Inference on Network Density

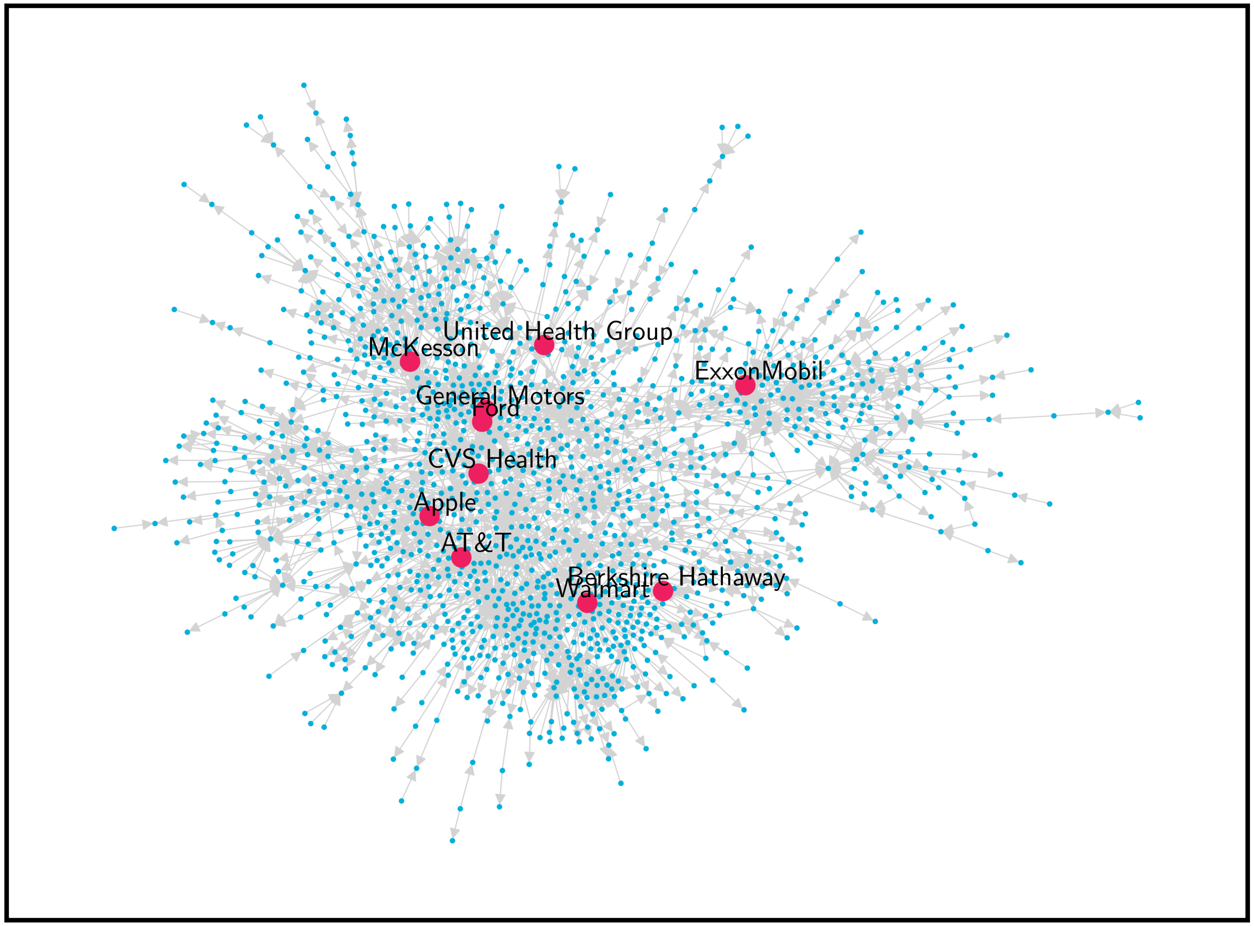
Jinan University Econometrics Camp
Guangzhou, China, December 9-12, 2019

(Based on IAAE Plenary Lecture in June 2019)

Bryan S. Graham

University of California - Berkeley

US Buyer-Supplier Network, 2015



“Networks”

You’ve drawn the picture, now what?

1. Find a question you care about (good luck, but you are on your own).

2. Compute some network summary statistics, attach measures of statistical precision to them and put together “Table 1” (helping with this is my – *our* – job).

3. Translate a scientific conjecture into a statistical hypothesis and test it (I have some ideas and points of departure to share).

Examples

[#1] How to estimate, and conduct inference on, network density and/or average degree?

The transitivity index would be a more interesting, but also far more complicated, example.

[#2] How to test for "strategic interdependencies" in link formation across "heterogenous" agents?

Sources: Graham (forthcoming, *Handbook of Econometrics*), Graham, Niu and Powell (2019, *WP*) and Pelican and Graham (2019, *WP*)...other researchers' work which I will cite as I go along.

Example 1: Network Density

(a “reduced form” example)

Density

Draw two agents, i and j, independently at random, from a large network. What is the (*ex ante*) chance they will be linked ($D_{ij} = 1$) versus not ($D_{ij} = 0$)?

We call this probability, $\rho_N = \Pr\left(D_{ij} = 1\right)$, *network density*.

Its analog estimate is the empirical frequency

$$\begin{aligned}\hat{\rho}_N &= \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} D_{ij} \\ &= \binom{N}{2}^{-1} \sum_{i<j} D_{ij}\end{aligned}$$

I am going to talk about this average for the next ~30 minutes.

Average Degree & Sparseness

Draw an agent at random. What is her expected number of links?

We call this expectation *average degree* ($\lambda_N = (N-1)\,\rho_N$) and estimate it by

$$\widehat{\lambda}_N = (N-1)\,\widehat{\rho}_N.$$

If $\lambda_N \to \lambda$ with $0 < \lambda < \infty$ as $N \to \infty$, then the graph is *sparse*. Sparsity requires that $\rho_N = O\left(N^{-1}\right)$.

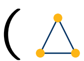
If $\rho_N = \rho$ with $0 < \rho < 1$, then $\lambda_N = O(N)$ and the graph is *dense*.

Density and Degree Redux

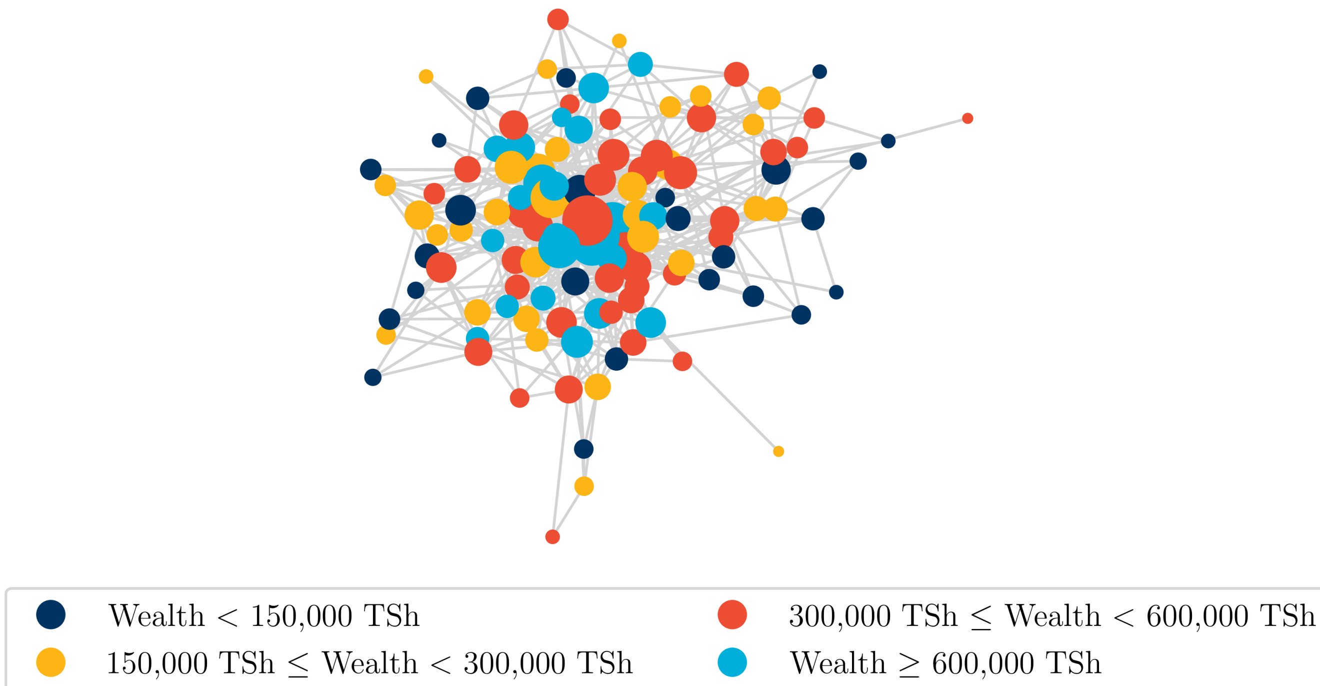
Density and average degree are (arguably) the most basic network summary statistics we can compute.

Understanding how to estimate, and conduct inference on ρ_N, is a prerequisite for understanding how to (for example)

1. undertake (dyadic) regression analysis (e.g., gravity models);

2. analyze other network moments (e.g., triangle () and two-star () frequencies; transitivity index).

Nyakatoke Risk-Sharing Network



$$\left(N = 119,\ n \overset{def}{\equiv} \binom{N}{2} = 7,021\right)$$

Density/Average Degree in Nyakatoke

In Nyakatoke we have:

$$\underset{\text{(a.s.e)}}{\widehat{\rho}_N} = \underset{(0.0072)}{0.0698}, \quad \underset{\text{(a.s.e)}}{\widehat{\lambda}_N} = \underset{(0.8459)}{8.2364}$$

How were these asymptotic standard errors (a.s.e) calculated?

What is their justification?

Can we use quantiles of the normal distribution as critical values?

Some Literature

Key references: Holland and Leinhardt (1976, *Sociological Methodology*) and Bickel, Chen and Levina (2011, *Annals of Statistics*)

Additional references: Nowicki (1991, *Statistica Neerlandica*), Bhattacharya and Bickel (2015, *Annals of Statistics*), Menzel (2017, *arXiv*), Davezies, D'Haultfoeuille and Guyonvarch (2019, *WP*), Auerbach (2019, *arXiv*), Graham (forthcoming, *Handbook of Econometrics*).

Conditional Edge Independence

Let $\{A_i\}_{i=1}^N$ be i.i.d uniform random *latent* agent-specific variables.

Today we will proceed 'as if' links/edges form independently *conditional* on the $\{A_i\}_{i=1}^N$ with

$$D_{ij}\Big| A_i, A_j \sim \text{Bernoulli}\left(h_N\left(A_i, A_j\right)\right)$$

for every dyad $\{i, j\}$ with $i < j$.

$h_N(a_1, a_2)$ is a symmetric edge probability function; typically called a *graphon*.

Unconditionally D_{ij} and D_{ik} may covary, but conditional on the latent A_i, A_j and A_k they do not.

Conditional Edge Independence (continued)

Graphon structure induces a particular form of dependence across the various elements of $\mathbf{D}_N = \left[D_{ij}\right]$, the adjacency matrix in hand.

We can motivate graphon approach formally using exchangeability arguments (Aldous-Hoover Theorem).

This involves viewing $\mathbf{D}_N$ as a record of all links amongst a random sample of N agents drawn from a large (infinite) unlabeled graph (cf., Bickel and Chen, 2009; Lovasz, 2012).

Network Density: Decomposition

We can decompose our density estimate as

$$\widehat{\rho}_N = U_N + V_N$$

with

$$U_N = \binom{N}{2}^{-1} \sum_{i<j} h_N\left(A_i, A_j\right)$$

$$V_N = \binom{N}{2}^{-1} \sum_{i<j} \left\{D_{ij} - h_N\left(A_i, A_j\right)\right\}.$$

1. U_N is a U-Statistic (with sample size dependent kernel);

2. V_N is a sum of uncorrelated random variables.

Network Density: Decomposition

Further decomposing U_N into its Hájek Projection and a reminder term yields

$$U_N = \rho_N + U_{1N} + U_{2N}$$

with, for $h_{1N}(a) = \mathbb{E}\left[h_N(a, A)\right]$,

$$U_{1N} = \frac{2}{N} \sum_{i=1}^{N} \left\{h_{1N}(A_i) - \rho_N\right\}$$

$$U_{2N} = \frac{2}{N(N-1)} \sum_{i<j} \Big\{h_N\left(A_i, A_j\right) - h_{1N}(A_i) \\ -h_{1N}\left(A_j\right) + \rho_N\Big\}.$$

Note that the U_{1N}, U_{2N} and V_N are all uncorrelated with one another.

Network Density: Decomposition

Putting things together yields the final decomposition,

$$\widehat{\rho}_N - \rho_N = U_{1N} + U_{2N} + V_N$$

consisting of

1. V_N: Projection Error #1: $\widehat{\rho}_N - \underbrace{\mathbb{E}\left[\widehat{\rho}_N \middle| A_1, \ldots, A_N\right]}_{\text{U-Statistic}}$;

2. U_{1N} : Hájek Projection;

3. U_{2N}: Projection Error #2.

Network Density: Variance

Define the notation:

$$P\left(\bullet\!-\!\bullet\right) = \mathbb{E}\left[D_{12}\right] \quad (= \rho_N)$$
$$Q\left(\wedge\right) = \mathbb{E}\left[D_{12}D_{13}\right]$$

and $\tilde{P}\left(\bullet\!-\!\bullet\right) = P\left(\bullet\!-\!\bullet\right)/\rho_N$ $(= 1)$ and $\tilde{Q}\left(\wedge\right) = Q\left(\wedge\right)/\rho_N^2$.

Dividing by (powers of) ρ_N stabilizes such that, for example, $\tilde{Q}\left(\wedge\right)$ does not vanish as $N \to \infty$.

Further define:

$$\Omega_{1N} = \rho_N^2\left\{\tilde{Q}\left(\wedge\right) - \tilde{P}\left(\bullet\!-\!\bullet\right)\tilde{P}\left(\bullet\!-\!\bullet\right)\right\} = O\left(\rho_N^2\right)$$
$$\Omega_{2N} = \mathbb{V}\left(\mathbb{E}\left[D_{12}|\,\mathbf{A}\right]\right) = O\left(\rho_N^2\right)$$
$$\Omega_{3N} = \mathbb{E}\left[\mathbb{V}\left(D_{12}|\,\mathbf{A}\right)\right] = O\left(\rho_N\right).$$

Network Density: Variance

Applying the variance operator we get

$$
\begin{aligned}
\mathbb{V}\left(\widehat{\rho}_N\right) &= \mathbb{V}\left(U_{1N}\right) + \mathbb{V}\left(U_{2N}\right) + \mathbb{V}\left(V_N\right) \\
&= \frac{4\Omega_{1N}}{N} + \binom{N}{2}^{-1}\left[\Omega_{2N} - 2\Omega_{1N}\right] + \binom{N}{2}^{-1}\Omega_{3N} \\
&= O\left(\rho_N^2 N^{-1}\right) + O\left(\rho_N^2 N^{-2}\right) + O\left(\rho_N N^{-2}\right)
\end{aligned}
$$

Which, after normalizing, yields

$$
\begin{aligned}
\mathbb{V}\left(\frac{\widehat{\rho}_N}{\rho_N}\right) =& \mathbb{V}\left(\frac{U_{1N}}{\rho_N}\right) + \mathbb{V}\left(\frac{U_{2N}}{\rho_N}\right) + \mathbb{V}\left(\frac{V_N}{\rho_N}\right) \\
=& O\left(\frac{1}{N}\right) + O\left(\frac{1}{N^2}\right) + O\left(\frac{1}{N\lambda_N}\right)
\end{aligned}
$$

Network Density: Sparsity and Degeneracy

The network is sparse if $(N-1)\,\rho_N \to \lambda > 0$ as $N \to \infty$.

The network is dense if $(N-1)\,\rho_N = O\,(N)$.

The graphon is degenerate if $\mathbb{V}\,(h_{1N}\,(A)) = 0$ (e.g., $\mathbb{E}\left[\left. D_{12} \right| A_1\right] = 0$).

Degeneracy $\Rightarrow \Omega_{1N} = 0$.

The Erdos-Renyi graphon is "degenerate".

Network Density: Rates-of-Convergence

	Degenerate	**Non-Degenerate**
Sparse	$\sqrt{N\lambda_N}$	$\sqrt{N}$ or $\sqrt{N\lambda_N}$
Dense	N	$\sqrt{N}$

Variance Estimation

Some options:

1. Holland-Leinhardt (1976)/Fafchamps-Gubert (2007) variance estimate;

2. Jack-knife variance estimate;

3. Corrected Jack-knife;

4. Bootstrap (cf., Menzel, 2019).

Will just discuss the first option today.

Fafchamps-Gubert (2007) Variance Estimate

Observe that

$$\Sigma_{1N} = \mathbb{E}\left[\left(D_{12} - \rho_N\right)\left(D_{13} - \rho_N\right)\right] = \Omega_{1N}$$
$$\Sigma_{2N} = \mathbb{E}\left[\left(D_{12} - \rho_N\right)\left(D_{12} - \rho_N\right)\right] = \Omega_{2N} + \Omega_{3N}$$

Natural analog estimates for these two terms are

$$\begin{aligned}\widehat{\Sigma}_{1N} =& \binom{N}{3} \sum_{i<j<k} \frac{1}{3}\Big\{\left(D_{ij} - \widehat{\rho}_N\right)\left(D_{ik} - \widehat{\rho}_N\right) \\ &+ \left(D_{ij} - \widehat{\rho}_N\right)\left(D_{jk} - \widehat{\rho}_N\right) \\ &+ \left(D_{ik} - \widehat{\rho}_N\right)\left(D_{jk} - \widehat{\rho}_N\right)\Big\} \\ \widehat{\Sigma}_{2N} =& \binom{N}{2}^{-1} \sum_{i<j} \left(D_{ij} - \widehat{\rho}_N\right)^2 = \widehat{\rho}_N\left(1 - \widehat{\rho}_N\right)\end{aligned}$$

Degrees of freedom corrections? These are unbiased variance estimates when $\widehat{\rho}_N$ is replaced by ρ_N.

Limit Distribution

Let $\hat{\sigma}_N^2 = \frac{4}{N}\hat{\Sigma}_{1N} + \frac{2}{N(N-1)}\left(\hat{\Sigma}_{2N} - 2\hat{\Sigma}_{1N}\right)$ be the Fafchamps and Gubert (2007) variance estimate.

For all cases, except the dense/degenerate one, we can use martingale triangular array ideas to show that (cf., Graham, Niu and Powell, 2019).

$$\frac{\hat{\rho}_N}{\hat{\sigma}_N} \xrightarrow{D} \mathcal{N}(0,1).$$

In the dense/degenerate case the limit distribution may be non-Gaussian (see Menzel (2019) for examples).

What did we learn?

Holland and Leinhardt (1976) correctly calculated the variance of $\widehat{\rho}_N$ over 40 years ago!

Their variance estimator is closely related to the one proposed by Fafchamps and Gubert (2007) in a regression context.

No results on the limit distribution of $\widehat{\rho}_N$ (except in some very special cases) were available until Bickel, Chen and Levina (2011); whose results I have generalized here.

See Menzel (2019) for some great results on the use of the bootstrap, as well as the dense/degenerate case.

What did we learn? (continued)

Now you can report a statistic and standard error with your network dataset (Row 1 of Table 1 is in the bag!)

The density case is canonical, although each example has its own twists and terms.

For some information on dyadic regression (lecture notes, computer code and examples) see the materials in the "/St. Gallen/2018" folder at

`https://github.com/bryangraham/short_courses`

Practitioners need these results, getting them is fun piece of mathematical statistics.

Example 2: Testing for Interdependencies

(a “structural” example)

This example is based on joint work with Andrin Pelican.

Two classes of network formation models

Null model: the utility i generates by linking with j depends upon ego (i) and alter (j) attributes alone.

- Stochastic Block Model

- β-Model

Alternative model: the utility generated by an i to j link *additionally varies* with the presence or absence of *other links* in the network.

- Strategic models (Jackson and Wolisky, 1996, *Journal of Economic Theory*)

Research question

Can we determine whether the network in hand was generated according to null or alternative model?

Very little prior work in this space (see de Paula et al., 2018, *Econometrica*; Sheng, forthcoming, *Econometrica*).

Why I care and you should (might?) too

With strategic behavior:

1. There may be multiple equilibrium network configurations.

2. The observed configuration may not maximize welfare.

3. Vertex removal (and/or local re-wirings) can trigger a process of link revision global in scope.

The effect of policies on the form of a network are very different under the null vs. the alternative.

Utility

Random utility framework *a la* McFadden (1973).

Let $\mathbf{d} \in \mathbb{D}$ be a (directed) adjacency matrix. The utility agent i gets from some feasible network wiring $\mathbf{d}$ is

$$\nu_i\left(\mathbf{d}_i, \mathbf{d}_{-i}; \mathbf{U}\right) = \sum_j d_{ij}\left[A_i + B_j + X_i'\Lambda X_j + \gamma_0 s_{ij}\left(\mathbf{d}\right) - U_{ij}\right],$$

where:

1. A_i is a "ego/sender effect" (out-degree heterogeneity);

2. B_j a "alter/receiver effect" (in-degree heterogeneity);

Utility (continued)

3. X_i equals a vector of K observed community membership dummies (the $K \times K$ matrix Λ parameterizes homophily);

4. $s_{ij}(\mathbf{d}) = s_{ij}(\mathbf{d} - ij) = s_{ij}(\mathbf{d} + ij)$ is a network/strategic effect; can be used to model:

   (a) reciprocity: $s_{ij}(\mathbf{d}) = d_{ji}$;

   (b) transitivity: $s_{ij}(\mathbf{d}) = \sum_k d_{ik}d_{kj}$

5. $\left\{U_{ij}\right\}_{i \neq i}$ idiosyncratic utility shifter (i.i.d. logistic)

Alternative model is an N-player game

$\mathbf{d} \in \mathbb{D}$ - a candidate network wiring – is a *pure strategy combination* (each agent decides which, out of $N-1$ choices, links to send).

A (pure strategy) Nash equilibrium (NE) is a pure strategy combination $\mathbf{d}^*$ where, for $\mathbf{U} = \mathbf{u}$ and all $i = 1, \ldots, N$,

$$\nu_i \left(\mathbf{d}_i^*, \mathbf{d}_{-i}^*, \mathbf{u} \right) \geq \nu_i \left(\mathbf{d}_i, \mathbf{d}_{-i}^*, \mathbf{u} \right) \tag{1}$$

for all possible (other) linking strategies $\mathbf{d}_i$.

We assume that $\mathbf{D}$ – the *observed* network – satisfies (1) at the realized $\mathbf{U}$.

Testing goal: challenges

Goal is to construct a test of the no strategic interaction ($\gamma_0 = 0$) null.

Three key challenges:

1. null is composite – nuisance parameter $\delta = (\mathbf{A}, \mathbf{B}, \Lambda)$ is high dimensional (worry: size distortion);

2. can't evaluate likelihood under the alternative (worry: how to maximize power?);

3. characterizing/simulating null distribution (worry: feasibility).

Testing goal: solutions

1. Apply exponential family/sufficiency theory (Ferguson, 1967; Lehmann & Romano, 2005).

2. Find *locally* best test:

   (a) derivative of likelihood w.r.t to γ difficult to compute (incompleteness);

   (b) exploit insights from the econometrics of games (e.g., Tamer, 2003; Bajari *et al*. 2010a,b).

3. Extend MCMC methods for (constrained) network simulation (e.g., Sinclair, 1993).

Results (ongoing)

Test works well on example networks.

Consistent with theory:

- size properties are excellent;

- more powerful than naive approaches (e.g., counting triangles).

Scalability to large networks is a challenge.

Closing comments

This project draws on (i) classical exponential family theory, (ii) game theory and (iii) simulation methods developed by computer scientists.

Under the null estimation of the model parameters is a solved problem (e.g., Graham, 2017, *Econometrica*; Jochmans, forthcoming, *Journal of Business and Economic Statistics*).

Under the alternative (set) estimation is challenging, but see work of de Paula et al. (2018) and Sheng (forthcoming) for ideas.

Overall closing comments

Many datasets have natural, and illuminating, graph-theoretic representations.

Broadly defined, the range of applications for "network econometrics" is massive.

There are many open questions regarding modeling, computation and basic statistical theory.

Tremendous opportunities for empirical work.

Resources

http://bryangraham.github.io/econometrics/

https://github.com/bryangraham/short_courses/