Network Moments

Econometric Methods for Networks,

University of Warwick, May 30th to June 1st, 2018

Bryan S. Graham

University of California - Berkeley

This Lecture

We need simple (nonparametric) graph generating processes (GGPs).

Such models facilitate an understanding of the properties of various descriptive network statistics.

Also allow for a rigorous analysis of more complex semiparametric models of network formation.

Network/Graph

$G(\mathcal{V}, \mathcal{E})$ is a finite undirected network or graph defined on

1. $N = |\mathcal{V}(G)|$ *vertices* or agents; here $\mathcal{V}(G) = \{1, \ldots, N\}$ denotes the set of all agents in the network.

2. Any two agents (*dyad*) in G are connected or not; the *edge* list $\mathcal{E}(G) = \{\{i, j\}, \{k, l\}, \ldots\}$, consists of the (unordered) indices of all connected dyads.

Call N the *order* of the network and $|\mathcal{E}(G)|$ its *size.*

Adjacency Matrix

We can represent $G(\mathcal{V}, \mathcal{E})$ by the $N \times N$ adjacency matrix $\mathbf{D} = \left[D_{ij}\right]$ where

$$D_{ij} = \begin{cases} 1, & \{i,j\} \in \mathcal{E}(G) \\ 0, & \text{otherwise} \end{cases} .$$

For an undirected network, with self-ties ruled out, $\mathbf{D}$ is a symmetric binary matrix with a diagonal of structural zeros.

Joint Exchangeability

Let $\pi : \{1, \ldots, N\} \mapsto \{1, \ldots, N\}$ be a permutation of the node labels of $G(\mathcal{V}, \mathcal{E})$ and Π the set of all such permutations.

The random graph G is *jointly exchangeable* if

$$\left[D_{ij}\right] \stackrel{D}{=} \left[D_{\pi(i)\pi(j)}\right] \tag{1}$$

for every permutation $\pi \in \Pi$.

Ignoring node covariates, exchangeability is a feature of most economic networks.

Joint Exchangeability (continued)

The same probability is attached to all isomorphisms.

Motivates a focus on certain network statistics.

Degree sequences and subgraph counts do not vary across isomorphisms

Conditional Edge Independence

Agents are independent random draws from some population.

Conditional on the agent-specific latent variables

$$\mathbf{U} = (U_1, \ldots, U_N)'$$

edges form independently with

$$D_{ij} \Big| \, h(\cdot), U_i, U_j \sim \text{Bernoulli} \left(h\left(U_i, U_j\right) \right),$$

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

Conditional Edge Independence (continued)

Here $h(u,v) = h(v,u)$ for all $(u,v) \in \mathbb{U} \times \mathbb{U}$ is a symmetric edge probability function: **graph** functi**on** or *graphon*.

The conditional likelihood of the network is

$$\Pr(\mathbf{D} = \mathbf{d} | \mathbf{U} = \mathbf{u}) = \prod_{i<j} h\left(u_i, u_j\right)^{d_{ij}} \left[1 - h\left(u_i, u_j\right)\right]^{1-d_{ij}}.$$

Conditional Edge Independence (continued)

Unconditional on $\mathbf{U}$, the likelihood equals

$$\Pr(\mathbf{D}=\mathbf{d}) = \int \cdots \int \left\{ \prod_{i<j} h\left(u_i, u_j\right)^{d_{ij}} \left[1 - h\left(u_i, u_j\right)\right]^{1-d_{ij}} \right\} \quad (2)$$
$$\times \prod_{i=1}^{N} f_U(u_i)\, \mathrm{d}u_i,$$

where $f_U(u)$ is the density of U.

Dependence across dyads which share agents in common.

Independence is conditional on the latent agent attributes.

cf. panel data

Conditional Edge Independence: Examples

β-model with graphon

$$h(u,v) = \frac{\exp(u+v)}{1+\exp(u+v)}$$

Random threshold graphs with graphon

$$h(u,v) = \mathbf{1}(F_U(u) + F_U(v) \geq \alpha),$$

and $F_U(u)$ the CDF of U.

Conditional Edge Independence & Exchangeability

Conditional edge independence models satisfy the finite joint exchangeability condition.

The Aldous-Hoover Theorem implies that the relationship goes in the other direction.

Aldous-Hoover

Strengthen finite exchangeability to infinite exchangeability.

(Aldous-Hoover) A random adjacency matrix $\left[D_{ij}\right]_{i,j\in\mathbb{N}}$ is jointly exchangeable if and only if there is a measureable function $g : [0,1]^4 \to \{0,1\}$ such that

$$\left[D_{ij}\right] \stackrel{D}{=} \left[g\left(\alpha, U_i, U_j, V_{ij}\right)\right]$$

for α, $\{U_i\}_{i\in\mathbb{N}}$,and $\left\{V_{ij}\right\}_{i<j\in\mathbb{N}}$ independently and identically distributed $\mathcal{U}[0,1]$ random variables with $V_{ij} = V_{ji}$.

Here α is a mixing parameter (cf., de Finetti,1931).

Aldous-Hoover (continued)

(Apparently) this representation result is non-trivial (cf., Kallenberg, 2005).

Averaging over V_{ij} yields

$$h\left(\alpha, u_i, u_j\right) \stackrel{def}{=} \int_0^1 g\left(\alpha, u_i, u_j, v\right) \mathrm{d}v$$

from which we get the more convenient representation, for $i < j$,

$$\left[D_{ij}\right] \stackrel{D}{=} \left[\mathbf{1}\left(V_{ij} \leq h\left(\alpha, U_i, U_j\right)\right)\right]. \tag{3}$$

This is, of course, just a conditional edge independence model (or, more precisely, a mixture of such models).

Aldous-Hoover (continued)

The Aldous-Hoover Theorem provides a justification for modeling large exchangeable random graphs using graphons (cf., Menzel, 2018).

This imposes useful structure on the dependence across dyads.

Sparse vs. Dense Graphs

A network is *dense* if its size, or number of edges, is $O\left(N^2\right)$.

It is *sparse* if it size is $O\left(N\right)$.

A sequence of graphs is sparse in the limit if the number of edges in it grows linearly with N...

...dense if this growth is quadratic.

Bickel-Chen Model

Let G_N be a random network of order N generated using some graphon.

A*verage degree* in this network equals

$$\lambda_N = (N-1)\,\rho_\alpha \tag{4}$$

for $\rho_\alpha = \int h\,(\alpha, u, v)\,\mathrm{d}u\mathrm{d}v$.

Average degree (4) either tends toward infinity or is zero, depending on whether ρ_α is greater than or equal to zero.

"Aldous-Hoover" graphs are either dense or empty.

Bickel-Chen Model (continued)

To extend model (3) so that it can accommodate sparse graph sequences Bickel and Chen (2009) define the conditional density

$$w_\alpha(u,v) = f_{U_i,U_j|D_{ij},\alpha}\left(u,v|\, D_{ij}=1,\alpha\right).$$

Next observe that since $f_{U_i,U_j|\alpha}(u,v|\,\alpha) = 1$ on $[0,1]^2$ we can decompose the graphon as

$$h(\alpha,u,v) = \rho_\alpha w_\alpha(u,v). \tag{5}$$

Bickel-Chen Model (continued)

With this parameterization, Bickel and Chen (2009) argue that it is natural to let $\rho_\alpha = \rho_{\alpha,N}$, but retain independence of $w_\alpha(u,v)$ from N.

Suppressing the α argument (it is never identifiable and ignored from here onwards), they write

$$\Pr\left(D_{ij}=1\middle|U_i=u,U_j=v\right)=h_N(u,v)=\rho_N w(u,v)\,\mathbf{1}\left(w(u,v)\leq\rho_N^{-1}\right)$$

with $w(u,v)$ symmetric.

This is restrictive, but useful.

Network Moments

Almost 50 years ago Holland and Leinhardt (1970) introduce the *triad census* to network analysis.

Today, characterizing a network by the frequency with which certain lower order subgraph configurations occur, is common across many fields.

Motifs

(Partial Subgraph) Let $\mathcal{V}(S) \subseteq \mathcal{V}(G)$ be any subset of the vertices of G and $\mathcal{E}(S) \subseteq \mathcal{E}(G) \cap \mathcal{V}(S) \times \mathcal{V}(S)$, then $S = (\mathcal{V}(S), \mathcal{E}(S))$ is an *partial subgraph* of G.

A partial subgraph S of G consists of a subset of agents in G and a *subset* of all edges among $\mathcal{V}(S)$ also appearing in G.

Counts of partial subgraphs are often referred to as *network motif* counts (e.g., Milo et al., 2002), although this terminology is not used consistently.

The two star motif $S =$  is a partial subgraph of $G =$  . Note that in this example S does not include the edge between agents, numbered clockwise from the top, 2 and 3.

Graphlets

(Induced Subgraph) Let $\mathcal{V}(S) \subseteq \mathcal{V}(G)$ be any subset of the vertices of G and $\mathcal{E}(S) = \mathcal{E}(G) \cap \mathcal{V}(S) \times \mathcal{V}(S)$, then $S = (\mathcal{V}(S), \mathcal{E}(S))$ is an *induced subgraph* of G.

An induced subgraph S includes *all* edges in G connecting any two agents in $\mathcal{V}(S)$. Although $S =$  is a partial subgraph of $G =$  , it is not an induced one.

Counts of induced subgraphs are often referred to as *graphlet* counts (e.f., Přzulj et al., 2004), although again not consistently so.

Graph Isomorphism

Consider two graphs, R and S, of the same order. Let $\varphi : \mathcal{V}(R) \to \mathcal{V}(S)$ be a bijection from the nodes of R to those of S.

The bijection $\varphi : \mathcal{V}(R) \to \mathcal{V}(S)$ *maintains adjacency* if for every dyad $i, j \in \mathcal{V}(R)$ if $(i, j) \in \mathcal{E}(R)$, then $(\varphi(i), \varphi(j)) \in \mathcal{E}(S)$.

It *maintains non-adjacency* if for every dyad $i, j \in \mathcal{V}(R)$ if $(i, j) \notin \mathcal{E}(R)$, then $(\varphi(i), \varphi(j)) \notin \mathcal{E}(S)$.

If the bijection maintains both adjacency and non-adjacency we say it *maintains structure*.

Graph Isomorphism (continued)

(Graph Isomorphism) The graphs R and S are *isomorphic* if there exists a structure-maintaining bijection $\varphi : \mathcal{V}(R) \to \mathcal{V}(S)$.

Use the notation $R \cong S$ to denote that "R is isomorphic to S."

p-cycles & trees

A p-*cycle* is p^{th} order graphlet with nodes labeled (or relabeled) such that its edges form a cycle:

$$\mathcal{E}(S) = \{(i_1, i_2), (i_2, i_3), \ldots, (i_p, i_1)\}.$$

A p-cycle is a connected graphlet with p edges on p nodes. Important examples of p-cycles are triangles ($S =$) and 4-cycles ($S =$ ).

A *tree* is a connected graph with no cycles. The number of edges on a p^{th} order tree is $p - 1$.

Important examples of trees are p-star graphlets, such as two-stars ($S =$ ) and three-stars ($S =$ ).

Induced Subgraph Density

Let S be a p^{th}-order graphlet of interest (e.g., $S =$  or $S =$ ), G_N is the real world network under study, and $\mathbf{i}_p \subseteq \{1, 2, \ldots, N\}$ a set of p integers with $i_1 < i_2 < \cdots < i_p$.

The set of all $\binom{N}{p}$ such integer sets is denoted by $\mathcal{C}_{p,N}$.

Finally let $G[\mathbf{i}_p]$ denote the induced subgraph of G associated with vertex set $\mathbf{i}_p$.

Induced Subgraph Density (continued)

Consider the event that the subgraph induced by the first p vertices in G_N *coincides* with S:

$$P_N(S) = \mathsf{Pr}(S = G_N[\{1, 2, \ldots, p\}]). \tag{6}$$

Implicit in (6) is the assumption that S is a labeled isomorphism of the graphlet of interest with vertex set $\{1, 2, \ldots, p\}$.

Call (6) the *induced subgraph density* of S in G_N.

It gives the frequency that an induced subgraph of G_N equals a labeled isomorphism of S.

Induced Subgraph Density (continued)

Let $\mathsf{iso}\,(S)$ be the group of isomorphisms of S, and $|\mathsf{iso}\,(S)|$ its cardinality.

$|\mathsf{iso}\,(S)|$ equals the number of (partial) subgraphs of K_p that are isomorphic to S.

We have that

$$\begin{aligned} P_N(S) &= \frac{1}{\binom{N}{p}\,|\mathsf{iso}\,(S)|} \sum_{\mathbf{i}_p \in \mathcal{C}_{p,N}} \mathbf{1}\left(S \cong G_N[\mathbf{i}_p]\right) \qquad (7) \\ &= \mathsf{Pr}\left(S = G_N[\mathbf{i}_p]\right) \\ &\stackrel{def}{=} t_{\mathsf{ind}}(S, G_N) \end{aligned}$$

Induced Subgraph Density (Examples)

- $t_{\text{ind}}($  ,  $) = \frac{2}{4}$, $t_{\text{ind}}($  ,  $) = \frac{2}{4}$

  and $t_{\text{ind}}($  ,  $) = \frac{0}{4}$

- $t_{\text{ind}}($  ,  $) = \frac{1}{4}$, $t_{\text{ind}}($  ,  $) = \frac{2}{4}$

  and $t_{\text{ind}}($  ,  $) = \frac{1}{4}$

Induced Subgraph Density (continued)

Let $\mathbf{D}_{[\mathbf{i}_p,\mathbf{i}_p]}$ be the $p \times p$ sub-adjacency matrix constructed by removing all rows and columns of $\mathbf{D}$ except those in $\mathbf{i}_p = \{i_1, \ldots, i_p\}$.

Let S be a graphlet of interest, defined on vertices $\mathbf{i}_p$.

We can check for whether S is an isomorphism of $G[\mathbf{i}_p]$ by inspecting the elements of the $\mathbf{D}_{[\mathbf{i}_p,\mathbf{i}_p]}$ sub-adjacency matrix.

Consider the two star triad $S =$  , we can express $\mathbf{1}(S \cong G_N[\mathbf{i}_p])$ in terms of $\mathbf{D}_{[\mathbf{i}_p,\mathbf{i}_p]}$ as

$$\begin{aligned}\mathbf{1}\left(\text{  } \cong G_N[\mathbf{i}_p]\right) =& D_{i_1i_2}D_{i_1i_3}\left(1 - D_{i_2i_3}\right) \\ &+ D_{i_1i_2}\left(1 - D_{i_1i_3}\right)D_{i_2i_3} + \left(1 - D_{i_1i_2}\right)D_{i_1i_3}D_{i_2i_3}.\end{aligned} \tag{8}$$

Induced Subgraph Density (continued)

Assume an Aldous-Hoover GGP.

Consider the two star configuration; iterated expectations and conditional independence of edges given $\mathbf{U} = (U_1, \ldots, U_N)'$ yield

$$\begin{aligned}
\mathbb{E}\left[D_{i_1 i_2} D_{i_1 i_3}\left(1 - D_{i_2 i_3}\right)\right] &= \mathbb{E}\left[\mathbb{E}\left[D_{i_1 i_2} D_{i_1 i_3}\left(1 - D_{i_2 i_3}\right)\middle|\, \mathbf{U}\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[D_{i_1 i_2} D_{i_1 i_3}\left(1 - D_{i_2 i_3}\right)\middle|\, U_{i_i}, U_{i_2}, U_{i_3}\right]\right] \\
&= \mathbb{E}\left[h\left(U_{i_1}, U_{i_2}\right) h\left(U_{i_1}, U_{i_3}\right)\left[1 - h\left(U_{i_2}, U_{i_3}\right)\right]\right] \\
&= \int\int\int h\left(t, u\right) h\left(t, v\right)\left[1 - h\left(u, v\right)\right] \mathrm{d}t \mathrm{d}u \mathrm{d}v
\end{aligned}$$

Induced Subgraph Density (continued)

The value of $\mathbb{E}\left[D_{i_1 i_2} D_{i_1 i_3}\left(1 - D_{i_2 i_3}\right)\right]$ is invariant to permutations of its indices.

Recalling that $|\text{iso}($ $)| = 3$, finally gives

$$\mathbb{E}\left[\mathbf{1}\left(\text{  } \cong G_N\left[\mathbf{i}_p\right]\right)\right] = 3 \cdot \int\int\int h\left(t,u\right) h\left(t,v\right) \left[1 - h\left(u,v\right)\right] \mathrm{d}t\mathrm{d}u\mathrm{d}v.$$

We can use the graphon to calculate expectations of $P_N\left(S\right)$.

Induced Subgraph Density (continued)

For a generic graphlet configuration we have

$$\begin{aligned}\mathbb{E}\left[t_{\mathsf{ind}}\left(S, G_N\right)\right] &= \left|\mathsf{iso}\left(S\right)\right|^{-1} \mathbb{E}\left[\mathbf{1}\left(S \cong G_N\left[\mathbf{i}_p\right]\right)\right] \\ &= \mathbb{E}\left[\prod_{\{i,j\}\in\mathcal{E}(S)} h\left(U_i, U_j\right) \prod_{\{i,j\}\in\mathcal{E}(\bar{S})} \left[1 - h\left(U_i, U_j\right)\right]\right] \\ &\stackrel{def}{\equiv} P\left(S\right)\end{aligned} \tag{9}$$

where $\bar{G}$ denotes the complement of the graph G.

Note the role placed by the Aldous-Hoover structure (i.e., graphon, conditional independence etc.).

Graph Limits

Let $\{G_N\}_{N=1}^{\infty}$ be a sequence of networks. If

$$\lim_{N\to\infty} t_{\text{ind}}(S, G_N) = t_{\text{ind}}(S, h)$$

for some graphon $h(\cdot,\cdot)$ and *all* fixed subgraphs S, then we say that G_N converges to $h(\cdot,\cdot)$.

- Lovász (2012) for complete development.
- Diaconis and Janson (2008) for connections with Aldous-Hoover Theorem.
- Result establishes a connection between subgraph counts and the graphon.

Goal

Using the infrastructure outlined above...

...we would like a result of the form...

$$\sqrt{N}\left(\left(\begin{array}{c} P_N\left(\begin{array}{c}\\ \end{array}\right) \\ P_N\left(\begin{array}{c}\\ \end{array}\right) \end{array}\right) - \left(\begin{array}{c} P\left(\begin{array}{c}\\ \end{array}\right) \\ P\left(\begin{array}{c}\\ \end{array}\right) \end{array}\right)\right) \xrightarrow{D} \mathcal{N}(0, \Sigma))$$

...under conditions we can understand

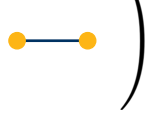
...with a covariance Σ we can estimate

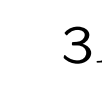
Bickel, Chen and Levina (2011) and Bhattachary and Bickel (2015) provide such a result.

Goal (continued)

With this result we can conduct inference on *transitivity...*

$$\text{TI} = \frac{3P\left(\ \right)}{P\left(\ \right) + 3P\left(\ \right)}$$





Is TI $> P\left(\ \right)$ (see Jackson et al. (2012) for some motivation)?

cf., Blitzstein and Diaconis (2011)

Network Moments: Sampling Distribution

The goal is to characterize the distribution of induced subgraph frequencies across replications of networks generated using a certain graphon.

The sample or experiment consists of a *single network realization*.

The *sampling distribution* of an empirical subgraph frequency refers to its random variation across independent replications of networks generated using a common graphon.

The *limiting distribution* refers to the distribution that characterizes this frequency – appropriately normalized – as the number of agents, N, in the network grows large.

Network Moments: Estimation

The obvious estimate is, for example,

$$P_N\left(\text{  }\right) = \binom{N}{3}^{-1} \frac{1}{3} \sum_{\mathbf{i}_3 \in \mathcal{C}_{3,N}} \left[D_{i_1 i_2} D_{i_1 i_3} \left(1 - D_{i_2 i_3}\right) + D_{i_1 i_2} \left(1 - D_{i_1 i_3}\right) D_{i_2 i_3} + \left(1 - D_{i_1 i_2}\right) D_{i_1 i_3} D_{i_2 i_3} \right]$$

This estimate is *unbiased*

$$\mathbb{E}\left[P_N\left(S\right)\right] = P\left(S\right)$$

Network Moments: Large Network Variance

Holland and Leinhardt (1976) provided an expression for the sampling variance of subgraph counts.

They used Hoeffding (1948) arguments to study the variance-covariance of network moments.

We can use the probability infrastructure established above to put their calculations on firmer footing (cf., Bhattacharya and Bickel, 2015).

Network moments: Large N behavior

Projecting, for example, $P_N\left(\triangle\right)$ on $\mathbf{U} = (U_1, \ldots, U_N)'$ gives:

$$\begin{aligned} P_N\left(\triangle\right) =& \binom{N}{3}^{-1} \sum_{i_1<i_2<i_3} h\left(U_{i_1}, U_{i_2}\right) h\left(U_{i_1}, U_{i_3}\right) h\left(U_{i_2}, U_{i_3}\right) \\ &+ \binom{N}{3}^{-1} \sum_{i_1<i_2<i_3} \Big\{ D_{i_1 i_2} D_{i_1 i_3} D_{i_2 i_3} \\ &- h\left(U_{i_1}, U_{i_2}\right) h\left(U_{i_1}, U_{i_3}\right) h\left(U_{i_2}, U_{i_3}\right) \Big\}. \end{aligned}$$

Second term is mean independent of first with conditionally independent summands.

First term is a 3^{rd} order U-Statistic.

Network moments: Large N behavior (continued)

Under some conditions (most important of which is that average degree grows with N) $\hat{P}_N\left(\text{  }\right)$ behaves like a U-Statistic s.t.

$$\sqrt{N}\left(\begin{pmatrix} P_N\left(\text{  }\right) \\ P_N\left(\text{  }\right) \end{pmatrix} - \begin{pmatrix} P\left(\text{  }\right) \\ P\left(\text{  }\right) \end{pmatrix}\right) \xrightarrow{D} \mathcal{N}\left(0, 9\Sigma_1\right)$$

...with Σ_1 estimable (analog estimate involves $O\left(N^5\right)$ operations!).

Use delta method to conduct inference on transitivity.

Network moments: Large N behavior (continued)

Depending on subgraph and ρ_N, other terms in the Hoeffding variance decomposition may be of equal order to the first one.

In some cases all terms are of equal order.

Makes variance estimation and establishing asymptotic normality more complicated.

If graphon is constant in U_i and U_j (random graph case), then $\Sigma_1 = 0$!

Transitivity in Nyakatoke

For the Nyakatoke network we have

$$P_N\left(\triangle\right) = \underset{(0.00030)}{0.00115}, \quad P_N\left(\wedge\right) = \underset{(0.00100)}{0.00496}$$

The transitivity index in Nyakatoke is 0.188, almost three times the magnitude of its density ($P_N\left(\text{—}\right) = 0.070$).

Applying the delta method I get an estimate standard error for the transitivity index of 0.011.

Transitivity is significantly greater than what we would expect to observe under the Erdös-Renyi random graph null.

Intellectual history

Some basic ideas (e.g., use of Hoeffding-like variance decompositions) go back (at least) to Holland and Leinhardt (1976).

Subsequent work by Nowicki (1991), Picard et al. (2008) and others.

Big breakthrough by Bickel et al. (2011) – abstract (proof uses lots of "tricks") and limiting variance is not characterized.

Bhattacharya and Bickel (2015) – explicit characterization of variance and an estimator (cf., Menzel, 2017).

Some (interesting and empirically-relevant) subtleties ignored today.

Intellectual history (continued)

My exposition (anchored in textbook U-Statistic theory) is based on basic approach of Graham (2017).

Challenge is finding a notation that can neatly handle all cases.

Some open questions regarding sparse graph sequences.

Wrapping Up

Exchangeability + Aldous Theorem + Graph Limit Theory motivates a graphon based approach to modeling large exchangeable networks.

These tools provide rigorous foundation for studying the large network properties of so called network moments.

Network moments underlie many foundational network statistics (e.g., transitivity).

The degree distribution can also be described in terms of such moments (not shown).

...building blocks for dyadic regression.