Econometrics of Network Models

Áureo de Paula[1]

[1] UCL, São Paulo School of Economics, CeMMAP, IFS and CEPR

ESWC (August 2015)

Networks are ...

... **vulgar**!

of or relating to the common people; generally current; of the usual, typical, or ordinary kind. ("Vulgar" in Merriam-Webster.com, 2011)

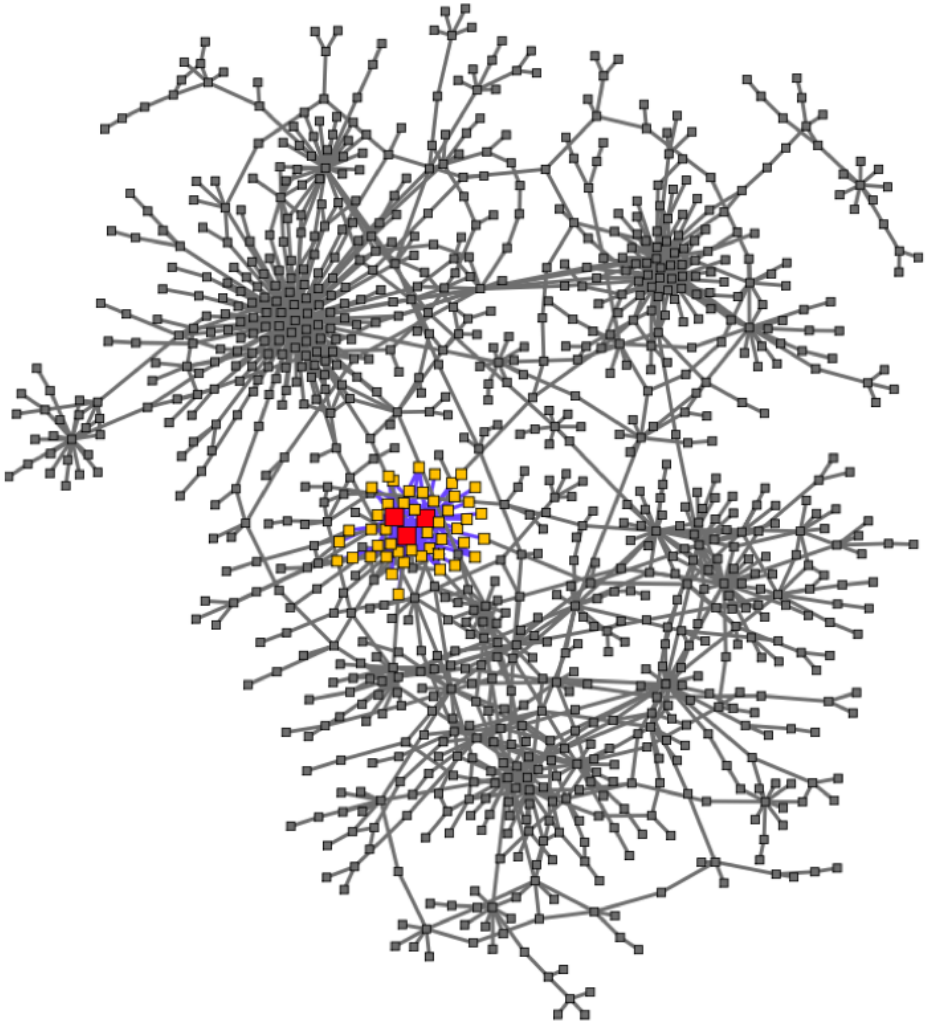
. . . everywhere.

Many social and economic activities are not directly mediated by prices (e.g., spillovers in education, labor market search, non-cognitive outcomes).

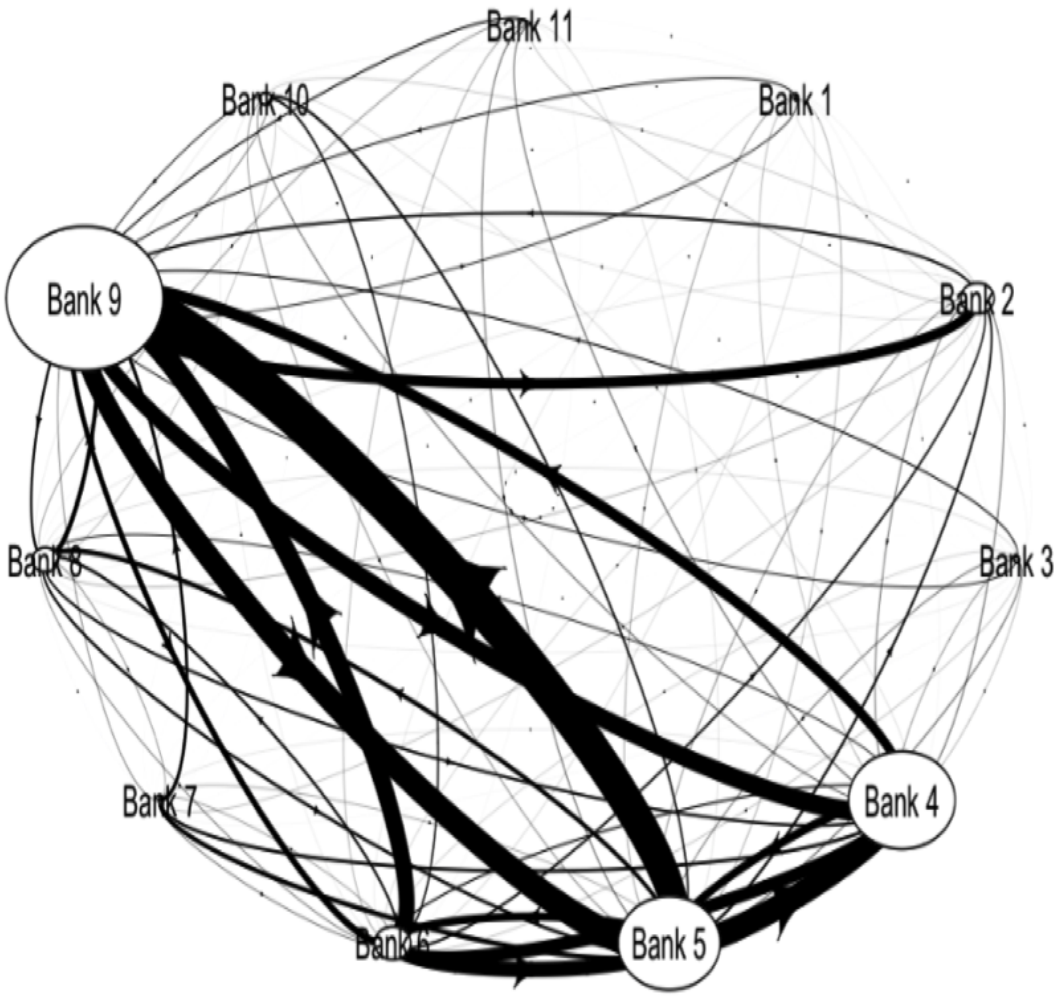
Many other behaviors are, but it also matters how agents are in contact (e.g., production and financial networks).

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

Look around!



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



Source: Denbee, Julliard, Li and Yuan (2014)

Source: Paul Butler

What this review does and does not cover.

- This is a selective overview of recent advances in the applied econometric literature.

  My focus is on econometric models of network formation and econometric models where outcome determination is mediated by networks.

- It does not elaborate on the measurement of network-related phenomena in macroeconomics, industrial organisation, finance, and trade. Nor does it aspires to cover econometric issues in subclasses of network models (e.g., barganing and matching in graphs).

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])

- $N_i(g)$: set of neighbours incident with node i in g.
  (degree of node $i = |N_i(g)|$)

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

- walks, paths, distance, cycles, clustering ...

- Centrality measures register the "importance" of a node.
  - degree centrality: how many neighbours;
  - closeness centrality: how far from any other node;
  - betweeness centrality: how crucial in connecting other nodes;
  - "centrality-referencing" measures.

    eigenvector centrality (Gould [1967], Bonacich [1972]);

    Katz centrality (Katz [1953]): ascribe $\tilde{\beta}^k$ to connections of length k.

    Bonacich centrality (Bonacich [1987]): $\alpha(\mathbf{I} - \tilde{\beta} W)^{-1} W \mathbf{1}$;

    ...

    diffusion centrality (Banerjee et al. [2014]);

    Google's PageRank index (Brin and Page [1998]).

- Other measures registering different features (e.g., clustering, etc.)

  (see, e.g., Jackson [2009])

- We can also define probabilistic graph models:

$$(\mathcal{G}, \sigma(\mathcal{G}), \mathbb{P}).$$

- Example: i.i.d. link formation on n nodes (Gilbert [1959], Erdös-Rényi [1960]). ($N \uparrow \infty \Rightarrow$ Poi degree distr.)
- Example: "iterative" models like preferential attachment (Barabási-Albert [1999]).
- Bollobás [2001], Jackson [2009], Kolaczyk [2009]

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

$$\mathbf{y}_{N\times 1} = \alpha \mathbf{1}_{N\times 1} + \beta W_{N\times N} \mathbf{y}_{N\times 1} + \eta \mathbf{x}_{N\times 1} + \gamma W_{N\times N} \mathbf{x}_{N\times 1} + \epsilon_{N\times 1}$$

$$\Leftrightarrow$$

$$\mathbf{y} = \alpha (\mathbf{I} - \beta W)^{-1} \mathbf{1} + (\mathbf{I} - \beta W)^{-1} (\eta \mathbf{I} + \gamma W) \mathbf{x} + (\mathbf{I} - \beta W)^{-1} \epsilon$$

- 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: Calvo-Armengól, 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 [2014] and other studies.)

- 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].

- 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.

- 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])

- $W_{ij} = (N-1)^{-1}, i \neq j$; $W_{ii} = 0 \Rightarrow W^2 = (N-1)^{-1}\mathbf{I} + (N-2)/(N-1)W$

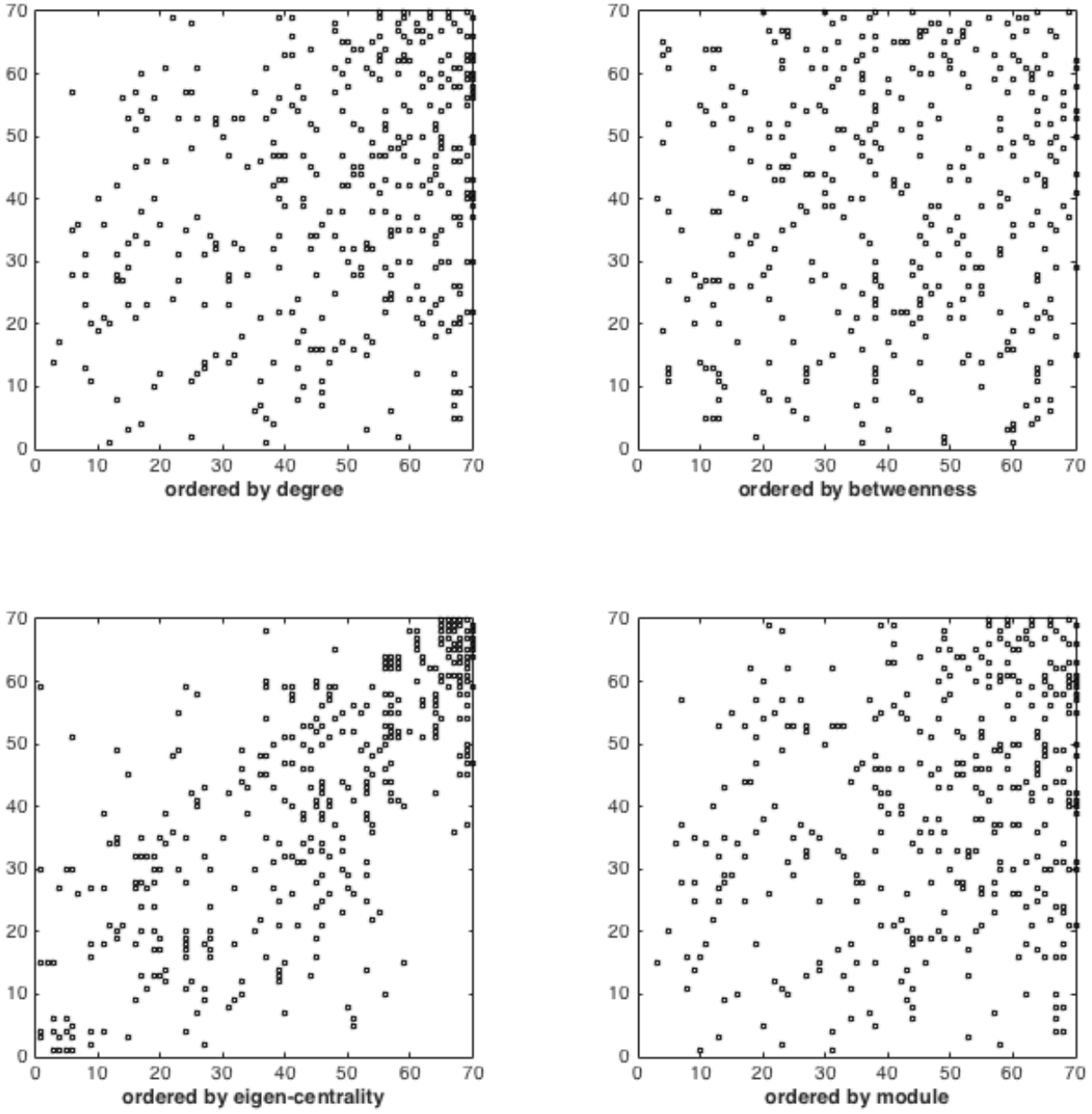
- W block diagonal and two blocks of different sizes $\Rightarrow$

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

  (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

- 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])

- In fact ...

  If W is diagonalizable such that $\sum_{j=1}^{N} W_{ij} = 1$ and $W_{ii} = 0$ for any $i \in \{1, \ldots, N\}$, $\mathbf{I}$, W and W^2 are linearly independent and $\beta\eta + \gamma \neq 0$, then $\alpha, \beta, \eta, \delta$ and W are uniquely determined by the reduced-form equation system.

  (de Paula, Rasul and Souza [2015])

- The reduced form coefficient matrix

$$\Pi = (\mathbf{I} - \beta W)^{-1}(\eta \mathbf{I} + \gamma W)$$

  can be estimated by OLS if repeated observations (T) are available and $T > N$.

- $T > N$ not necessarily attainable . . . but notice that (observed) Ws are "sparse" in many cases.
  (e.g., Atalay et al. [2011] $<$ 1%; Carvalho [2014] $\approx$ 3%; AddHealth $\approx$ 2%).

- If Π is itself sparse, one can then estimate

$$\hat{\pi}_i = \text{argmin}_{\pi_i} \frac{1}{T} \sum_t (y_{it} - \pi_i^\top \mathbf{x}_t)^2 + \lambda \sum_j p_T(\pi_{ij}).$$

(e.g., Manresa [2014] ($\beta = 0$, LASSO), Bonaldi, Hortacsu and Kastl [2014] (elastic net).)

- If $\beta \neq 0$, Π will not necessarily be sparse. In this case, one can focus on

$$\begin{gathered} \min_{(W,\beta,\delta,\gamma)} \tfrac{1}{T} \textstyle\sum_t \|\mathbf{y}_t - \Pi \mathbf{x}_t\|_2^2 + \lambda \sum_{i \neq j} p_T(W_{ij}) \\ \text{s.t. } (\mathbf{I} - \beta W)\Pi - (\eta \mathbf{I} + \gamma W) = 0 \end{gathered}$$

- Monte Carlo results in de Paula, Rasul and Souza [2015] are encouraging.

- 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 [2007], Tincani [2015].
- Example: Brock and Durlauf [2001, 2007], Xu and Lee [forthcoming] $\leftarrow$ 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 [2014]).

Other Considerations

- Spillovers mediated through networks: myriad of economic and social circumstances.
  - Example: Input-output networks. Suppliers and clients.
    (Carvalho et al. [2014], Bernard et al. [2014])
  - Example: Taxation networks. VAT “binds” compliance through network.
    (de Paula and Scheinkman [2010], Pomeranz [forthcoming])
  - Example: Propagation of micro shocks in production and financial systems.
    (Carvalho [2014], Acemoglu et al. [forthcoming])

. . . econometric insights may be useful in these and other contexts.

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.

  (Instruments (when available) can possibly be used in the previous models (see, e.g., Qu and Lee [2015]).)

- 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 [2014])

- 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 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 generating function;
  - ...
- 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 main alternative avenues:
  > Variational principles (Jordan and Wainwright [2008]);
  > MCMC (Kolaczyk [2009], recent articles).

- $\mathbb{P}(W_{ij} = 1 | W_{-ij} = w_{-ij}; \alpha) = \mathbb{P}(W_{ij} = 1; \alpha) \Rightarrow$ 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 [2015] takes αs to be "fixed effects."

  - Example: Chatterjee et al. [2011], Yan and Xu [2013] (undirected network, β-model). Graham [2014] characterizes MLE (with covar) and studies a conditional ML (using sufficient stats for α_i).

- Chandrasekhar and Jackson [2014]. Use additional subgraphs $(G_l)_{l=1}^{K}$ (beyond pairs): SUGM.
  (e.g., $K = 2$, $G_1 =$ pair, $G_2 =$ triangle)

Strategic Formation

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

- Typically, $u_i(g)$ (in undirected network) is a variation of

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

- Similar specifications for directed networks.

- Transferable or non-transferable utility.

- Network formation:
  - iterative;
    (Blume [1993], Watts [2001], Jackson and Watts [2002])
  - static.
    (Jackson and Wolinsky [1996], Bala and Goyal [2000])

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

  - Example: Christakis, Fowler, Imbens and Kalyanaraman [2010], undirected.
    (formation $\approx$ stochastic stability analysis in Jackson and Watts [2002])

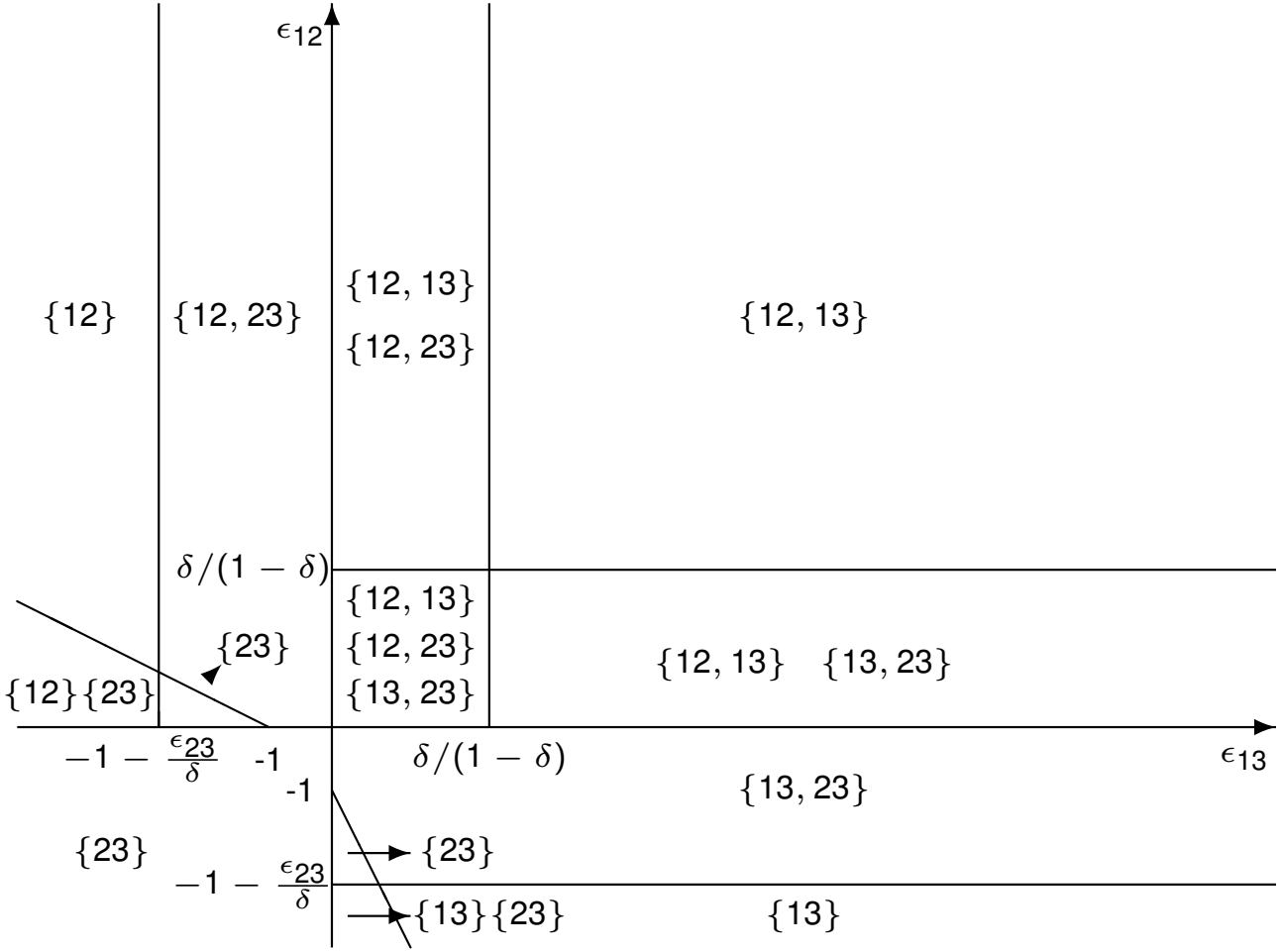
  - Example: Mele [2013], Badev [2013], directed.
    (Potential function $\Rightarrow$ NE or k-Nash stable equilibria w/o unobservables)
    (Meeting protocol + myopic updating $\Rightarrow$ unique invariant distr on graphs)

  > i.i.d. EV unobservables $\Rightarrow$ ERGM...
    (Mele [2013] suggests MC scheme to improve on performance)

  - Models are fitted to AddHealth data on friendships and outcomes (smoking, Badev [2013]) using Bayesian methods or ML.

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

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



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

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

- Sheng [2014]. Use small size subnetworks consistent with PS + additional payoff structures $\Rightarrow$ bounds. (Maybe too conservative if $N \gg$ subnetwork size.)

- Miyauchi [2014]. Payoff restrictions $\Rightarrow$ complementarity (supermodularity). Use lattice structure of equilibrium set to improve computation.

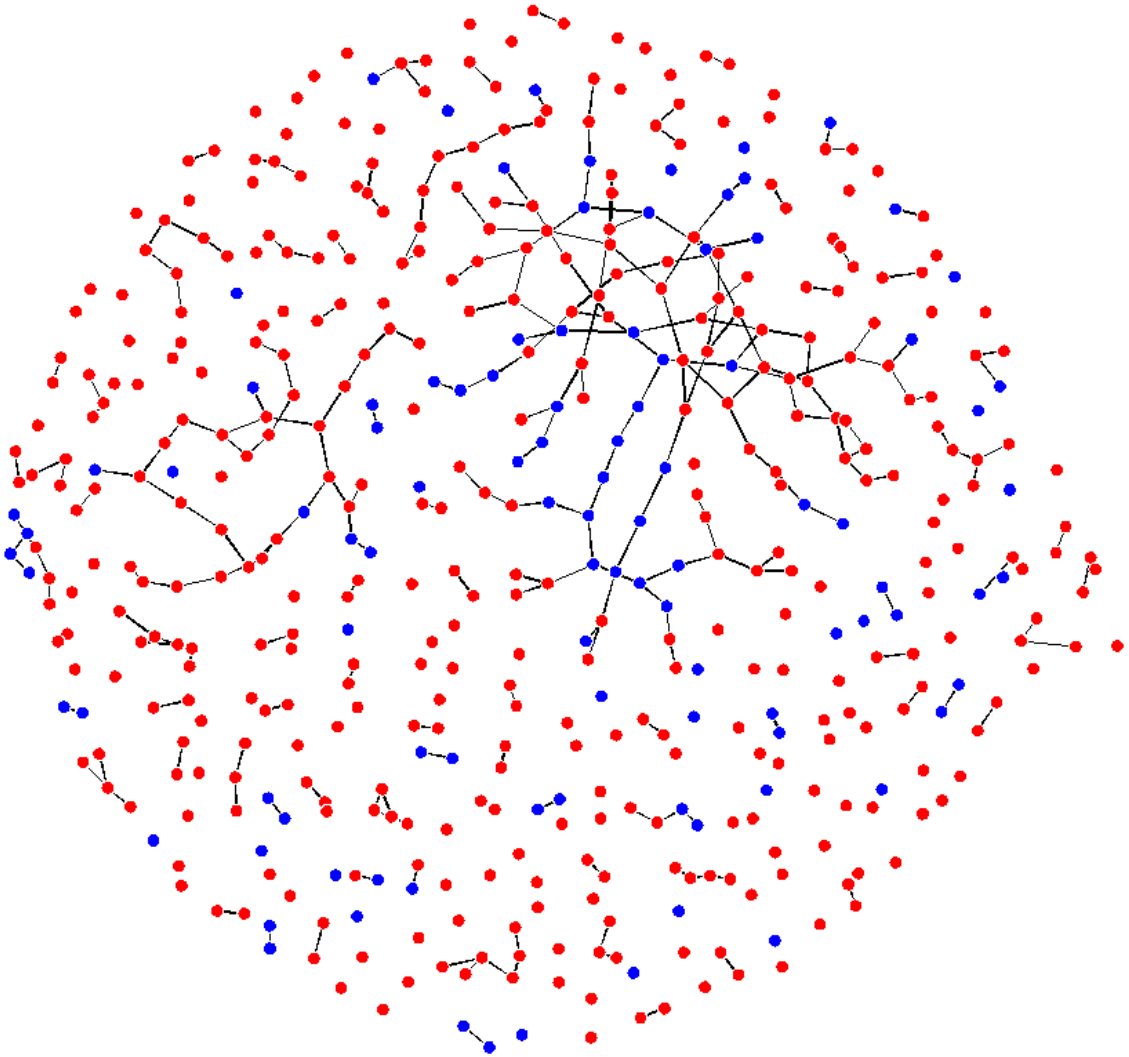
- Other examples: Boucher and Mourifié [2013], Leung [2015] ...

- de Paula, Richards-Shubik and Tamer [2015]: pairwise stability in (non-traferable 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.

  - $\star$ 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).



- Incomplete information: e.g., Gilleskie and Zheng [2009], Leung [2015].

- Dynamic (farsighted) network formation: e.g., Lee and Fong [2011] (bipartite), Johnson [2012].
  ($\approx$ empirical dynamic games)

- Network formation and outcomes: Gilleskie and Zheng [2009], Badev [2013]. Goldsmith-Pinkham and Imbens [2013] (dyadic formation + linear-in-means), Hsieh and Lee [2013] (ERGM + linear-in-means).
  (Partial identification in formation model $\Rightarrow$ partial identification in outcome model parameters. E.g., Ciliberto, Murry and Tamer [2015], Chesher and Rosen [2014].)

- Surveys: Kolaczyk [2009], Goldenberg et al. [2009], Hunter et al. [2012] (statistics); Graham [2015], Chandrasekhar [2015] (econometrics).

Measurement

- Measurement is (obviously) essential!

- Complete networks are not always and costly to observe.
  (e.g., ERGM not projective (Rinaldo and Shalizi [2013]); Handcock and Gile [2010], Koskinen et al. [2010])

  (Sampling and inference schemes for network features from incompletely observed networks (see Kolaczyk [2009] could be used in certain models of interest (e.g., Chandrasekhar and Jackson [2014], de Paula, Richards-Shubik and Tamer [2015])

- Measurement error in outcomes and/or covariates complicate inference strategies in interaction systems.
  (e.g., Moffitt [2001], Ammermueller and Pischke [2009], Angrist [2014]).

- Some results are still possible:

  Suppose there are two groups $g = 1, 2$ such that $N_1 = 2$ and $N_2 = 3$. If $|\beta| < 1$, $W_{ij,g} = (N_g - 1)^{-1}$ if $i \neq j$, $W_{ii,g} = 0$, $\mathbb{V}(\epsilon_g | \mathbf{x}_g) = \sigma^2 \mathbf{I}_g$ and $\tilde{y}_{i,g} = y_{i,g} + v_{i,g}$ where $v_{i,g} \perp\!\!\!\perp v_{j,h}$ and $v_{i,g} \perp\!\!\!\perp y_{j,h}$ for any i, j, g and h, then β is identified.

- This would not work if measurement error is in the covariates (as highlighted by the papers above).

  ...but nevertheless demonstrate that the network interaction structure may itself be used to handle the measurement error.

Looking ahead

- Networks are everywhere and research has tackled many questions and contexts.
- More is needed.
  - Heterogeneity (e.g., Masten [2015]);
  - Nonlinearities (e.g., Tincani [2015]);
  - Dynamics (forward-looking);
  - Measurement.
- Many more details in the paper!