# Inference on Network Density

## Econometric Methods for Social Spillovers and Networks

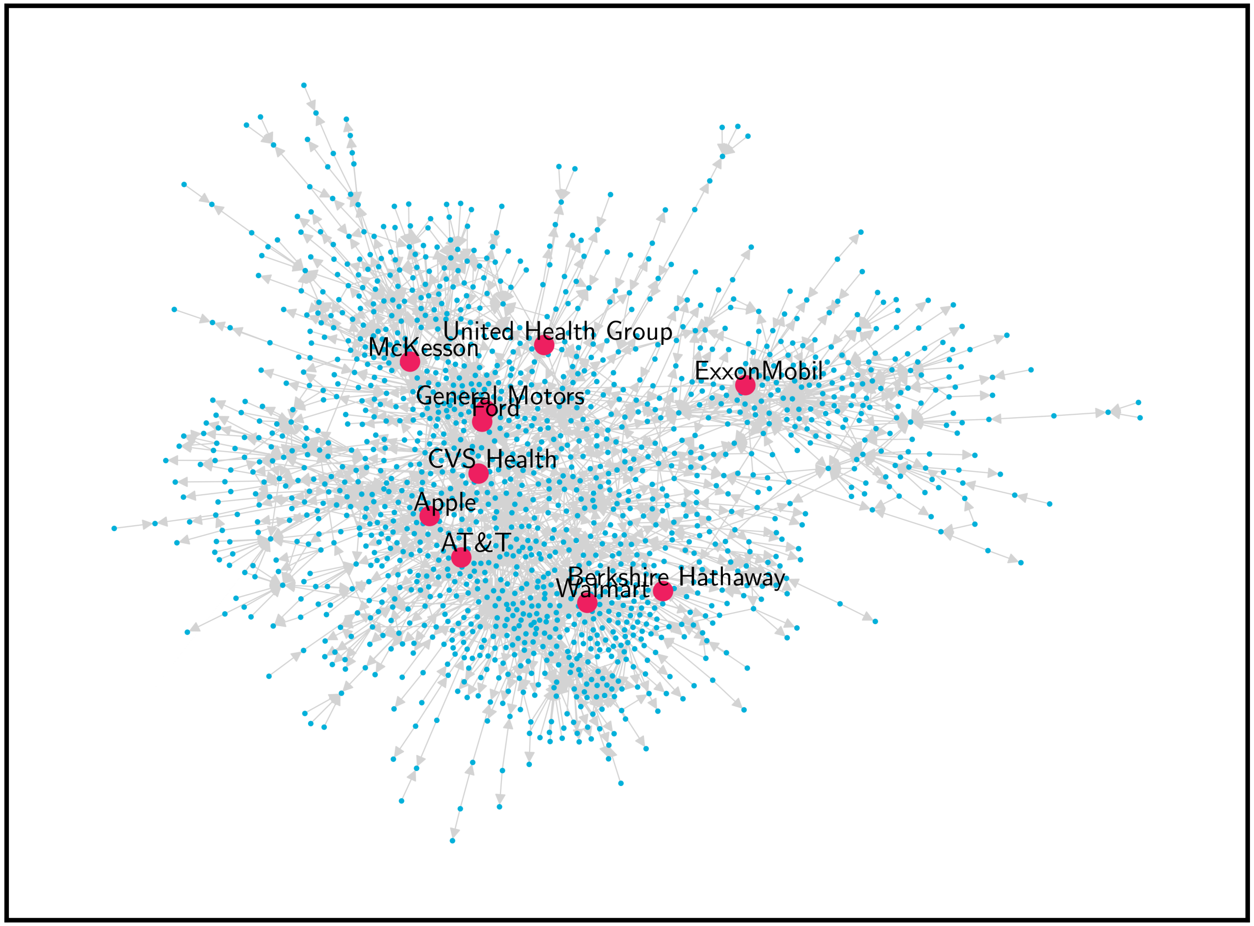
Universität St.Gallen, October 7 to 11, 2024

(Based on "Part "A of an 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).

## 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.

Sources: Graham (2020, *Handbook of Econometrics*), Graham, Niu and Powell (2023, *JOE*)

## 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

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

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

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

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

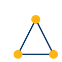
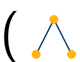
If $\rho_N = \rho$ with $0 < \rho < 1$, then $\lambda_N = O(N)$ and the graph is *dense*.

## Density and Degree Redux

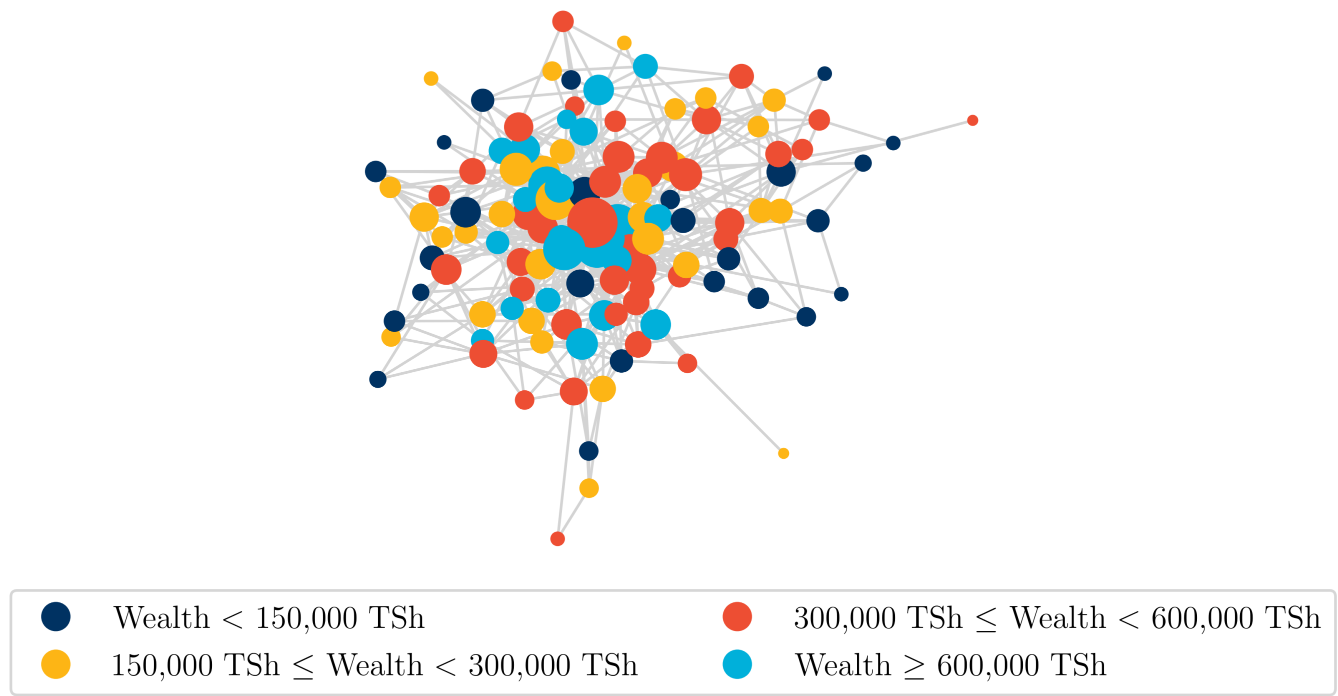
Density and average degree are (arguably) the most basic network summary statistics we can compute.

Understanding how to estimate, and conduct inference on $\rho_N$, 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:

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

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 (2020, *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 $i < j$.

$h_N\left(a_1, a_2\right)$ 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 and Chen, 2009; Lovasz, 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 $h_{1N}(a) = \mathbb{E}\left[h_N(a, A)\right]$,

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

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,

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

consisting of

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

2. $U_{1N}$ : Hájek Projection;

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

## Network Density: Variance

Define the notation:

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

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) $\rho_N$ stabilizes such that, for example, $\tilde{Q}\left(\wedge\right)$ does not vanish as $N \rightarrow \infty$.

Further define:

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

## Network Density: Variance

Applying the variance operator we get

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

Which, after normalizing, yields

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

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

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, 2019).

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

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

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

## 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).

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

In the dense/degenerate case the limit distribution may be non-Gaussian (see Menzel (2017) for examples).

## What did we learn?

Holland and Leinhardt (1976) correctly calculated the variance of $\hat{\rho}_N$ 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 $\hat{\rho}_N$ (except in some very special cases) were available until Bickel, Chen and Levina (2011); whose results I have generalized here.

See Menzel (2017) 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.