An Brief Overview of Empirical Models of Network Formation

Econometric Methods for Networks,
CORE, Dec 12th - 14th, 2016

Bryan S. Graham

University of California - Berkeley

Overview

- Why study models of network formation?
  - equilibrium $\mathbf{D}$ may be inefficient.
  - planner may have preferences over $\mathbf{D}$ and hence is interested in policies which influence it.
  - we might view network manipulations as a mechanism for influencing other outcomes.
  - correct for "endogenous network formation bias" (cf., Auerbach, 2016)

Existing Approaches

- "Applied theory approach": posit generative models of network formation that match "stylized facts" (Albert and Barabasi, 2002).

- ERGM: directly write down likelihoods for $\Pr(\mathbf{D} = \mathbf{d})$ and try to maximize them (cf., Shalizi and Rinaldo, 2013, *Annals of Statistics*).
  - generally no micro-foundations...
  - ...but see Mele (2016)

Existing Approaches

- Random Utility Models (RUM): Sheng (2012), Christakis *et al.* (2012), Imbens and Goldsmith-Pinkham (2013), Graham (2013, 2014, 2016), de Paula et al. (2015)

- Specialized structural modes: Banerjee *et al.* (2012).

Random Utility Approach

- This approach is both natural, and familiar, to economists

- Provides a framework for modelling the effect on link surplus (i.e., utility) of
  - observed agent/dyad covariates
  - unobserved agent attributes (heterogeneity)
  - preference interdependencies (e.g., a taste for transitivity)

A Simple Model of Network Formation

- Consider a network of three agents: i, j, k
  - Link formation: $D_{ij} = \mathbf{1}\left(\alpha + \beta D_{ik} D_{jk} - U_{ij} \geq 0\right)$ with $\beta \geq 0$ (returns to transitivity).
  - Three "types" of U_{ij} draws: $\mathbb{U}_L = (-\infty, \alpha]$, $\mathbb{U}_M = (\alpha, \alpha + \beta]$ or $\mathbb{U}_H = (\alpha + \beta, \infty)$.
  - Positive measure on the subset of the support of $\mathbf{U} = \left(U_{ij}, U_{ik}, U_{jk}\right)'$ with multiple NE networks.
- The model is *incomplete* (cf., Bresnahan and Reiss, 1991; Tamer, 2003).

A Simple Model of Network Formation (continued)

- There are $3^3 = 27$ "configurations" of $\mathbf{U}$...

- ...but only $\frac{(3+3-1)!}{3!(3-1)!} = 10$ non-isomorphic ones

- Two of these configurations admit multiple NE networks
  - if $U_{ij} \in \mathbb{U}_L$, $U_{ik} \in \mathbb{U}_M$ and $U_{jk} \in \mathbb{U}_M$
  - if $U_{ij} \in \mathbb{U}_M$, $U_{ik} \in \mathbb{U}_M$ and $U_{jk} \in \mathbb{U}_M$

A Simple Model of Network Formation (continued)

- Only one realization (out of 4 possible realizations of $\mathbf{D}$) is uniquely predicted
  - if $U_{ij} \in \mathbb{U}_L$, $U_{ik} \in \mathbb{U}_L$ and $U_{jk} \in \mathbb{U}_H$, then links $\{i,j\}$ and $\{i,k\}$ form and $\{j,k\}$ does not

- cf., Ciliberto and Tamer (2009), Sheng (2012), de Paula et al. (2015) provide methods for analyzing incomplete models of network formation

- serious challenges to implementation at scale

Christakis, Fowler, Imbens and Kalyanaraman (2010)

- Dyads form links *sequentially* and *myopically.*

- If the linking order is ij, ik and jk we have

  - $D_{ij} = \mathbf{1}\left(\alpha - U_{ij} \geq 0\right)$

  - $D_{ik} = \mathbf{1}\left(\alpha - U_{ik} \geq 0\right)$

  - $D_{jk} = \mathbf{1}\left(\alpha + \beta \mathbf{1}\left(\alpha - U_{ij} \geq 0\right) \mathbf{1}\left(\alpha - U_{ik} \geq 0\right) - U_{jk} \geq 0\right)$

- Conditional on the ij, ik and jk the realization of $\mathbf{U}$ delivers a unique prediction of $\mathbf{D}$.

Christakis, Fowler, Imbens and Kalyanaraman (2010)

- Since we don't observe the order of link formation we
  - assign a (prior) distribution to it and
  - work with an integrated likelihood.

- With three dyads there are $3! = 6$ possible link orderings. Let $O \in \mathbb{O} = \{1, 2, 3, 4, 5, 6\}$ be the possible orderings. Our integrated likelihood is

$$\Pr(\mathbf{D} = \mathbf{d}) = \sum_{o \in \mathbb{O}} \Pr(\mathbf{D} = \mathbf{d} | O = o) \Pr(O = o)$$

Christakis, Fowler, Imbens and Kalyanaraman (2010)

- Christakis et al. (2010) use Bayesian MCMC methods
  - provides a method of inference as well
  - (no large sample theory for their estimator)
- Simulation methods, and assumptions about the timing of link formation, are also central to work by Goldsmith-Imbens and Pinkham (2013) and Mele (2016)

Graham (2014)

- Simple model of dyadic link formation with *unobserved* agent-level degree heterogeneity

- Related to the β-model introduced in Lecture 4

- Natural generalization of modelling approaches currently used in empirical work

- Model does not allow for network interdependencies

Graham (2016)

- Focuses of discriminating between *homophilous sorting* on unobserved attributes and a structural taste for *transitivity*
  - both effects generated clustered links
  - heterogeneity and interdependencies
  - timing used to side-step issues of incompleteness

- Uses availability of multiple network observations over time