Social connectedness

[and its influence on participation in risky behaviour]

A. Advani

University College London and Institute for Fiscal Studies

Irish Society of New Economists Conference, 2011

Motivation

- Lots of informal discussion of the influence of connections (e.g. Granovetter, Christakis and Fowler).
- To see whether connections have influence on behaviour or outcomes, need them to be exogenous.
- Therefore need to first work out how likely people are to be friends.

Literature - Network Theory

- Early contribution: Goyal (1993)
- Seminal paper: Jackson and Wolinsky (1996) - *Connections Model*
- More recently: Bala and Goyal (2000); Bramoullé *et al.* (2004); Fafchamps, van der Leij, and Goyal (2010); Galeotti *et al.* (2010)

Literature - Network Empirics (1)

Four classes of empirical work:

1. **"Peer group effects"** – doesn't use information on network structure, often because data aren't available.
   Problem: Manski (1993) highlights important identification problems with these structures, showing that it is rarely possible to draw any conclusions as to *why* we observe a correlation in outcomes.
   Example: Duflo, Dupas and Kremer (2008)

Literature - Network Empirics (1)

Four classes of empirical work:

1. **"Peer group effects"** – doesn't use information on network structure, often because data aren't available.
   Problem: Manski (1993) highlights important identification problems with these structures, showing that it is rarely possible to draw any conclusions as to *why* we observe a correlation in outcomes.
   Example: Duflo, Dupas and Kremer (2008)

2. "**Fixed network**" – takes the network as given (wrt action we are considering), potentially with component fixed effects.
   Problem: In many cases the network being exogenously fixed and given is unlikely to be a reasonable assumption, in which case not controlling for the selection in connection formation will bias the results.
   Examples: Conley & Udry (2010), Calvó-Armengol, Patacchini, and Zenou (2009)

Literature - Network Empirics (2)

3. "**Randomised network**" – uses data where the author can argue that network formation occurred in an random (exogenous) way.
   Problem: Low external validity - even if we believe randomisation has been successful in creating the network, it is hard to think about how we can learn from these results since the formation process bears no relation to reality.
   Example: Sacerdote (2001)

Literature - Network Empirics (2)

3. "**Randomised network**" – uses data where the author can argue that network formation occurred in an random (exogenous) way.
   Problem: Low external validity - even if we believe randomisation has been successful in creating the network, it is hard to think about how we can learn from these results since the formation process bears no relation to reality.
   Example: Sacerdote (2001)

4. "**Structural network**" – directly models the network formation process.
   Problem: Relies on the model being a valid abstraction of reality.
   Example: Christakis, Fowler, Imbens & Kalyanaraman (2010)

Network Model (1) - Notation

The model we consider is based on Jackson and Wolinsky (1996).

The Network

Let $\mathcal{N} = \{1, 2, ...N\}$ be a finite set of agents.
Let $\mathbf{g} \in \mathcal{G}_N$ denote a directed network, in the set of all possible directed networks on nodes (agents) $\mathcal{N}$.
For each pair of individuals $i, j \in \mathcal{N}$, $g_{ij} = 1$ iff individual i has chosen individual j as a direct friend, 0 otherwise.

Network Model (2) - Definitions

Definition

A *path* between agents i and j is a sequence of agents beginning with i and ending with j such that each agent is connected to the next.

Definition

$d(i, j; \mathbf{g})$ is the *geodesic distance* – the length of the shortest path – between agents i and j in network $\mathbf{g}$.

Definition

$N_i(\mathbf{g}) := \{j | g_{ij} = 1\}$ as the set of neighbours of i; the people to whom i has chosen to link.

Network Model (3) - Preferences

Assumption 1: Individual i receives a benefit from having a path to another individual j. The magnitude of this benefit depends on the length of the shortest path between i and 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 (3) - Preferences

Assumption 1: Individual i receives a benefit from having a path to another individual j. The magnitude of this benefit depends on the length of the shortest path between i and j in network $\mathbf{g}$, $d(i, j; \mathbf{g})$, with values $\delta_{d(i,j;\mathbf{g})}$.

Assumption 2: Link formation is costly.

Under these two assumptions, we can define our generalised connections model of utility as:

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

This is a generalisation of the Jackson and Wolinsky *connections model*, which assumes benefits decline geometrically in the geodesic distance.

Network Model (4) - Costs

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

$$c_{ij}(\mathbf{g}) = c + |\mathbf{z}_i - \mathbf{z}_j|\prime\gamma - \nu_i - \nu_j - \varepsilon_{ij} \tag{2}$$

where:

- $\mathbf{z}_j$ is a vector of observable characteristics for individual j,
- ν's are unobserved individual-specfic costs for each agent, and
- ε_{ij} is an unobserved link-specific cost representing unobserved characteristics.

This includes the restriction that differences in observed characteristics are treated symmetrically, so that only the absolute difference matters.

Implied Behavioural Assumptions

Inherent in the modelling decisions above we have made a number of restrictions on individual behaviour:

1. The benefit to individual i from having a path to some other agent j depends only on the distance of j from i in the network; it is independent of the characteristics of i, j, and all other agents in the network.
2. All heterogeneity, observed and unobserved, has been placed in the cost function.
3. We have so far assumed that the network is directed, so if $g_{ij} = 1$ and $g_{ji} = 0$ then i receives benefit δ from j, but j only receives a benefit from i if he is connected to i via a directed path. This could be easily relaxed.

Network Model (5) - Equilibrium Friendships

We assume the observed network is the equilibrium outcome of individuals' utility maximising link formation decisions.
We consider possible deviations from the observed network on a link-by-link basis. We define *single-link equilibrium* as:

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

i.e. no agent could be better off by either adding a link to someone he is currently not linked to or removing a link to someone he is currently linked to.

Estimation (1)

From Assumptions 1-3 and our definition of equilibrium we get a series of inequalities, which can be simplified to give:

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

where $n_i(d;\mathbf{g})$ is the number of people in network $\mathbf{g}$ such that the shortest path from i to each of those people has length d, and $\mathbf{g}\prime = \mathbf{g} + g_{ij}$.

Estimation (1)

From Assumptions 1-3 and our definition of equilibrium we get a series of inequalities, which can be simplified to give:

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

where $n_i(d;\mathbf{g})$ is the number of people in network $\mathbf{g}$ such that the shortest path from i to each of those people has length d, and $\mathbf{g'} = \mathbf{g} + g_{ij}$.
Intuitively, the net costs of forming an additional link, must be higher than the unobserved link-specific benefit of that link.
The net costs are composed of:

- the change in network benefits
- the additional link costs: a constant, the observable components, and the individual-specific unobserveables, which are replaced with fixed effects.

Estimation (2)

With some distributional assumption on the ε_{ij} we have a standard binary choice set up.

$$g_{ij} = 1(\Delta \mathbf{n}_{ij}\prime\delta + c + |\mathbf{z}_i - \mathbf{z}_j|\prime\gamma - \nu_i - \nu_j \geq \varepsilon_{ij}) \qquad (5)$$

So we can consistently estimate the parameters: $\{\delta, c, \gamma, \nu_i\}$.

Simulation (1)

- Having recovered estimates of the parameters of interest $\{\hat{\delta}, \hat{c}, \hat{\gamma}, \hat{\nu}\}$ we want to estimate $\mathbb{E}g_{ij} \, \forall i \neq j$.
- Simply using the predicted g_{ij} from Equation5 is not sufficient, since these are conditional on $\mathbf{g} \backslash g_{ij}$, the other links in the network.
- We can simulate a number of possible equilibrium networks, and then use an empirical average to estimate $\mathbb{E}g_{ij}$.

Simulation (2)

Simulation

1. Begin with the empty network, $\mathbf{g}_0$.
2. Pairs of agents are randomly selected to meet.
3. When a pair of agents meets, each can unilaterally decide to make/break a link with the other, and we update the network accordingly.
4. This process continues until no agent wants to adjust his links.
5. We store this outcome as one equilibrium network.
6. Repeat 1-5 until we have a set of S equilibrium networks. Then $\hat{\mathbb{E}}\mathbf{g} = S^{-1} \sum_{s=1}^{S} \mathbf{g}(s)$.

Data

- We use a subsample of data from the National Longitudinal Survey of Adolescent Health (Add Health).
- This subsample contains about 2,750 students, spread across 16 high schools in the US, collected 1994/5.
- Since we have data on (almost) all the pupils in the school, we have complete network information.

Since the schools are unrelated and geographically seperate, there are no cross-school links. This allows us to run 16 seperate estimations (with an average of 171 children in each).

Outlook

Next steps

- Implement the method described on the data (Add Health).
- Use the results to assess the validity of the workhorse theoretical model of network economics - test the restrictions of geometric decline in benefits, and undirectionality.
- Use the estimated probability of having a link in further work examining the impact of social connectedness on individual decision-making, particularly participation in risky behaviour.

Discussion and Questions