Inference on Network Density
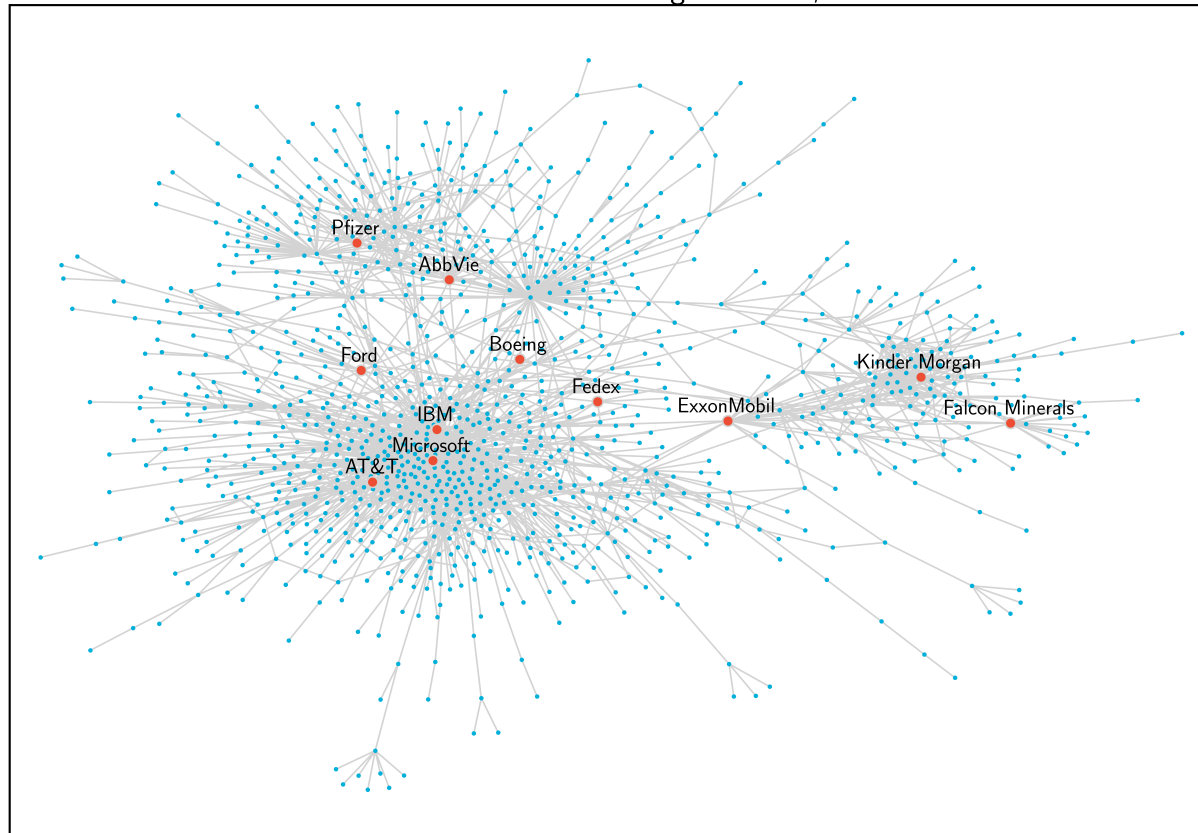
An Introduction to the Econometrics of Networks

CEMFI, November 28th to December 2nd, 2022

Bryan S. Graham

University of California - Berkeley

US R&D Collaboration Network, 2020

R & D Collaboration Among US Firms, 2020

"Networks"

You've drawn the picture, now what?

1. Find a question you, and preferably someone else, 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. .....

A starting point

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.

<u>Sources</u>: Graham (2020, Handbook of Econometrics), Graham et al. (2022, Journal of Econometrics)

Network 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

$$\widehat{\rho}_N = \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} D_{ij}$$
$$= \binom{N}{2}^{-1} \sum_{i<j} D_{ij}$$

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

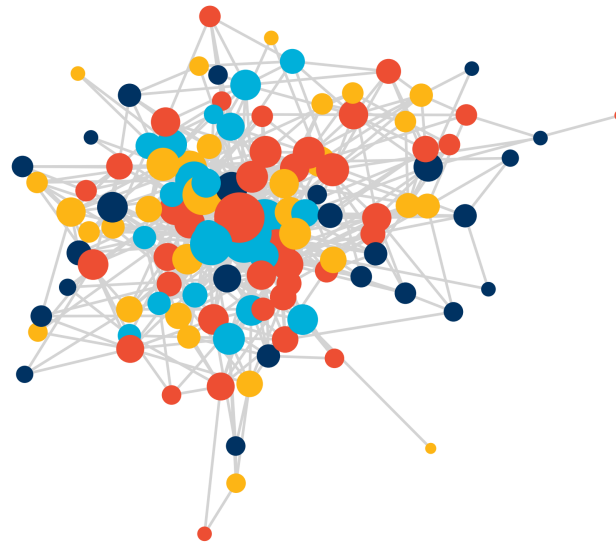Density and average degree are (arguably) the most basic network summary statistics we can compute.

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

Legend:
- ● Wealth $< 150{,}000$ TSh
- ● $150{,}000$ TSh $\leq$ Wealth $< 300{,}000$ TSh
- ● $300{,}000$ TSh $\leq$ Wealth $< 600{,}000$ TSh
- ● Wealth $\geq 600{,}000$ TSh

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

Density/Average Degree in Nyakatoke

In Nyakatoke we have:

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

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 & Leinhardt (1976, Sociological Methodology), Bickel et al. (2011, Annals of Statistics)

Additional references: Nowicki (1991, Statistica Neerlandica), Bhattacharya & Bickel (2015, Annals of Statistics), Menzel (2021, Econometrica), Davezies et al. (2021, Annals of Statistics)

Conditional Edge Independence

Let $\{A_i\}_{i=1}^N$ be i.i.d uniform random <u>latent</u> 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 $i < j$.

$h_N\left(a_1, a_2\right)$ 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 & Chen, 2009; Lovász, 2012).

Network Density: Decomposition

We can decompose our density estimate as

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

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

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

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

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: $\widehat{\rho}_N - \underbrace{\mathbb{E}\left[\widehat{\rho}_N \mid A_1, \ldots, A_N\right]}_{\text{U-Statistic}}$;

2. U_{1N} : Hájek Projection;

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

Network Density: Variance

Define the notation:

$$P\left(\text{╾╼}\right) = \mathbb{E}\left[D_{12}\right] \ \left(= \rho_N\right)$$

$$Q\left(\wedge\right) = \mathbb{E}\left[D_{12}D_{13}\right]$$

and $\tilde{P}\left(\text{╾╼}\right) = P\left(\text{╾╼}\right)/\rho_N \ \left(= 1\right)$ 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:

$$\Omega_{1N} = \rho_N^2 \left\{\tilde{Q}\left(\wedge\right) - \tilde{P}\left(\text{╾╼}\right)\tilde{P}\left(\text{╾╼}\right)\right\} = O\left(\rho_N^2\right)$$

$$\Omega_{2N} = \mathbb{V}\left(\mathbb{E}\left[D_{12}|\mathbf{A}\right]\right) = O\left(\rho_N^2\right)$$

$$\Omega_{3N} = \mathbb{E}\left[\mathbb{V}\left(D_{12}|\mathbf{A}\right)\right] = O\left(\rho_N\right).$$

Network Density: Variance

Applying the variance operator we get

$$\mathbb{V}\left(\widehat{\rho}_N\right) = \mathbb{V}\left(U_{1N}\right) + \mathbb{V}\left(U_{2N}\right) + \mathbb{V}\left(V_N\right)$$

$$= \frac{4\Omega_{1N}}{N} + \binom{N}{2}^{-1}\left[\Omega_{2N} - 2\Omega_{1N}\right] + \binom{N}{2}^{-1}\Omega_{3N}$$

$$= O\left(\rho_N^2 N^{-1}\right) + O\left(\rho_N^2 N^{-2}\right) + O\left(\rho_N N^{-2}\right)$$

Which, after normalizing, yields

$$\mathbb{V}\left(\frac{\widehat{\rho}_N}{\rho_N}\right) = \mathbb{V}\left(\frac{U_{1N}}{\rho_N}\right) + \mathbb{V}\left(\frac{U_{2N}}{\rho_N}\right) + \mathbb{V}\left(\frac{V_N}{\rho_N}\right)$$

$$= O\left(\frac{1}{N}\right) + O\left(\frac{1}{N^2}\right) + O\left(\frac{1}{N\lambda_N}\right)$$

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., $\mathbb{E}\,[D_{12}|\,A_1] = 0$).

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, 2021; Davezies et al., 2021).

Will just discuss the first option today.

Fafchamps-Gubert (2007) Variance Estimate

Observe that

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

Natural analog estimates for these two terms are

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

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

Limit Distribution

Let $\widehat{\sigma}_N^2 = \frac{4}{N}\widehat{\Sigma}_{1N} + \frac{2}{N(N-1)}\left(\widehat{\Sigma}_{2N} - 2\widehat{\Sigma}_{1N}\right)$ 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 et al., 2022).

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

In the dense/degenerate case the limit distribution may be non-Gaussian (see Menzel, 2021 for examples).

What did we learn?

Holland & 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 & 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 et al. (2011); whose results I have generalized here.

See Menzel (2021) 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.

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

*References

Bhattacharya, S. & Bickel, P. J. (2015). Subsampling bootstrap of count features of networks. *Annals of Statistics*, 43(6), $2384 - 2411$.

Bickel, P. J. & Chen, A. (2009). A nonparametric view of network models and newman-girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50), $21068 - 21073$.

Bickel, P. J., Chen, A., & Levina, E. (2011). The method of moments and degree distributions for network models. *Annals of Statistics*, 39(5), $2280 - 2301$.

Davezies, L., d'Haultfoeuille, X., & Guyonvarch, Y. (2021). Empirical process results for exchangeable arrays. *Annals of Statistics*, 49(2), 845 − 862.

Fafchamps, M. & Gubert, F. (2007). The formation of risk sharing networks. *Journal of Development Economics*, 83(2), 326 − 350.

Graham, B. S. (2020). *Handbook of Econometrics*, volume 7, chapter Network data, (pp. 111 − 218). North-Holland: Amsterdam, 1st edition.

Graham, B. S., Niu, F., & Powell, J. L. (2022). Kernel density estimation for undirected dyadic data. *Journal of Econometrics (forthcoming)*.

Holland, P. W. & Leinhardt, S. (1976). Local structure in social networks. *Sociological Methodology*, 7, 1 − 45.

Lovász, L. (2012). *Large Networks and Graph Limits*, volume 60 of *American Mathematical Society Colloquium Publications*. American Mathematical Society.

Menzel, K. (2021). Bootstrap with cluster-dependence in two or more dimensions. *Econometrica*, 89(5), 2143 − 2188.

Nowicki, K. (1991). Asymptotic distributions in random graphs with applications to social networks. *Statistica Neerlandica*, 45(3), 295 − 325.