Econ 7217 Economic Analysis of Social Networks Introduction and characterization of social networks

Chih-Sheng Hsieh

Department of Economics
National Taiwan University

February 19, 2024

Significance of Network Studies

- Network studies are of multidisciplinary nature. Well-known studies were covered by news media and published in top academic journals of various disciplines, e.g., *Nature*, *Science*, *Journal of American Statistical Association*, *American Economic Review*, *Social Networks*, *Journal of Finance*, *Journal of Financial Economics*, *Journal of Marketing Research*, *Management Science*, etc.
- Economic applications of networks are now the research focus of several groups leaded by well-known economists such as Daron Acemoglu (MIT), Esther Duflo (MIT), James Heckman (U Chicago), Steven Durlauf (Chicago), Guido Imbens (Standford), Matthew Jackson (Standford), Yves Zenou (Monash), etc.
- Network study is part of the frontier "Big Data Analytics" due to its complexity and variety (Pržulj and Malod-Dognin, 2016).

Economic applications of social networks

- Researchers have long believed that social networks are important in influencing economic behaviors, e.g., word-of-mouth communication on affecting economic decisions (Katz and Paul, 1955); the spread of (job) information (Myers et al., 1951), and providing a form of social capital (Coleman, 1988).
- Two questions: how do network structures affect economic outcomes? how do networks form?
- The first question ties closely to the economic literature on social interactions or network effects (Blume et al., 2010; Jackson, 2010; Jackson et al., 2017).
- The second question opens the research pathway on studying network formation (Goldenberg et al., 2010; De Paula, 2016).
- A growing interest on answering these two questions jointly (Badev, 2013; Goldsmith-Pinkham and Imbens, 2013; Hsieh and Lee, 2016; Boucher, 2016; Hsieh et al., 2019).

Economic applications of networks

A selected list of network studies in the economic literature:

- job finding and labor force participation (Calvo-Armengol and Jackson, 2004; Calvó-Armengol and Jackson, 2007; Bayer et al., 2008);
- social learning and knowledge diffusion (Conley and Udry, 2001, 2010);
- risk sharing and insurance in rural villages (Fafchamps and Gubert, 2007a,b);
- obesity and happiness transmission (Christakis and Fowler, 2007; Fowler and Christakis, 2008) ;
- program participation (Duflo and Saez, 2003; Dahl et al., 2014);
- peer effects on students' academic achievement (Calvó-Armengol et al., 2009; Lin, 2010);

Economic applications of networks

- peer effects on smoking and drinking behaviors (Clark and Lohéac, 2007; Kremer and Levy, 2008)
- peer effects on students' sport club participation and sleeping (Bramoullé et al., 2009; Liu et al., 2017);
- juvenile delinquencies or criminal activities (Ballester et al., 2010; Patacchini and Zenou, 2009; Bayer et al., 2009);
- homophily and segregation (Currarini et al., 2009; Boucher, 2015).

This is a short and incomplete list. Network studies on macroeconomics, international economics, banking and finance are out of this list. There is a huge literature on the structure of financial networks and financial stability in the 2007-2008 financial crisis. There are also network studies on COVID pandemics.

Topics to be covered in this lecture

- Presentation of relational (network) data
  - directed versus undirected networks
  - binary versus valued networks
  - socio-centric (complete) network versus ego-centric (Ego) networks
  - one-mode (unipartite), two-mode (bipartite) and multipartite networks
- Descriptive network statistics
  - path, walk, cycle, geodesic distance, diameter, and average path length
  - (giant) components
  - individual centrality - degree, betweenness, closeness, eigenvector, Bonacich centralities
  - global assortativity and clustering coefficient
- Application: How central are clients in sexual networks created by commercial sex?

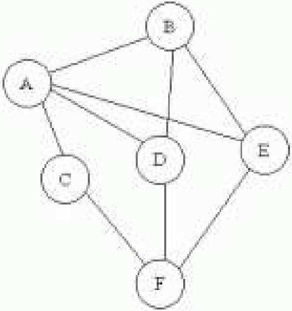
Presentation of relational (network) data

- Where do you see networks: communication network (telephone networks, Internet); transportation network (airline routes, railroad networks); social network (friendship networks, sexual contacts); coauthor or citation networks; financial networks (borrowing and lending); interlock networks of corporate boards; biological networks; trading networks.
- Some public assessable social network datasets:
  Stanford Large Network Dataset Collection
  Social Networks and Microfinance
  Siena Homepage
  Book "Statistical Analysis of Network Data with R"
  UCINET

Presentation of relational (network) data

- A network (graph) can be expressed as $G = (V, E)$ which consists of a set V of vertices (nodes) and a set E of edges (links).
- Network data can also be represented by either an adjacency (socio) matrix or a graph.

	A	B	C	D	E	F
A	-	1	1	1	1	
B	1	-		1	1	
C	1		-			1
D	1	1		-		1
E	1	1			-	1
F			1	1	1	-



Presentation of relational (network) data

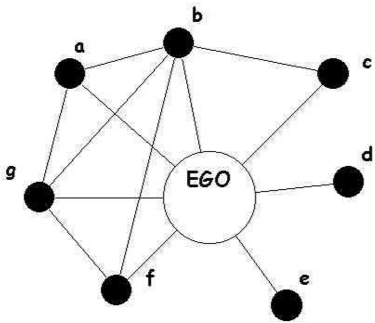
Understanding your network data:

- Are links in your network directed (*directed graph*, *digraph*) or undirected? For example, communication networks such as telephone, email, twitter, etc. should be directed. Trade networks could be directed or undirected. Sexual networks are not directed.
- Are links in your network valued or unvalued? For example, values on links may represent strength of relationships, contact frequency, closeness, etc.

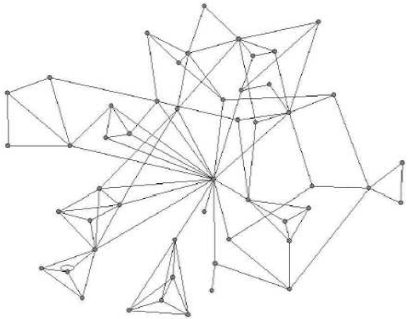
Presentation of relational (network) data

Understanding your network data:

- Are your networks ego-centric or socio-centric (complete)?



(a) Ego network



(b) Complete network

Presentation of relational (network) data

Understanding your network data:

- Are links in your network belonging to one mode (unipartite), two mode (bipartite), or multiple mode (multipartite)?
- In two mode networks, $G = (V_1, V_2, E)$, nodes of the same type (mode) cannot connect with each other.
- Sources of two mode networks:
  - affiliations: attendance at events, memberships in organizations.
  - correspondences: authors & topics, illnesses and treatments.
  - sellers and buyers.
  - sexual transactions by clients and escorts.
- Multiple mode networks, $G = (V_1, V_2, \cdots, V_k, E)$.

Presentation of relational (network) data

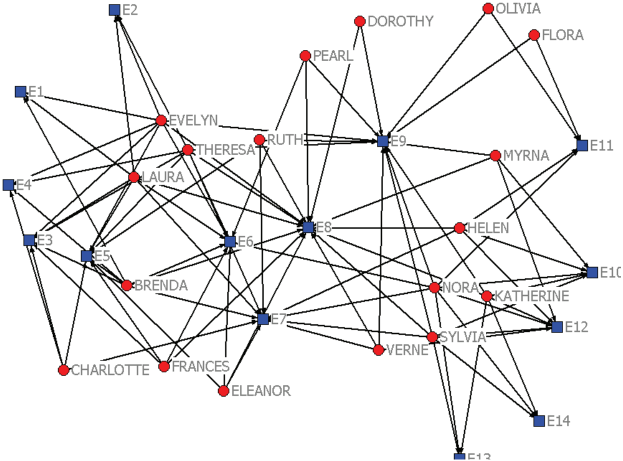
The classic example

Names of Participants of Group I	Code Numbers and Dates of Social Events Reported in *Old City Herald*													
	(1) 6/27	(2) 3/2	(3) 4/12	(4) 9/26	(5) 2/25	(6) 5/19	(7) 3/15	(8) 9/16	(9) 4/8	(10) 6/10	(11) 2/23	(12) 4/7	(13) 11/21	(14) 8/3
1. Mrs. Evelyn Jefferson	×	×	×	×	×	×		×	×					
2. Miss Laura Mandeville	×	×	×		×	×	×	×						
3. Miss Theresa Anderson		×	×	×	×	×	×	×	×					
4. Miss Brenda Rogers	×		×	×	×	×	×	×						
5. Miss Charlotte McDowd			×	×	×		×							
6. Miss Frances Anderson			×		×	×		×						
7. Miss Eleanor Nye					×	×	×	×						
8. Miss Pearl Oglethorpe						×		×	×					
9. Miss Ruth DeSand					×		×	×	×					
10. Miss Verne Sanderson							×	×	×			×		
11. Miss Myra Liddell								×	×	×		×		
12. Miss Katherine Rogers								×	×	×		×	×	×
13. Mrs. Sylvia Avondale							×	×	×	×		×	×	×
14. Mrs. Nora Fayette						×	×		×	×	×	×	×	×
15. Mrs. Helen Lloyd							×	×		×	×	×		
16. Mrs. Dorothy Murchison								×	×					
17. Mrs. Olivia Carleton									×		×			
18. Mrs. Flora Price									×		×			

Figure 1. Davis, Gardner and Gardner (1941) *Deep South* women-by-events matrix.

Presentation of relational (network) data

Canonical visualization

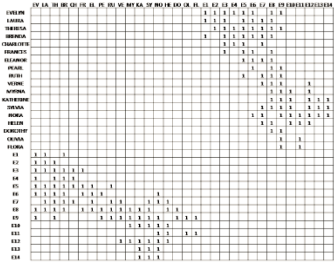


Presentation of relational (network) data

Two mode network data can have an adjacency matrix presentation.

	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14
EVELYN	1	1	1	1	1	1	0	1	1	0	0	0	0	0
LAURA	1	1	1	0	1	1	1	1	0	0	0	0	0	0
THERESA	0	1	1	1	1	1	1	1	1	0	0	0	0	0
BRENDA	1	0	1	1	1	1	1	1	0	0	0	0	0	0
CHARLOTTE	0	0	1	1	1	0	1	0	0	0	0	0	0	0
FRANCES	0	0	1	0	1	1	0	1	0	0	0	0	0	0
ELEANOR	0	0	0	0	1	1	1	1	0	0	0	0	0	0
PEARL	0	0	0	0	0	1	0	1	1	0	0	0	0	0
RUTH	0	0	0	0	1	0	1	1	1	0	0	0	0	0
VERNE	0	0	0	0	0	0	1	1	1	0	0	1	0	0
MYRNA	0	0	0	0	0	0	0	1	1	1	0	1	0	0
KATHERINE	0	0	0	0	0	0	0	1	1	1	0	1	1	1
SYLVIA	0	0	0	0	0	0	1	1	1	1	0	1	1	1
NORA	0	0	0	0	0	1	1	0	1	1	1	1	1	1
HELEN	0	0	0	0	0	0	1	1	0	1	1	1	0	0
DOROTHY	0	0	0	0	0	0	0	1	1	0	0	0	0	0
OLIVIA	0	0	0	0	0	0	0	0	1	0	1	0	0	0
FLORA	0	0	0	0	0	0	0	0	1	0	1	0	0	0

incidence matrix



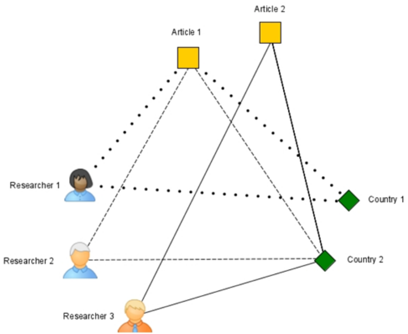
bi-adjacency matrix

Transform two mode network to one mode network in excel

https://www.youtube.com/watch?v=w9NG7C6O9mg

Presentation of relational (network) data

- Multipartite graphs (hypergraphs) consider that the vertices can be partitioned into k independent sets, and the edges can be used to describe more-than-dyadic situations involving either three or more actors, characteristics of actors or situations, and time and place.
- By comparison, unipartite or bipartite graphs only allow the relationships (edges) between vertices to be dyadic.



Presentation of relational (network) data

Handy network analysis and visualization software:

- Pajek: Program for large network analysis
- UCINET: comprehensive social network analysis software
- (R)SNA: social network analysis tools
- (R)igraph: network analysis tool used in R, Python, and C.
- PNet: Estimating exponential random graph model
- (R)SIENA: Statistical network analysis
- (R)statnet: R package for statistical analysis
- NetDraw: network drawing associated with UCINET
- Gephi: network graphing tool for dynamic and hierarchical networks

Descriptive Network Statistics

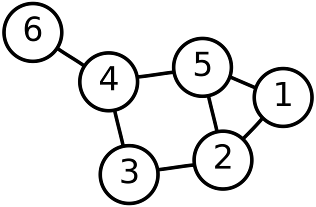
- Given a complex network, we need statistics to characterize the properties of nodes, links, and the structure of the network.
- Some network statistics have microfoundations, e.g., In Ballester et al. (2006), they show that individual behaviors are proportional to their Bonacich centrality in a network. Also, a denser and larger network would typically increase aggregated economic outcomes.
- Policy experiments in networks: target on changing certain network statistics to improve economic outcomes. For example, Mele (2013) simulate the policy of desegregation busing to examine the change of the network clustering coefficient with respect to the change of network composition.
- Jackson et al. (2017) provide a survey on the economic consequence of network structures. We will review some applications which use network statistics as explanatory variables in the next lecture.

Descriptive Network Statistics – definitions

- **Path**: A path in a network $g \in G(N)$ between nodes i and j is a sequence of links $i_1 i_2, i_2 i_3, \cdots, i_{K-1} i_K$ such that $i_k i_{k+1} \in g$ for each $k \in \{1, \cdots, K-1\}$ with $i_1 = i$ and $i_K = j$, and such each node in the sequence $i_1, \cdots, i_K$ is distinct.
- **Walk**: Walk is similar to Path except each node in the sequence is not required to be distinct, i.e., a walk can come back to a given node more than once.
- **Cycle**: A cycle is a walk that starts and ends at the same node, with all other nodes appearing only once.
- it is not trivial to find all k-cycles, $k \geq 2$ and even, in undirected graphs. See discussion in Yuster and Zwick (1997).

Descriptive Network Statistics – definitions

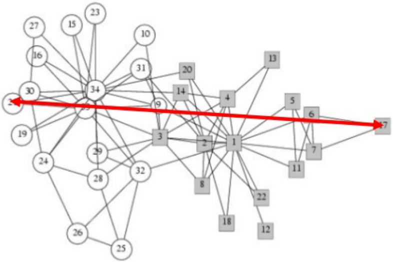
- **Geodesic distance**: a shortest path between two nodes i and j.
- **Nodal eccentricity**: the
  eccentricity of a node is the largest distance from it to any other node.



- In the above figure, the geodesic distance between node 6 and node 1 is 3 and the geodesic distance between node 4 and node 1 is 2.
- $e = (3, 3, 2, 2, 2, 3)$

Descriptive Network Statistics – definitions

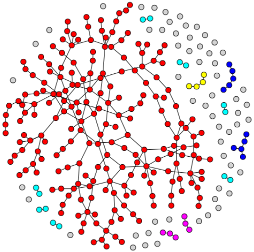
- $W^n = W \times W \times \cdots W$ tells how many walks there are of length n between any two nodes in an adjacency matrix W.
- The **diameter** of a network is the largest distance between any two nodes in the network.
- The **average path length** is taken over geodesics. It is surely bounded by the diameter.
- Comparing diameter with the average path length can reveal whether the diameter is determined by a few outliers or not.



Descriptive Network Statistics – definitions

Component

- A network (V, E) is connected if for each $i \in V$ and $j \in V$ there exists a path in (V, E) between i and j.
- A component of a network (V, E) is a nonempty subnetwork (V', E') $(\emptyset \neq V' \in V, E' \in E)$ such that i) (V', E') is connected and ii) if $i \in V'$ and $ij \in E$, then $j \in V'$ and $ij \in E'$.
- An emergence of a giant component is a necessary condition for epidemic.



Descriptive Network Statistics – degree

Degree: The degree of a node is the number of links that involve that node, which is the cardinality of the node's neighborhood.

- Indegree: $d_{i,in}(w) = \#\{j : w_{ji} = 1\}$.
- Outdegree: $d_{i,out}(w) = \#\{j : w_{ij} = 1\}$.
- If links are undirected, indegree=outdegree.
- Network density: $\sum_{i=1}^{n} d_{i,out}/(n(n-1))$.
- Degree is one of the centrality measure.

Descriptive Network Statistics – centrality

Degree centrality: $d_i(w)/(n-1)$.

- measure how connected a node is.
- however, it does not measure how well located a node is in a network.

Closeness Centrality: $(n-1)/\sum_{j\neq i}\ell(i,j)$, where $\ell(i,j)$ is the number of links in the shortest path (i.e., geodesic) between i and j.

- measure how close a given node is to any other node.

Descriptive Network Statistics – centrality

Betweenness centrality: $\frac{\sum_{k\neq j: i\notin\{k,j\}} P_i(kj)/P(kj)}{(n-1)(n-2)/2}$, where $P(kj)$ is the total number of geodesics between k and j and $P_i(kj)$ denotes the number of geodesics between k and j that i lies on.

- measure how well situated a node is in terms of the paths that it lies on.

Eigenvector centrality: Let $C^e(w)$ denote the eigenvector centrality associated with network matrix w. The centrality of a node is proportional to the sum of the centrality of its neighborhoods: $\lambda C_i^e(w) = \sum_j w_{ij} C_j^e(w)$. In a matrix form, $\lambda C^e(w) = wC^e(w)$. Conventionally we choose λ as the largest eigenvalue.

- The more central the neighborhoods of a node are, the more central that node itself is.
- A scalar λ is called an eigenvalue of the $n \times n$ matrix A if there is a nontrivial solution x of $Ax = \lambda x$. Such an x is called an eigenvector corresponding to the eigenvalue λ.

Descriptive Network Statistics – centrality

Bonacich centrality: $C^B(w) = (I - \lambda w)^{-1}\lambda w \ell$, where I is a $n \times n$ identity matrix and ℓ is a n vector of ones.

- $(I - \lambda w)^{-1}\lambda w \ell = (1 + \lambda w + \lambda^2 w^2 + \cdots)\lambda w \ell = \lambda w \ell + \lambda^2 w^2 \ell + \cdots$.
- the centrality of a node is a weighted sum of the walks that emanate from it.
- It is proportional to individual economic behavior in a network (Ballester et al., 2006).
- It is relevant to estimating the peer effect from the social interactions model (Lee et al., 2010). We will discuss it later in the class.

Descriptive Network Statistics – centrality

PageRank centrality: $PR(i) = \frac{1-d}{N} + d \sum_{j \in M(i)} \frac{PR(j)}{L(j)}$, where $PR(i)$ denotes the page rank value of node i. The constant d is a damping parameter and usually set at 0.85. $L(j)$ is the outdegree of node j. and $M(i)$ is the set of nodes that send link to i.

- The value of $PR(i)$ is between 0 and 1.
- By iteration, the formula will lead to a stable PR which is close to the true theoretical value.
- Page rank is named after Larry Page, the co-founder of Google.
- Page rank was once the main algorithm used by Google to evaluate importance of each webpage.

Descriptive Network Statistics – Assortivity and clustering

Assortivity: measure the correlation between pairs of linked nodes based on a certain characteristics. It is equivalent to the Pearson correlation coefficients for continuous characteristics.

- Positive values indicate a correlation between nodes of similar characteristics, while negative values indicate relationships between nodes of different characteristics. In general, it lies between -1 and 1.

Clustering coefficient:

- global clustering coef. $\frac{\sum_{i;j\neq i;k\neq j;k\neq i} w_{ij}w_{ik}w_{jk}}{\sum_{i;j\neq i;k\neq j;k\neq i} w_{ij}w_{ik}}$. The numerator is the number of triangles. The denominator is the number of two stars.
- individual clustering coef. $\frac{\sum_{j\neq i;k\neq j;k\neq i} w_{ij}w_{ik}w_{jk}}{\sum_{j\neq i;k\neq j;k\neq i} w_{ij}w_{ik}}$.
- one may calculate the average of individual clustering coef. and compare it with the global clustering coef. More weights are given to low-degree nodes in the average clustering coef.

Example: How central are clients in sexual networks created by commercial sex?

- Source: Hsieh et al. (2014) in *Scientific Reports*
- We study the network of male sex workers (MSW) (Logan, 2010; Logan and Shah, 2013).
- The network data we use comes from http://www.daddysreviews.com, a website that has been in existence since 1998 and provides a means for clients to review MSW services (Now it is offline).
- Both MSWs and clients are identified by unique usernames which they use on the website. It allows us to build the network based on the review records.
- Reviews of MSW services provide information on location, so we can analyze multilevel networks – national level and city level.

Example: Hsieh, Kovářík, and Logan (2014)

- Understanding the structure of MSW networks provides us valuable policy suggestions to fight against sexually transmitted diseases (STD).
- Identify the key players and their characteristics.
- The role played by travelers (no matter MSWs or clients) in affecting the structure of the national network.
- Evaluating different immunization strategies on the network structure – random or specific target

Example: How central are clients in sexual networks created by commercial sex?

Table: Statistics in city MSW networks

	Escort	Client	N	Edges	GC	GC(%)	Assor	Thrd
NY - New York City	476	1004	1480	1333	787	0.532	-0.2308	0.2377
CA - Los Angeles/Long Beach/Orange County	246	531	777	668	388	0.499	-0.2631	0.2419
IL - Chicago	120	275	395	323	104	0.263	-0.3016	0.2818
FL - Miami/Fort Lauderdale	103	172	275	193	15	0.055	-0.2348	0.4142
DC - Washington DC	114	209	323	261	86	0.266	-0.2077	0.2816
GA - Atlanta	92	138	230	183	26	0.113	-0.3042	0.2914
MA - Boston	70	127	197	161	57	0.289	-0.2078	0.3009
MI - Detroit/Ann Arbor	14	22	36	25	12	0.333	-0.2755	0.3731
CA - San Francisco/Oakland	190	421	611	500	230	0.376	-0.2781	0.2362
WA - Seattle/Bellevue/Everett	43	71	114	84	24	0.211	-0.2563	0.3500
MN - Minneapolis/St. Paul	32	59	91	71	20	0.220	-0.3087	0.3087
MO - St. Louis	10	16	26	16	6	0.231	-0.2981	0.5517
FL - Tampa/St. Petersburg	22	49	71	52	11	0.155	-0.4071	0.3662
CO - Denver	21	35	56	37	6	0.107	-0.3524	0.5441
OR - Portland	17	48	65	54	31	0.477	-0.3940	0.2022
CA - Sacramento/Yolo/Yuba City	10	22	32	24	9	0.281	-0.4968	0.3529
MO - Kansas City	11	22	33	25	11	0.333	-0.2757	0.3521
OH - Columbus	19	29	48	30	5	0.104	-0.2821	0.6250
IN - Indianapolis	13	27	40	29	8	0.200	-0.4160	0.4028
NC - Charlotte/Gastonia/Rock Hill	11	18	29	18	5	0.172	-0.3487	0.6207
TX - Austin/San Marcos	18	31	49	33	10	0.204	-0.3327	0.3837
TN - Nashville/Davidson	11	30	41	31	14	0.341	-0.4044	0.2480
OK - Oklahoma City	2	2	4	2	2	0.500	NA	1.0000
NY - Buffalo/Niagara Falls	2	3	5	3	3	0.600	-0.5000	0.7500
NY - Rochester	4	5	9	5	3	0.333	-0.2500	0.8333
NY - Albany/Schenectady/Troy	1	1	2	1	2	1.000	NA	1.0000
Nation wide	1778	3900	5678	5817	3965	0.698	-0.119	0.1700

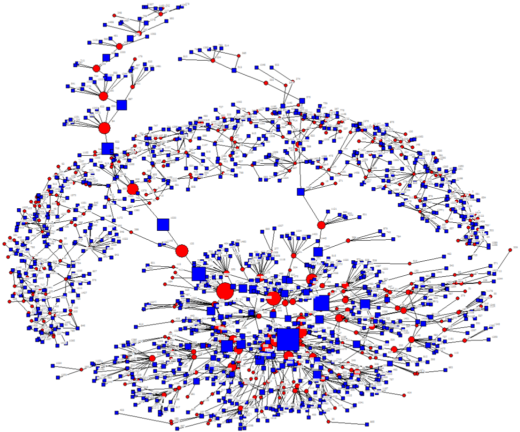


Figure: MSW network – New York

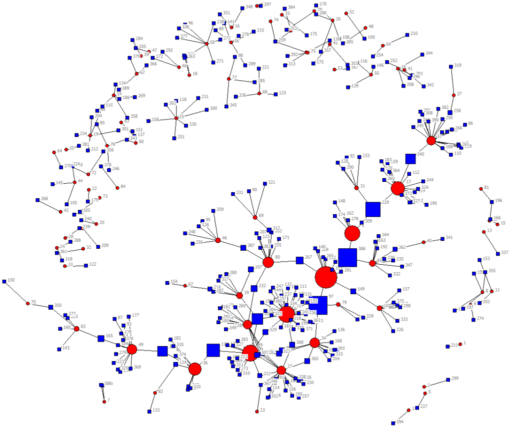


Figure: MSW network – Chicago

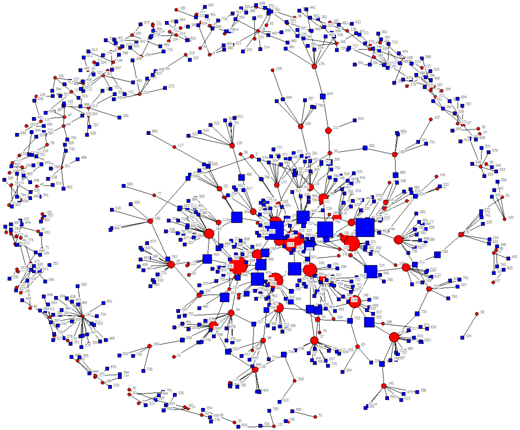


Figure: MSW network – Las Angeles

Example: How central are clients in sexual networks created by commercial sex?

Table: The immunization experiment

Travel	Escort	Client	N	Edges	GC	GC(%)	Assor.	Thrd
Original national data	1778	3900	5678	5817	3965	0.698	-0.119	0.1701
remove top 0.5% (28)	1763	3887	5650	5307	3592	0.636	-0.178	0.2017
remove top 1% (57)	1743	3878	5621	4957	3232	0.575	-0.176	0.2168
remove top 2% (114)	1718	3846	5564	4416	2712	0.487	-0.216	0.2561
remove top 3%(171)	1702	3805	5507	4112	2263	0.411	-0.228	0.2737
remove top 4%(228)	1677	3773	5450	3757	1706	0.313	-0.219	0.3010
remove top 5%(285)	1677	3716	5393	3575	1288	0.239	-0.213	0.3128
Random	Escort	Client	N	Edges	GC	GC(%)	Assor.	Thrd
remove top 0.5% (28)	1768	3882	5650	5748	3920	0.694	-0.124	0.1710
remove top 1% (57)	1760	3861	5621	5682	3878	0.690	-0.121	0.1716
remove top 2% (114)	1744	3820	5564	5580	3799	0.683	-0.120	0.1727
remove top 3%(171)	1722	3785	5507	5478	3729	0.677	-0.120	0.1731
remove top 4%(228)	1706	3744	5450	5371	3664	0.672	-0.119	0.1746
remove top 5%(285)	1686	3707	5393	5262	3573	0.663	-0.119	0.1764

Example: How central are clients in sexual networks created by commercial sex?

Table: The immunization experiment – continued

Travel	Max closeness	Max betweenness	diameter	Average Path Length
Original national data	0.000583	1038137.000	20	7.722642
remove top 0.5% (28)	0.000485	734212.500	24	8.681758
remove top 1% (57)	0.000418	555730.4	24	9.224694
remove top 2% (114)	0.000350	477722.400	30	11.07613
remove top 3%(171)	0.000308	510114.8	41	13.40275
remove top 4%(228)	0.000267	468840.800	36	14.57983
remove top 5%(285)	0.000243	237607.200	40	13.98779
Random	Max closeness	Max betweenness	diameter	Average Path Length
remove top 0.5% (28)	0.000577	988240.500	20	7.730389
remove top 1% (57)	0.000573	978207.000	20	7.73738
remove top 2% (114)	0.000565	935739.900	20	7.713839
remove top 3%(171)	0.000561	921086.600	20	7.730685
remove top 4%(228)	0.000559	904569.200	20	7.744975
remove top 5%(285)	0.000548	857372.700	21	7.774552

Reference

Badev, Anton (2013) “Discrete games in endogenous networks: Theory and policy.”

Ballester, Coralio, Antoni Calvó-Armengol, and Yves Zenou (2006) “Who’s who in networks. Wanted: The key player,” *Econometrica*, Vol. 74, pp. 1403–1417.

Ballester, Coralio, Yves Zenou, and Antoni Calvó-Armengol (2010) “Delinquent networks,” *Journal of the European Economic Association*, Vol. 8, pp. 34–61.

Bayer, Patrick, Randi Hjalmarsson, and David Pozen (2009) “Building criminal capital behind bars: Peer effects in juvenile corrections,” *The Quarterly Journal of Economics*, Vol. 124, pp. 105–147.

Bayer, Patrick, Stephen L Ross, and Giorgio Topa (2008) “Place of work and place of residence: Informal hiring networks and labor market outcomes,” *Journal of Political Economy*, Vol. 116, pp. 1150–1196.

Blume, Lawrence E, William A Brock, Steven N Durlauf, and Yannis M Ioannides (2010) "Identification of social interactions."

Boucher, Vincent (2015) "Structural homophily," *International Economic Review*, Vol. 56, pp. 235–264.

——— (2016) "Conformism and self-selection in social networks," *Journal of Public Economics*, Vol. 136, pp. 30–44.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin (2009) "Identification of peer effects through social networks," *Journal of Econometrics*, Vol. 150, pp. 41–55.

Calvo-Armengol, Antoni and Matthew O Jackson (2004) "The effects of social networks on employment and inequality," *The American Economic Review*, Vol. 94, pp. 426–454.

Calvó-Armengol, Antoni and Matthew O Jackson (2007) "Networks in labor markets: Wage and employment dynamics and inequality," *Journal of Economic Theory*, Vol. 132, pp. 27–46.

Calvó-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou (2009) "Peer effects and social networks in education," *The Review of Economic Studies*, Vol. 76, pp. 1239–1267.

Christakis, Nicholas A and James H Fowler (2007) "The spread of obesity in a large social network over 32 years," *New England Journal of Medicine*, Vol. 2007, pp. 370–379.

Clark, Andrew E and Youenn Lohéac (2007) ""It wasn't me, it was them!" Social influence in risky behavior by adolescents," *Journal of Health Economics*, Vol. 26, pp. 763–784.

Coleman, James S (1988) "Social capital in the creation of human capital," *American Journal of Sociology*, Vol. 94, pp. S95–S120.

Conley, Timothy G and Christopher R Udry (2010) "Learning about a new technology: Pineapple in Ghana," *The American Economic Review*, Vol. 100, pp. 35–69.

Conley, Timothy and Christopher Udry (2001) "Social learning through networks: The adoption of new agricultural technologies in Ghana," *American Journal of Agricultural Economics*, Vol. 83, pp. 668–673.

Currarini, Sergio, Matthew O Jackson, and Paolo Pin (2009) "An economic model of friendship: Homophily, minorities, and segregation," *Econometrica*, Vol. 77, pp. 1003–1045.

Dahl, Gordon B, Katrine V Løken, and Magne Mogstad (2014) "Peer effects in program participation," *The American Economic Review*, Vol. 104, pp. 2049–2074.

De Paula, Aureo (2016) "Econometrics of network models," Technical report, cemmap working paper, Centre for Microdata Methods and Practice.

Duflo, Esther and Emmanuel Saez (2003) "The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment," *The Quarterly Journal of Economics*, Vol. 118, pp. 815–842.

Fafchamps, Marcel and Flore Gubert (2007a) "The formation of risk sharing networks," *Journal of Development Economics*, Vol. 83, pp. 326–350.

——— (2007b) "Risk sharing and network formation," *The American Economic Review*, Vol. 97, pp. 75–79.

Fowler, James H and Nicholas A Christakis (2008) "Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study," *BMJ*, Vol. 337, p. a2338.

Goldenberg, Anna, Alice X Zheng, Stephen E Fienberg, Edoardo M Airoldi et al. (2010) "A survey of statistical network models," *Foundations and Trends® in Machine Learning*, Vol. 2, pp. 129–233.

Goldsmith-Pinkham, Paul and Guido W Imbens (2013) "Social networks and the identification of peer effects," *Journal of Business & Economic Statistics*, Vol. 31, pp. 253–264.

Hsieh, Chih-Sheng, Jaromír Kovářík, and Trevon Logan (2014) "How central are clients in sexual networks created by commercial sex?" *Scientific reports*, Vol. 4, p. 7540.

Hsieh, Chih-Sheng and Lung Fei Lee (2016) "A social interactions model with endogenous friendship formation and selectivity," *Journal of Applied Econometrics*, Vol. 31, pp. 301–319.

Hsieh, Chih-Sheng, Lung-Fei Lee, and Vincent Boucher (2019) "Specification and estimation of network formation and network interaction models with the exponential probability distribution."

Jackson, Matthew O (2010) "An overview of social networks and economic applications," *The handbook of social economics*, Vol. 1, pp. 511–85.

Jackson, Matthew O, Brian W Rogers, and Yves Zenou (2017) "The economic consequences of social-network structure," *Journal of Economic Literature*, Vol. 55, pp. 49–95.

Katz, Elihu and F Paul (1955) "Lazarsfeld (1955), Personal Influence," *The Part Played by People in the Flow of Mass Communication. New York*.

Kremer, Michael and Dan Levy (2008) "Peer effects and alcohol use among college students," *Journal of Economic perspectives*, Vol. 22, pp. 189–206.

Lee, Lung-fei, Xiaodong Liu, and Xu Lin (2010) "Specification and estimation of social interaction models with network structures," *The Econometrics Journal*, Vol. 13, pp. 145–176.

Lin, Xu (2010) "Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables," *Journal of Labor Economics*, Vol. 28, pp. 825–860.

Liu, Xiaodong, Eleonora Patacchini, and Edoardo Rainone (2017) "Peer effects in bedtime decisions among adolescents: a social network model with sampled data," *The Econometrics Journal*.

Logan, Trevon D (2010) "Personal characteristics, sexual behaviors, and male sex work: A quantitative approach," *American sociological review*, Vol. 75, pp. 679–704.

Logan, Trevon D and Manisha Shah (2013) "Face value: information and signaling in an illegal market," *Southern Economic Journal*, Vol. 79, pp. 529–564.

Mele, Angelo (2013) "A structural model of segregation in social networks."

Myers, Charles A Shultz, George Pratt Charles A Myers, and George P Shultz (1951) "The dynamics of a labor market: a study of the impact of employment changes on labor mobility, job satisfactions, and company and union policies,"Technical report.

Patacchini, Eleonora and Yves Zenou (2009) "Juvenile delinquency and conformism," *The Journal of Law, Economics, & Organization*, Vol. 28, pp. 1–31.

Pržulj, Nataša and Noël Malod-Dognin (2016) "Network analytics in the age of big data," *Science*, Vol. 353, pp. 123–124.

Yuster, Raphael and Uri Zwick (1997) "Finding even cycles even faster," *SIAM Journal on Discrete Mathematics*, Vol. 10, pp. 209–222.