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.
Three classes of empirical work:

1. “Fixed network”
   - **Problem:** Selection?
Three classes of empirical work:

1. “Fixed network”
   - **Problem**: Selection?

2. “Randomised network”
   - **Problem**: Success of randomisation?
   - **Problem**: Low external validity.
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?
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 $g$ 
$d(i, j; g)$ 
with values $\delta_{d(i,j;g)}$. 

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 $g$ – $d(i, j; g)$ – with values $\delta_{d(i, j; g)}$.

**Assumption 2:** Link formation is costly.
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 $g$ – $d(i, j; g)$ – with values $\delta_{d(i,j;g)}$.

Assumption 2: Link formation is costly.

Generalised “connections” model of utility:

$$u_i(g) = \sum_{j \neq i} \delta_{d(i,j;g)} - \sum_{j \in N_i(g)} c_{ij}(g)$$  \hspace{1cm} (1)
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.

$$c_{ij}(g) = c^* + x_i'\beta + (z_i-z_j)'\Omega(z_i-z_j) + \nu_i + \nu_j + \varepsilon_{ij} \quad (2)$$

where:

- $c^*$: constant,
- $x_i$: vector of observables for individual “paying” for the link,
- $z_j$: vector of observable characteristics for $j$,
- $\nu_j$: unobserved individual-specific net costs for $j$, and
- $\varepsilon_{ij}$ 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” ≡ (pure strategy) Nash equilibrium

\[
\begin{align*}
  u_i(g) & \geq u_i(g + g_{ij}) \quad \forall j \notin N_i(g), \forall i \\
  u_i(g) & \geq u_i(g - g_{ik}) \quad \forall k \in N_i(g), \forall i
\end{align*}
\]  

- 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.
Combining Assumptions 1-3 + Equilibrium gives:

\[
\sum_{d=1}^{D} [n_i(d; g) - n_i(d; g')] \delta_d + c^* + x_i' \beta + (z_i - z_j)' \Omega (z_i - z_j) + \nu_i + \nu_j \geq -\varepsilon_{ij}
\]

(4)

where:
- \( n_i(d; g) \): number of people in network \( g \) s.t. shortest path from \( i \) to each of them has length \( d \),
- \( c = c^* - \delta_1 \),
- \( g' = g + g_{ij} \), and
- \( D \): maximum geodesic in network, “diameter”.
Combining **Assumptions 1-3 + Equilibrium** gives:

\[
\sum_{d=1}^{D} [n_i(d; g) - n_i(d; g')] \delta_d + c^* + x_i' \beta + (z_i - z_j)' \Omega (z_i - z_j) + \nu_i + \nu_j \geq -\varepsilon_{ij}
\]

(4)

- \{\delta, c, \beta, \gamma, \nu\} are identified by the model, up to some normalisation of \(\sigma_\varepsilon\).
  - \(\gamma = diag(\Omega)\).

- Estimation: standard binary choice set up.
Simulation

- Predicted $g_{ij}$ from Equation 4 are conditional on $g \setminus g_{ij}$.

- Want to know $\mathbb{E}[g_{ij} | x_i, z_i, z_j]$.

- 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

- Christakis and Fowler (2007) “The Spread of Obesity in a Large Social Network over 32 Years” *NJEM*
- 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*