Beta Model & Network Simulation

Econometric Methods for Networks,

CORE, Dec 12th - 14th, 2016

Bryan S. Graham

University of California - Berkeley

Introduction

- In practice methods of inference used in "network science" are decidedly heuristic/approximate.

- Widespread awareness among researchers about the lack of a coherent body of large network distribution theory.

Introduction (continued)

- Is the network in hand especially "transitive"?

  - Compare with Erdos-Renyi random graph; but are measured differences statistically significant? (how do we conceptualize "sampling error")

  - Compare with a large set of (empirical) reference graphs. Is the graph of interest usual?

  - Combine an ad hoc and/or approximate variance estimate with a normal reference distribution (w/o limit theory it is difficult to evaluate this approach; see last lecture)

Inference: Exact

- Blitzstein and Diaconis (2011) − additional work in both machine learning and statistics

- Look at a reference set of graphs (e.g., all graphs with degree sequences identical to the graph of interest)

  − Is transitivity (for example) in the graph at hand high relative to this reference group? (exact p-value approach)

  − Computational challenge: how to enumerate, or draw uniformly, from reference graph distribution

Inference: Asymptotic

- <u>Last lecture</u>: Bickel, Chen & Levina and Bhattacharya and Bickel (2015):

  - Derive limit theory for network statistics (specifically normalized subgraph counts)

  - Challenge is also computational − both statistics and their variance estimates hard to construct

Beta Model

- Models with network externalities are attractive because

  – they capture what is believed to be an *a priori* important feature of link formation,

  – they generate clustering, which we observe in real word networks.

Beta Model (continued)

- An alternative (ideally complementary) way to generate clustering is to introduce unobserved, agent-level, heterogeneity.

  - beta model: $D_{ij} = \mathbf{1}\left(A_i + A_j - U_{ij} \leq 0\right)$

  - A_i measures attractiveness, trustworthiness, productivity etc.

  - Distribution of $\mathbf{A}$ is unrestricted; components of $\mathbf{U}$ are i.i.d. (logistic).

- cf. "state dependence vs. heterogeneity" in dynamic discrete choice analysis (Heckman, 1978; 1981a-c; Chamberlain, 1985).

Beta Model (continued)

- Assuming U_{ij} i.i.d. logistic yields a link probability of

$$\Pr\left(D_{ij} = 1 \,\middle|\, \mathbf{A}\right) = \frac{\exp\left(A_i + A_j\right)}{1 + \exp\left(A_i + A_j\right)} = \frac{\exp\left(W_{ij}'\mathbf{A}\right)}{1 + \exp\left(W_{ij}'\mathbf{A}\right)}$$

  with W_{ij} the $N \times 1$ vector with a one for its ith and j^{th} elements and zeros elsewhere.

- Choosing $A_i = -\frac{1}{2}\ln\left(\frac{p}{1-p}\right)$ for $i = 1, \ldots, N$ yields the Erdos-Renyi random graph model.

Beta Model (continued)

- The likelihood, $\Pr\left(\mathbf{D} = \mathbf{d} \mid \mathbf{A}\right)$, includes $\binom{N}{2}$ conditional independent components

$$\prod_{i=1}^{N} \prod_{j<i} \left[\frac{\exp\left(W'_{ij}\mathbf{A}\right)}{1 + \exp\left(W'_{ij}\mathbf{A}\right)} \right]^{d_{ij}} \left[\frac{1}{1 + \exp\left(W'_{ij}\mathbf{A}\right)} \right]^{1-d_{ij}} .$$

- ...but all N parameters

- model is non-standard since the dimension of the parameter space grows with N

Beta Model (continued)

- Manipulating the likelihood gives the exponential family representation

$$\Pr\left(\mathbf{D} = \mathbf{d}\,|\, \mathbf{A}\right) = c\left(\mathbf{Z}; \mathbf{A}\right) \exp\left(T\left(\mathbf{d}\right)' \mathbf{A}\right) \qquad (1)$$

  where

$$T\left(\mathbf{d}\right) \;=\; \left(\begin{array}{ccc} d_{1+} & \cdots & d_{N+} \end{array}\right)' = \mathbf{d}_{+}.$$

- The network's *degree sequence*, is a sufficient statistic for $\mathbf{A}$.

Beta Model (continued)

- The beta model allows for networks with arbitrary degree distributions.

- Despite its simplicity it is reasonable flexible (N parameters) and provides a useful benchmark.

Beta Model (continued)

- Let $\mathbb{D}_{N,\mathbf{d}_+}$ denote the set of all networks with N agents and degree sequence $\mathbf{D}_+ = \mathbf{d}_+$.

- Let $\left| \mathbb{D}_{N,\mathbf{d}_+} \right|$ denote the cardinality of $\mathbb{D}_{N,\mathbf{d}_+}$.

  - $\left| \mathbb{D}_{N,\mathbf{d}_+} \right|$ is generally huge, even for small N.

- Under the β-model the probability distribution of networks conditional on their degree sequence is uniform:

$$\Pr\left(\mathbf{D} = \mathbf{d} \mid \mathbf{d} \in \mathbb{D}_{N,\mathbf{d}_+} \right) = \frac{1}{\left| \mathbb{D}_{N,\mathbf{d}_+} \right|}.$$

Testing

- Let $f(\mathbf{D})$ be some function of the adjacency matrix

  - examples: transitivity index, diameter, number of K-length paths etc.

- Let $f(\mathbf{d})$ be the value of the statistic in the observed network.

- We week to evaluate

$$\Pr\left(f\left(\mathbf{D}\right) \leq f\left(\mathbf{d}\right)\right) = \frac{\sum_{\mathbf{v} \in \mathbb{D}_{\mathbf{N},\mathbf{d}_+}} \mathbf{1}\left(f\left(\mathbf{v}\right) \leq f\left(\mathbf{d}\right)\right)}{\left|\mathbb{D}_{N,\mathbf{d}_+}\right|}. \qquad (2)$$

Testing: Intuition

- If the probability that measured transitivity, in a network randomly drawn from the null distribution, lies above observed transitivity is very low, we take that as evidence against the β-model.

Testing

- This approach to testing is

  - very precise about its description of the null hypothesis

  - exact

- no alternative hypothesis is specified...

- ...however the choice of statistic should be guided by researcher intuitions about what departures from the null model are of concern.

Sampling from $\mathbb{D}_{N,\mathbf{d}_+}$

- Direction enumeration of all the elements of $\mathbb{D}_{N,\mathbf{d}_+}$ is generally not feasible.

- Need a method of sampling from $\mathbb{D}_{N,\mathbf{d}_+}$ uniformly and also estimating its size (implement an approximation of the ideal test).

Sampling from $\mathbb{D}_{N,\mathbf{d}_+}$ (continued)

- Blitzstein and Diaconis (2010) develop a sequential importance sampling algorithm for uniformly sampling from $\mathbb{D}_{N,\mathbf{d}_+}$

- Two challenges:

  - how to generate a random draw from $\mathbb{D}_{N,\mathbf{d}_+}$

  - how to do so uniformly (importance weights)

Graphical Integer Sequences

- To construct $\mathbf{D}$ we begin with a matrix of zeros and sequentially add links to it until its rows and columns sum to the target degree sequence.

- Problem is that unless links are added carefully it is easy to get "stuck" (cf., Snijders, 1991).

- The key is to check whether residual degree sequences are graphical as you add links (avoid dead ends).

- $\mathbf{D}_+ = (2, 2, 1)$ is not graphic

Graphical Integer Sequences (continued)

- Erdos and Gallai (1961) showed $\mathbf{D}_+$ is graphical if and only if $\sum_{i=1}^{N} D_{i+}$ is even and

$$\sum_{i=1}^{k} D_{i+} \leq k(k-1) + \sum_{i=k+1}^{N} \min\left(k, D_{i+}\right) \text{ for each } k \in \{1, \ldots, N\}.$$

Graphical Integer Sequences (continued)

Necessity:

- <u>even</u>: if i is linked to j, then the link is counted in both D_{i+} and D_{j+}

- For any set S of k agents, there can be at most $\binom{k}{2} = \frac{1}{2}k(k-1)$ links between them (first term).

- For the $N - k$ agents $i \notin S$, then can be at most $\min\left(k, D_{i+}\right)$ links for i to agents in S.

A Recursive Test

Theorem: *(Havel-Hakimi) Let $D_{i+} > 0$, if $\mathbf{D}_+$ does not have at least D_{i+} positive entries other than i it is not graphical. Assume this condition holds. Let $\tilde{\mathbf{D}}_+$ be a degree sequence of length $N-1$ obtained by*

   [i]   deleting the ith entry of $\mathbf{D}_+$ and

   [ii]   subtracting 1 from each of the D_{i+} highest elements in $\mathbf{D}_+$ (aside from the ith one).

$\mathbf{D}_+$ *is graphical if and only if $\tilde{\mathbf{D}}_+$ is graphical. If $\mathbf{D}_+$ is graphical, then it has a realization where agent i is connected to any of the D_{i+} highest degree agents (other than i).*

Blitzstein and Diaconis Procedure

- Start with lowest degree agent (with at least one link).

- (Randomly) Link this agent with high degree agents.

- A one is subtracted from the lowest degree agent's and her chosen partner's degree counts.

- Continue until the **residual degree sequence** is zero.

Blitzstein and Diaconis Procedure (continued)

- Consider the example

$$(3, 2, 2, 2, 1) \quad \rightarrow \quad (3, 1, 2, 2, 0) \rightarrow (2, 0, 2, 2, 0)$$
$$\rightarrow (1, 0, 2, 1, 0) \quad \rightarrow \quad (0, 0, 1, 1, 0) \rightarrow (0, 0, 0, 0, 0)$$

- Now imagine that in the fourth step instead of linking agent 1 with agent 3, agents 1 and 4 were linked.

  - this would have resulted in a residual degree sequence of $(0, 0, 2, 0, 0)$, which is not graphic.

  - algorithm doesn't allow this to occur by checking for whether the residual degree sequence associated with a candidate link is graphical.

Blitzstein and Diaconis Procedure (continued)

- Let $\left(\oplus_{i_1,\dots,i_k}\mathbf{D}_+\right)$ be the vector obtained by adding a one to the $i_1,\dots,i_k$ elements of $\mathbf{D}_+$:

$$\left(\oplus_{i_1,\dots,i_k}\mathbf{D}_+\right)_j = \begin{cases} D_{j+} + 1 & \text{for } j \in \{i_1,\dots,i_k\} \\ D_{j+} & \text{otherwise} \end{cases}$$

- Let $\left(\ominus_{i_1,\dots,i_k}\mathbf{D}_+\right)$ be the vector obtained by subtracting one from the $i_1,\dots,i_k$ elements of $\mathbf{D}_+$:

$$\left(\ominus_{i_1,\dots,i_k}\mathbf{D}_+\right)_j = \begin{cases} D_{j+} - 1 & \text{for } j \in \{i_1,\dots,i_k\} \\ D_{j+} & \text{otherwise} \end{cases}$$

Blitzstein and Diaconis Procedure (continued)

Algorithm: A sequential algorithm for constructing a random graph with degree sequence $\mathbf{D}_+ = \left(D_{1+}, \ldots, D_{N+}\right)'$ is

1. Let $\mathbf{G}$ be an empty adjacency matrix.

2. If $\mathbf{D}_+ = \mathbf{0}$ terminate with output $\mathbf{G}$

3. Choose the agent i with minimal positive degree D_{i+}.

4. Construct a list of candidate partners

$$J = \left\{ j \neq i \ : \ \mathbf{G}_{ij} = \mathbf{G}_{ji} = 0 \text{ and } \ominus_{i,j} \mathbf{D}_+ \text{ graphical} \right\}$$

5. Pick a partner $j \in J$ with probability proportional to its degree in $\mathbf{D}_+$.

6. Set $\mathbf{G}_{ij} = \mathbf{G}_{ji} = 1$ and update $\mathbf{D}_+$ to $\ominus_{i,j}\mathbf{D}_+$.

7. Repeat steps 4 to 6 until the degree of agent i is zero.

8. Return to step 2.

The input for the algorithm is the target degree sequence $\mathbf{D}_+$ and the output is an undirected adjacency matrix $\mathbf{G}$ with $\mathbf{G}'\iota = \mathbf{D}_+$.

Importance Weights

- The Blitzstein and Diaconis (2010) procedure delivers a random draw from $\mathbb{D}_{N,\mathbf{d}_+}$, but not a *uniform* random draw.

- Construct importance weights in order to compute expectations using the correct reference distribution.

- Let $\mathcal{Y}_{N,\mathbf{d}_+}$ denote the set of all possible sequences of links generated by the algorithm given input $\mathbf{D}_+ = \mathbf{d}_+$.

Importance Weights(continued)

- Let $\mathcal{G}(Y)$ be the adjacency matrix induced by link sequence Y.

  - Let Y and Y' are equivalent if $\mathcal{G}(Y) = \mathcal{G}(Y')$.

- We can partition $\mathcal{Y}_{N,\mathbf{d}_+}$ into a set of equivalence classes whose number coincides with the cardinality of $\mathbb{D}_{N,\mathbf{d}_+}$.

Importance Weights(continued)

- Let $c(Y)$ denote the number of possible link sequences produced by the algorithm that produce Y's end point adjacency matrix.

- Let $i_1, i_2, \ldots, i_M$ be the sequence of agents chosen in step 3 of the algorithm in which Y is the output.

Importance Weights(continued)

- Let $a_1, \ldots, a_m$ be the degree sequences of $i_1, \ldots, i_M$ at the time when each agent was *first* selected in step 3.

- Blitzstein and Diaconis show that:

$$c(Y) = \prod_{k=1}^{M} a_k!$$

Importance Weights(continued)

- Let $\sigma(Y)$ be the probability that the algorithm produces link sequence Y.

- $\sigma(Y)$ is easy to compute:

  - each time a link in step 5 is chosen record the probability with which it was chosen.

  - this equals the residual degree of the chosen agent divided by the sum of the residual degrees of all agents in the choice set.

  - the product of all these probabilities equals $\sigma(Y)$.

Importance Weights(continued)

- Let $f(\mathbf{G})$ be some function of the adjacency matrix and consider the expected value

$$
\begin{aligned}
\mathbb{E}\left[\frac{\pi(\mathcal{G}(Y))}{c(Y)\sigma(Y)}f(\mathcal{G}(Y))\right] &= \sum_{y\in\mathbb{Y}_{N,\mathbf{d}}}\frac{\pi(\mathcal{G}(y))}{c(y)\sigma(y)}f(\mathcal{G}(y))\sigma(y) \\
&= \sum_{y\in Y_{N,\mathbf{d}}}\frac{\pi(\mathcal{G}(y))}{c(y)}f(\mathcal{G}(y)) \\
&= \sum_{\mathbf{g}\in\mathbb{D}_{N,\mathbf{d}}}\sum_{\{y:G(y)=\mathbf{g}\}}\frac{\pi(\mathbf{g})}{c(y)}f(\mathbf{g}) \\
&= \sum_{\mathbf{g}\in\mathbb{D}_{N,\mathbf{d}}}\pi(\mathbf{g})f(\mathbf{g}) \\
&= \mathbb{E}_{\pi}[f(\mathbf{G})].
\end{aligned}
$$

Importance Weights (continued)

- The ratio $\pi\left(\mathbf{G}\left(Y_t\right)\right)/c\left(Y_t\right)\sigma\left(Y_t\right)$ is called the likelihood ratio or the **importance weight.**

- Our random network draws are not uniform from the set of interest

- The importance weights correct for the fact that we are sampling from the wrong distribution

Blitzstein and Diaconis Wrap-Up

- While using the β-model as a reference model is restrictive it

  - is a natural starting point for hypothesis testing.

  - suggests that an investment in computation skills is likely to be valuable to anyone doing empirical work.

- It might be of interest to condition on additional features of the network in hand...