Friendship Formation in a Network Context

ENTER Jamboree Presentation

A. Advani

University College London and Institute for Fiscal Studies

11 Feb 2012

Motivation

- Connections might influence behaviour.
- Causality needs exogeniety (pre-determinedness).
- My solution: learn likelihood of connection from data.
- Key challenge: single cross-section of data, possibly with missing links.

Literature - Network Empirics

Three classes of empirical work:

1. "**Fixed network**"
   - **Problem:** Selection?

Literature - Network Empirics

Three classes of empirical work:

1. "**Fixed network**"
   - **Problem:** Selection?

2. "**Randomised network**"
   - **Problem:** Success of randomisation?
   - **Problem:** Low external validity.

Literature - Network Empirics

Three classes of empirical work:

1. "**Fixed network**"
   - **Problem:** Selection?

2. "**Randomised network**"
   - **Problem:** Success of randomisation?
   - **Problem:** Low external validity.

3. "**Structural network**"
   - **Problem:** Capture all the features?

Network Model (1) - Preferences

Assumption 1: Agent i receives a benefit from having a path to another agent j.
Magnitude depends on length of shortest path from i to j in network $\mathbf{g}$ – $d(i, j; \mathbf{g})$ – with values $\delta_{d(i,j;\mathbf{g})}$.

Network Model (1) - Preferences

Assumption 1: Agent i receives a benefit from having a path to another agent j.
Magnitude depends on length of shortest path from i to j in network $\mathbf{g}$ – $d(i, j; \mathbf{g})$ – with values $\delta_{d(i,j;\mathbf{g})}$.

Assumption 2: Link formation is costly.

Network Model (1) - Preferences

Assumption 1: Agent i receives a benefit from having a path to another agent j.
Magnitude depends on length of shortest path from i to j in network $\mathbf{g}$ – $d(i,j;\mathbf{g})$ – with values $\delta_{d(i,j;\mathbf{g})}$.

Assumption 2: Link formation is costly.

Generalised "connections" model of utility:

$$u_i(\mathbf{g}) = \sum_{j \neq i} \delta_{d(i,j;\mathbf{g})} - \sum_{j \epsilon N_i(\mathbf{g})} c_{ij}(\mathbf{g}) \qquad (1)$$

Network Model (2) - Net Costs

Assumption 3: The net cost to i of link formation with j depends on some own characteristics, the similarity of some of their characteristics, and individual-specific costs.

$$c_{ij}(\mathbf{g}) = c^* + \mathbf{x}_i'\beta + (\mathbf{z}_i - \mathbf{z}_j)'\Omega(\mathbf{z}_i - \mathbf{z}_j) + \nu_i + \nu_j + \varepsilon_{ij} \qquad (2)$$

where:

- c^*: constant,
- $\mathbf{x}_i$: vector of observables for individual "paying" for the link,
- $\mathbf{z}_j$: vector of observable characteristics for j,
- ν_j: unobserved individual-specfic net costs for j, and
- ε_{ij} is an unobserved link-specific net cost.

Network Model (3) - Friendship Formation Process

- Individuals meet in pairs.
- Non-cooperatively decide on link.
- Decisions are "myopic".
- Interact many times.
- Baseline: assume uniform probability of any pair meeting.

Network Model (4) - Equilibrium Friendships

- Observe a single realisation of network.
- Assume this is **an** equilibrium outcome.
- "Equilibrium" $\equiv$ (pure strategy) Nash equilibrium

$$\begin{aligned} u_i(\mathbf{g}) &\geq u_i(\mathbf{g} + g_{ij}) \forall j \notin N_i(\mathbf{g}), \forall i \\ u_i(\mathbf{g}) &\geq u_i(\mathbf{g} - g_{ik}) \forall k \in N_i(\mathbf{g}), \forall i \end{aligned} \tag{3}$$

  - Decision on each link is optimal, given all others.

Implied Behavioural Assumptions

Inherent restrictions on individual behaviour:

1. Benefit to i from path to j depends only on distance.
   - it is independent of the characteristics.
2. Costs and benefits are additive.
3. Directedness.
4. *Strength* of ties.
5. Single-link deviations.

Identification and Estimation

Combining **Assumptions 1-3** + *Equilibrium* gives:

$$\sum_{d=1}^{D}[n_i(d;\mathbf{g})-n_i(d;\mathbf{g}')]\delta_d+c^*+\mathbf{x}_i'\beta+(\mathbf{z}_i-\mathbf{z}_j)'\Omega(\mathbf{z}_i-\mathbf{z}_j)+\nu_i+\nu_j \geq -\varepsilon_{ij} \quad (4)$$

where:
- $n_i(d;\mathbf{g})$: number of people in network $\mathbf{g}$ s.t. shortest path from i to each of them has length d,
- $c = c^* - \delta_1$,
- $\mathbf{g}\prime = \mathbf{g} + g_{ij}$, and
- D: maximum geodesic in network, "diameter".

Identification and Estimation

Combining **Assumptions 1-3** + *Equilibrium* gives:

$$\sum_{d=1}^{D}[n_i(d;\mathbf{g})-n_i(d;\mathbf{g}')]\delta_d+c^*+\mathbf{x}_i'\beta+(\mathbf{z}_i-\mathbf{z}_j)'\Omega(\mathbf{z}_i-\mathbf{z}_j)+\nu_i+\nu_j \geq -\varepsilon_{ij} \tag{4}$$

- $\{\delta, c, \beta, \gamma, \nu\}$ are identified by the model, up to some normalisation of σ_ε.
  - $\gamma = diag(\Omega)$.
- Estimation: standard binary choice set up.

Simulation

- Predicted g_{ij} from Equation 4 are conditional on $\mathbf{g} \setminus g_{ij}$.
- Want to know $\mathbb{E}[g_{ij} | \mathbf{x_i}, \mathbf{z}_i, \mathbf{z}_j]$.
- Simulate (some) possible equilibrium networks, and use an empirical average.
  - Process not (yet) informed by data.

Outlook

Next steps

- Test on simulated data.
- Implement on real data.
- Test restrictions imposed in theory.
- Use results in further work.

Discussion and Questions

References

- Bala and Goyal (2000) "Self-Organisation in Communication Networks" *Econometrica*
- Christakis and Fowler (2007) "The Spread of Obesity in a Large Social Network over 32 Years" *NJEM*
- Christakis, Fowler, Imbens, and Kalyanaraman (2010) "An Empirical Model for Network Formation" *NBER Working Paper 16039*
- Conley and Udry (2010) "Learning About a New Technology: Pineapple in Ghana" *AER*
- Duflo, Dupas, and Kremer (2011) "Peer Effects and the Impacts of Tracking: Evidence from a Randomized Evaluation in Kenya" *AER*
- Galeotti (2006) "One-way flow networks: the role of heterogeneity" *Economic Theory*
- Jackson and Wolinsky (1996) "A Strategic Model of Social and Economic Networks" *JET*
- Mele (2011) "A Structural Model of Segregation in Social Networks" *mimeo, University of Illinois at Urbana-Champaign*
- Sacerdote (2001) "Peer Effects With Random Assignment: Results For Dartmouth Roommates" *QJE*