha_{mh}(w_r, w') = \min\left\{1, \frac{\exp[Q(w', X; \theta)]}{\exp[Q(w_r, X; \theta)]} \frac{q_w(w_r|w')}{q_w(w'|w_r)}\right\}$$

$\Rightarrow$ The network transitions from w_r to w' with probability $T(w_r, w') = q_w(w'|w_r)\alpha_{mh}(w_r, w')$ and $\alpha_{mh}(w_r, w')$ guarantees "detailed balance":

$$\begin{aligned} \pi(w_r)T(w_r, w') &= \pi(w_r)q_w(w'|w_r)\alpha_{mh}(w_r, w') \\ &= \min\{\pi(w_r)q_w(w'|w_r), \pi(w')q_w(w_r|w')\} \\ &= \pi(w')q_w(w_r|w')\alpha_{mh}(w', w_r) = \pi(w')T(w', w_r) \end{aligned}$$

- It does not depend on the normalising constant.

In Mele's model, with $u^\theta = \alpha$ and $m^\theta = \beta$, we have:



(A) (B) (C)

FIGURE 1.—Network simulations at different parameter values. Traceplots of simulations of model (9) using Algorithm 1 with local chains. The simulations are obtained for a network with $n = 100$ players, with parameters $\alpha = -3$ and $\beta = \{1/n, 3/n, 7/n\}$ (Panel (A), (B), and (C), respectively). Each simulation is started at 10 different starting networks, each corresponding to a directed Erdős–Rényi network with probability of link $\mu = \{0, 0.111, 0.222, 0.333, 0.444, 0.555, 0.666, 0.777, 0.888, 1\}$.

- While these illustrate the “degeneracy” issue referred to above in (1.), Mele [2017] also establishes the existence of parameter regions where convergence is slow (for a given parametrisation) (Theorem 5).

  In this region, “once the sampler reaches a local maximum, there is probablity $\exp(-Cn^2)$ to escape such state of the network. As a consequence, the sampling is practically infeasible with a local sampler.”
- An algorithm with larger step sizes (Appendix B) is also shown to help with these issues.

- Likewise, he also demonstrates that when $\beta \geqslant 0$, the model is asymptotically (in *n*) indistinguishable from a directed Erdos-Renyi model or a mixture of such models (Theorem 2), but not when $\beta < 0$ (Theorem 3). (Similar results when include externality on cyclic triangles, Theorem 4.)

$\Rightarrow$ These problems are related to the existence of multiple NE in the game without shocks.

- Multiple networks: identification can be attained with variation in sufficient statistics (across networks).
  "If the sufficient statistics are not linearly dependent, then the exponential family is minimal and the likelihood is stricly concave, therefore the mode is unique." [Mele, 2017]

Mele [2020] studies racial segregation using this model and AddHealth data (14-16 schools).

Table 1: Descriptive Statistics for the schools in the Saturated Sample

School	1	2	3	7	8	28	58	77	81	88	106	115	126	175	194	369
Students	44	60	117	159	110	150	811	1664	98	90	81	20	53	52	43	52
Links	12	120	125	344	239	355	3290	3604	163	308	162	44	123	171	42	48
Females	0.5	0.517	0.419	0.44	0.5	0.587	0.473	0.483	0.531	0.522	0.531	0.55	0.491	0.538	0.512	0.654
Clustering	0.000	0.421	0.154	0.222	0.282	0.291	0.197	0.193	0.244	0.362	0.202	0.393	0.392	0.284	0.064	0.056
Density	0.006	0.034	0.009	0.014	0.020	0.016	0.005	0.001	0.017	0.038	0.024	0.116	0.045	0.064	0.023	0.018
A. Racial Composition																
Whites	0.5	0.95	0.983	0.981	0.973	0.42	0.978	0.055	0.98	0.989	0	1	0.472	0.769	0.977	0.942
Blacks	0.136	0	0	0.006	0.018	0.453	0.002	0.233	0	0	0.963	0	0.151	0.019	0	0
Asians	0	0	0	0	0.009	0.007	0.005	0.299	0.01	0	0	0	0.038	0.038	0	0
Hispanics	0.364	0.05	0.017	0.006	0	0.107	0.011	0.392	0.01	0	0.025	0	0.302	0.154	0.023	0.058
Others	0	0	0	0	0	0.013	0.004	0.02	0	0.011	0	0	0.038	0.019	0	0
Racial Fragm	0.599	0.095	0.034	0.037	0.053	0.606	0.044	0.699	0.04	0.022	0.072	0	0.661	0.382	0.045	0.109
B. Grade Composition																
7th Grade	0.159	0.2	0.128	0.145	0.227	0.173	0.002	0.001	0.112	0.144	0.506	0.4	0.491	0.462	0.488	0.538
8th Grade	0.159	0.217	0.154	0.157	0.2	0.173	0.004	0.003	0.153	0.178	0.481	0.6	0.472	0.538	0.488	0.462
9th Grade	0.114	0.2	0.12	0.214	0.136	0.2	0.289	0.004	0.153	0.122	0.012	0	0.038	0	0	0
10th Grade	0.273	0.133	0.205	0.157	0.182	0.167	0.277	0.346	0.214	0.167	0	0	0	0	0	0
11th Grade	0.136	0.167	0.179	0.164	0.118	0.14	0.223	0.345	0.265	0.211	0	0	0	0	0.023	0
12th Grade	0.159	0.083	0.214	0.164	0.136	0.147	0.205	0.301	0.102	0.178	0	0	0	0	0	0
C. Segregation																
FSI gender	0.348	0.035	0.095	0.263	0.100	0.206	0.142	0.228	0.196	0.107	0.186	0.123	0.095	0.050	0.186	0.000
FSI race	0.000	0.689	0.180	0.553	0.000	0.671	0.014	0.690	0.819	0.816	0.000		0.403	0.000	0.000	0.560
FSI income 90	0.596	0.332	0.000	0.189	0.000	0.118	0.024	0.000	0.000	0.077	0.000	0.000	0.000	0.272	0.000	0.384
FSI income 50	0.023	0.000	0.027	0.133	0.000	0.013	0.077	0.082	0.064	0.116	0.012	0.000	0.269	0.069	0.131	0.000
SSI gender	0.305	0.541	0.493	0.697	0.586	0.659	0.798	0.727	0.614	0.618	0.601	0.658	0.561	0.696	0.488	0.461
SSI race	0.146	0.862	0.791	0.894	0.865	0.754	0.927	0.748	0.805	0.921	0.761		0.550	0.632	0.786	0.817
SSI income 90	0.409	0.767	0.641	0.783	0.735	0.803	0.820	0.767	0.684	0.836	0.705	0.778	0.726	0.825	0.709	0.681
SSI income 50	0.214	0.469	0.444	0.601	0.501	0.547	0.726	0.620	0.439	0.563	0.460	0.419	0.535	0.611	0.484	0.421

Table 2: Posterior mean of estimated models

	(1)	(2)	(3)	(4)	(5)	(6)
			A. Direct utility (u_{ij})			
CONSTANT	-6.9201	-5.5381	-6.6500	-5.9132	-7.2182	-5.8070
MALE i			-0.1517	0.0463	-0.2718	0.2350
WHITE i			-0.1710	0.0044[a]	0.0440[a]	0.3023
BLACK i			1.0451	1.1310	0.7074	1.1801
HISP i			2.0990	2.2806	1.4590	2.0295
INCOME i (logs)			-2.0543	-1.6492	-1.8738	-1.4645
SAME GENDER	-0.4545	0.1850	0.2067	0.4851	0.3154	0.7644
SAME GRADE	2.3124	2.2384	2.3817	2.0113	2.5185	2.1800
WHITE-WHITE	0.3504	0.5414	1.0138	0.5720	0.9959	0.2739
BLACK-BLACK	0.1443	0.3660	1.6491	1.1445	1.5347	0.9405
HISP-HISP	1.8597	1.6794	0.3186	-0.2269	0.7130	-0.1394
ATTRACTIVE i (Phys)	0.2757	0.3068	-2.3568	-2.2413	-1.9291	-1.9430
ATTRACTIVE j (Phys)	-0.0410	0.2322	2.5166	1.5861	2.7615	1.2609
ATTRACTIVE i (Pers)	-0.4402	0.0063[a]	-0.4964	-0.1570	-0.8646	-0.1631
ATTRACTIVE j (Pers)	1.0672	0.8678	-1.0932	-0.7390	-0.6361	-0.3939
INCOME i - INCOME j (logs)	0.1793	0.1462	0.8883	0.9012	0.9938	0.7403
INCOME i + INCOME j (logs)	-0.0882	-0.0806	1.0947	0.9244	0.8977	0.6892
SHARE WHITES	0.9070	-0.4814	-1.7088	-1.4420	-1.5748	-1.6126
SHARE BLACKS	3.2238	3.0985	1.3416	1.8309	0.7645	1.9618
SHARE HISP	2.524	2.444	0.8397	0.7798	1.0078	0.7731
WHITE-WHITE * SHARE WHITES	1.3962	1.0094	4.3915	2.7840	4.7269	2.3272
BLACK-BLACK * SHARE BLACKS	0.4664	0.1478	0.2528	0.4028	0.1172	0.2516
HISP-HISP * SHARE HISP	-1.5643	-1.4255	-1.6908	-1.3630	-1.3872	-1.1400
			B. Mutual utility (m_{ij})			
CONSTANT		1.1853		6.1668		5.3139
SAME GENDER		1.1652		1.0716		1.1539
SAME GRADE		-1.6882		-3.0514		-3.0575
WHITE-WHITE		0.0073[a]		-0.6017		-0.4960
BLACK-BLACK		0.7468		1.1177		0.7067
HISP-HISP		0.7779		-1.4659		-1.4639
			C. Indirect utility and Popularity (v_{ij})			
CONSTANT		-0.2891		-0.4705		-0.4308
SAME GENDER		0.1721		-0.4074		-0.3987
SAME GRADE		-0.3145		0.1136		0.3266
WHITE-WHITE		0.2239		0.1856		0.2978
BLACK-BLACK		-0.1364		0.1372		0.1202
HISP-HISP		0.4328		-0.5067		-0.2859
SCHOOL DUMMIES	YES	YES	YES	YES	YES	YES
			D. Sample size			
# Schools	14	14	14	14	16	16
# Students	1129	1129	1129	1129	3604	3604
# Pairs/Dyads	112,751	112,751	112,751	112,751	3,536,893	3,536,893

Models (1)-(4): posterior sample of 100,000 parameter and 5000 network simulations per parameter. Models (5)-(6): posterior sample of 20,000 parameter and 10,000 network simulations per parameter. [a] credible 95% interval contains both positive and negative values.

Back