Prepared with SEVISLIDES

Building socio-economic networks: how many conferences should you attend?

Antonio Cabrales, Antoni Calvó-Armengol, Yves Zenou

October 2006

Summary

- Spillovers between different agents generate incentives for "linking."
  - Research and development.
  - Labor Market Information.
  - Friendships and "Social Capital."
- If linking is done "non-cooperatively," inefficiencies arise (overlinking - underwork), so role for policy.

- Previous "purely economic" work does not look very much at endogenous and costly network formation.

- The (more game-theoretic) work that does, simplifies away the game after forming the network.

- Reason: Analytical intractability.

Model

- We analyze a network formation game with two choices:
  - Socialization effort.
  - Productive effort.
- The key simplification is: undirected socialization.
  - Each link created with probability equal to product of socialization efforts.
  - Thus random network.
- Strategy space much simpler (one dimensional for each player - rather than $n-1$-dimensional), so equilibrium is a smaller-sized fixed point.

- As a result we can:
  - Discuss welfare and policies.
  - We can also replicate some-fat tails, short distance-(but not all-clustering) features of available data.
  - We can (and do) perform statistical (regression) analysis.

Results

- Equilibrium: for "large" groups - two stable (and one unstable).

- The equilibria are "ordered": both in actions and in welfare.

- An increase in returns increases (decreases) actions at Low (High) equilibrium.

- This increase in returns has stronger relative effect on socialization effort.
  - An explanation for the explotion of R&D collaboration.
  - Perhaps also for the decrease in social capital.

- Heterogeneity: A mean preserving spread in rewards, increases (decreases) payoffs at Low (High) equilibrium.

- Results robust to cost structure.

Prior work:

- Spillovers (theory): Marshall (1920), D'Aspremont and Jacquemin (1988), Bénabou (1993).
- Spillovers (empirics): Ciccone, Hall (1996); Cassiman, Veugelers (2002).
- Spillovers (policy): Motta (1996), Leahy and Neary (1997).
- Networks: Myerson (1981), Jackson and Wolinsky (1996).
- Replicating the features of data: Jackson and Rogers (2006).
- Play on fixed networks: Calvó-Armengol and Jackson (2004), Bramoullé and Kranton (2005).

The replica game

- $N = \{1, ..., n\}$ finite set of players, $T = \{1, \ldots, t\}$ finite set of types.
- There are exactly m players of each type $\tau \in T$.
- For each $i \in N$, $\tau(i) \in T$ is his type.
- Simultaneous move game of network formation and investment.
- Returns to investment are the sum of a private component and a synergistic component.

The game *(2/5)*

- Private returns are heterogeneous: $\mathbf{b} = (b_1, ..., b_t)$ where $0 < b_1 \leq b_2 \leq ... \leq b_t$.

- The synergistic returns depend on the network.

Network formation

- Each player i selects $k_i \geq 0$ a level of socialization effort. $\mathbf{k} = (k_1, ..., k_n)$.

- Then, i and j interact with link intensity ($g_{ij} = g_{ji}$):

$$g_{ij}(\mathbf{k}) = \rho(\mathbf{k})\, k_i k_j; \quad g_i(\mathbf{k}) = \sum_{j=1}^{n} g_{ij}(\mathbf{k}); \quad \rho(\mathbf{k}) = \begin{cases} 1/\sum_{j=1}^{n} k_j, & \text{if } \mathbf{k} \neq \mathbf{0} \\ 0, & \text{if } \mathbf{k} = \mathbf{0} \end{cases}$$

- When $\max_i k_i^2 < 1/\rho(\mathbf{k})$, network interpretable as random graph where $g_{ij}(\mathbf{k})$ probability of ij edge.

- Random graph model with expected degrees $\mathbf{k} = (k_1, ..., k_n)$ in Chung and Lu (2002) (can replicate Poisson distributions, power laws etc.)

Investment

- Each player i selects an investment level $s_i \geq 0$ and $\mathbf{s} = (s_1, ..., s_n)$.
- The choices of k_i and s_i are simultaneous.
- Individual investment yields private return and synergistic return.
- Private returns: $b_{\tau(i)}\, s_i - s_i^2/2$.
- Synergistic returns: $\frac{\partial^2 u_i(\mathbf{s},\mathbf{k})}{\partial s_i \partial s_j} = a g_{ij}(\mathbf{k})$, $a \geq 0$

Payoffs

Formally, let $p_{ij}\left(\mathbf{k}\right) = g_{ij}\left(\mathbf{k}\right)$ if $i \neq j$ and $p_{ii}\left(\mathbf{k}\right) = g_{ii}\left(\mathbf{k}\right)/2$. Player i's utility is given by:

$$u_i(\mathbf{s}, \mathbf{k}) = b_{\tau(i)}\, s_i + a \sum_{j=1}^{n} p_{ij}\left(\mathbf{k}\right) s_j s_i - \frac{1}{2} s_i^2 - \frac{1}{2} k_i^2 \qquad (1)$$

- We solve for Nash equilibria in pure strategies $(\mathbf{s}^*; \mathbf{k}^*) = (s_1^*, ..., s_n^*; k_1^*, ..., k_n^*)$ of $m-$replica game with m large enough.

- There are exactly three such equilibria.
  - One (partially corner) with null socialization.
  - Two interior equilibria.

Lemma 1 $\left(s_i^*, k_i^*\right) = \left(b_{\tau(i)}, 0\right)$ *for all* $i = 1, ..., mt$ *is a pure strategy Nash equilibrium with payoffs* $b_{\tau(i)}^2/2$.

- It is a strict equilibrium, but not stable for large populations, as we will show later.

Define:

$$a(\mathbf{b}) = a\frac{\sum_{\tau=1}^{t} b_\tau^2}{\sum_{\tau=1}^{t} b_\tau}. \tag{2}$$

- Holding average type $\sum_{\tau=1}^{t} b_\tau / t$ constant, $a(\mathbf{b})$ increases with heterogeneity in types.

- Many authors refer to $\sum_{\tau=1}^{t} b_\tau^2 / \sum_{\tau=1}^{t} b_\tau$, as the second-order average type (see e.g. Vega-Redondo 2006).

Theorem 2 *Suppose* $2/3\sqrt{3} > a(\mathbf{b}) > 0$*. Then, for* $m \geq m^*$*, there are exactly two interior pure strategy Nash equilibria.*
For these equilibria (s_i, k_i) *converge to* $(s^*_{\tau(i)}, k^*_{\tau(i)})$ *as* m *goes to infinity* $s^*_{\tau(i)} = b_{\tau(i)}s$, $k^*_{\tau(i)} = b_{\tau(i)}k$, *and* (s, k) *are positive solutions to:*

$$\begin{cases} k = a(\mathbf{b})s^2 \\ s\left[1 - a(\mathbf{b})k\right] = 1 \end{cases} \tag{3}$$

- Under $2/3\sqrt{3} > a(\mathbf{b}) > 0$. , the system (3) has exactly two positive solutions.

Simulations on Theorem 1 with $a = 2$, $t = 1$ and $b_1 = 0.1$. Numbers are multiplied by 10^4.

n	2	5	10	20	50	100	500	∞
				Low equilibrium				
s^*	1,898	1,195	1,101	1,065	1,049	1,046	1,046	1,046
k^*	2,366	815	458	303	234	222	218	219
				High equilibrium				
s^*	3346	4,643	4,591	4,508	4,444	4,420	4,400	4,394
k^*	3506	3,923	3,911	3,891	3,875	3,869	3,864	3,862

- The two equations (3) equalize marginal costs with marginal benefits at equilibrium.

- The marginal benefit from investment s_i^* is:

$$b_{\tau(i)}/(1-\frac{a(\mathbf{b})}{b_{\tau(i)}}k_i^*)$$

- When $a = 0$, this marginal benefit boils down to $b_{\tau(i)}$, the private return in (1).

- When $a \neq 0$, this is scaled up by *synergistic multiplier* $1/(1-\frac{a(\mathbf{b})}{b_{\tau(i)}}k_i^*)$, homogeneous across players and an increasing function of the second order average type $a(\mathbf{b})$.

- The marginal benefit of k_i^*, as the population size gets large, boils down to

$$a\rho\left(\mathbf{k}\right)\sum_{j=1}^{n}s_is_j$$

- The condition $2/3\sqrt{3} > a(\mathbf{b})$ is necessary and sufficient for (3) to have a non-negative solution.

- When $a(\mathbf{b})$ is too large, the synergistic multiplier operates too intensively and both k and s increase without bound.

- The socialization effort at equilibrium :

$$\frac{k_i^*}{k_j^*} = \frac{b_{\tau(i)}}{b_{\tau(j)}}.$$

- Thus, intensity of a link at approximate equilibrium is:

$$g_{ij}\left(\mathbf{k}^*\right) = k^* \frac{b_{\tau(i)} b_{\tau(j)}}{m \sum_{\tau=1}^{t} b_\tau}, \tag{4}$$

which decreases linearly with $1/m$.

- For this reason, the overall socialization effort $g_i\left(\mathbf{k}^*\right) = k^* b_{\tau(i)}$ is independent of the population size.

Proposition 3 *For m sufficiently large, the two interior equilibria are stable while the equilibrium with $\left(s_i^*, k_i^*\right) = \left(b_{\tau(i)}, 0\right)$ for all $i = 1, ..., mt$ is not stable.*

Proposition 4 *Let $(\mathbf{s}^*, \mathbf{k}^*)$ and $(\mathbf{s}^{**}, \mathbf{k}^{**})$ be the two different approximate equilibria of an $m-$replica game. Then, without loss of generality, $(\mathbf{s}^*, \mathbf{k}^*) \geq (\mathbf{s}^{**}, \mathbf{k}^{**})$ and $\mathbf{u}(\mathbf{s}^*, \mathbf{k}^*) \geq \mathbf{u}(\mathbf{s}^{**}, \mathbf{k}^{**})$, where $\geq$ is the componentwise ordering.*

Socialization and investment

Proposition 5 *Let* $(\mathbf{s}^*, \mathbf{k}^*) \geq (\mathbf{s}^{**}, \mathbf{k}^{**})$ *be the two ranked approximate equilibria of an* $m-$*replica game.*
Suppose that $a(\mathbf{b})$ *increases.*
Then, at the Pareto-superior approximate equilibrium $(\mathbf{s}^*, \mathbf{k}^*)$ *all the equilibrium actions decrease,*
while at the Pareto-inferior approximate equilibrium $(\mathbf{s}^{**}, \mathbf{k}^{**})$ *all the equilibrium actions increase.*
In both cases, the percentage change in k_i *is higher than that of* s_i *(in absolute values), for all* $i = 1, ..., mt$*.*

Equilibrium payoffs

When m gets large, equilibrium payoffs are:

$$u_i^* = \frac{b_{\tau(i)}^2}{2a(\mathbf{b})} \frac{k}{s} + o(1), \text{ for all } i = 1, ..., mt. \quad (5)$$

$$= \frac{b_{\tau(i)}^2}{2} s + o(1), \text{ for all } i = 1, ..., mt. \quad (6)$$

Proposition 6 *Let* $(\mathbf{s}^*, \mathbf{k}^*) \geq (\mathbf{s}^{**}, \mathbf{k}^{**})$ *be the two ranked approximated equilibria of an* $m-$*replica game.*

1. *Suppose that either only* a *increases, or* $(a; b_1, ..., b_t)$ *are all scaled up by a common multiplicative factor. Then, at the Pareto-superior approximated equilibrium all the payoffs* $u_i(\mathbf{s}^*, \mathbf{k}^*)$ *decrease, while at the Pareto-inferior approximated equilibrium all payoffs* $u_i(\mathbf{s}^{**}, \mathbf{k}^{**})$ *increase, for all* $i = 1, ..., mt$*.*

2. *Suppose that the vector* $(b_1, ..., b_t)$ *changes via a mean preserving spread (i.e. a change that holds* $\sum_{\tau=1}^{t} b_\tau$ *constant but increases* $\sum_{\tau=1}^{t} b_\tau^2$*). Then, at the Pareto-superior approximated equilibrium the sum of payoffs* $\sum_{i=1}^{mt} u_i(\mathbf{s}^*, \mathbf{k}^*)$ *decreases, as well as payoffs for types below the average. At the Pareto-inferior approximated equilibrium the sum of payoffs* $\sum_{i=1}^{mt} u_i(\mathbf{s}^{**}, \mathbf{k}^{**})$ *increases.*

Remark 7 *Let* $(\mathbf{s}^*, \mathbf{k}^*) \geq (\mathbf{s}^{**}, \mathbf{k}^{**})$ *be the two ranked approximated equilibria of an* $m-$*replica game. Fix* i *and let* $b'_{-\tau(i)}$ *and* $b_{-\tau(i)}$ *be two different population types (excluding* i*). If* $a(b_{\tau(i)}, b_{-\tau(i)}) \geq a\left(b_{\tau(i)}, b'_{-\tau(i)}\right)$, *then player* i *gets a lower (resp. higher) utility at the Pareto superior approximated equilibrium (resp. at the Pareto inferior approximated equilibrium) under* $(a, b_{\tau(i)}, b_{-\tau(i)})$ *that under* $(a, b_{\tau(i)}, b'_{-\tau(i)})$.

- Key network regularities:

  1. The distribution of connectivities is fat tailed. Higher proportion of nodes with many links than at random.

  2. Average distance (or shortest path) between nodes is very small and grows very slowly with network size. For Hollywood actors network is 225,226 individuals and average path length is 3.65.

  3. Third, the tendency of two linked nodes to be linked to a common third-party, (clustering coefficient), is much higher than at random. For the movie actors 3,000 times higher.

  4. Social networks exhibit internal (sometimes hierarchical) community structure, sometimes arranged hierarchically.

  5. Also, highly connected nodes tend to be connected with highly connected nodes (positive assortativity).

Topology: theory and empirics *(2/12)*

- Some mechanisms replicate this topological features, (Jackson and Rogers 2006).

- Basic ingredients are: a population growth process, and a link formation device for newcomers that combines random meetings with network search of the partner.

- Our model is static. It thus cannot replicate some things (e.g. high clustering).

- Yet, delivers some implications for topology, and relates it topology to incentives.

- Since $g_i(\mathbf{k}^*) = k_i^* = k^* b_{\tau(i)}$ when link intensities are smaller than one, we can interpret our network as a random graph.

- Then, the average connectivity is $\overline{\mathbf{k}^*} = k^* \overline{\mathbf{b}}$,

- The empirical variance of connectivities is $v(\mathbf{k}^*) = k^{2*} v(\mathbf{b})$. Therefore,

$$\frac{\sqrt{v(\mathbf{k}^*)}}{\overline{\mathbf{k}^*}} = \frac{\sqrt{v(\mathbf{b})}}{\overline{\mathbf{b}}}.$$

- Heteregeneity is driven by the heterogeneity in private returns. We can thus cover many topologies, including fat tailed connectivity.

- With Proposition 6, we can show impact on welfare of some changes in $\mathbf{b}$.

- Chung and Lu (2002) show that average distance in a random graph with expected connectivity $\left(k_1^*, ..., k_n^*\right) = k^*\left(b_{\tau(1)}, ..., b_{\tau(n)}\right)$ is:

$$(1 + o(1)) \frac{\log (mt)}{\log \left(k^* \overline{\mathbf{b}}\right)}.$$

Summarizing :

	Low equilibrium				High equilibrium			
	$\bar{\mathbf{k}}$	$v(\mathbf{k})$	$\overline{\text{distance}}$	payoffs	$\bar{\mathbf{k}}$	$v(\mathbf{k})$	$\overline{\text{distance}}$	payoffs
a up	$+$	$+$	$-$	$+$	$-$	$-$	$+$	$-$
$(a, \mathbf{b})$ all up	$+$	$++$	$-$	$++$	$\cdot$	$\cdot$	$\cdot$	$-$
$\mathbf{b}$ spread	$+$	$++$	$--$	$++$ (payoffs)	$-$	$\cdot$	$+$	$--$ (payoffs)

Topology: theory and empirics *(6/12)*

- Our static model does not generate networks with a high clustering.

- Yet, split the population into smaller subpopulations.

- $1 - \varepsilon$ of the socialization in-home, while a residual fraction ε is invested in the whole

- The smaller the size of each community, the bigger the clustering level (for identical average connectivity).

- This goes against our characterization of equilibrium actions.

- Finally, empirically observed social networks have a giant component Next section.

Empirics

- Data from the National Longitudinal Survey of Adolescent Health (AddHealth).
- Students in grades 7-12 from roughly 130 private and public schools in years 1994-95.
- Detailed information on friendship relationships.
- Detailed information on grades (math, history, social studies and science). We calculate an index.
- We take the network comprising the largest number of individuals for our exercise, with 107 nodes.

For this network, we focus on:

- the degree connectivity of each node k_i, $i = 1, ..., 107$
- the student achievement for each node e_i, $i = 1, ..., 107$

- We transform the performance measure. We write $e_i = s_i^{\beta} \exp(\varepsilon_i)$.
- In equilibrium $k_i/s_i = k/s$. Thus $e_i = \left(\frac{s}{k} k_i\right)^{\beta} \exp(\varepsilon_i)$.
- We run the regression: $\log(e_i) = \delta + \beta \log(k_i) + \varepsilon_i$.
- We find $\widehat{\delta} = .0686$ and $\widehat{\beta} = 1.3264$, significant at 10% and 1%.
- We then change variables: $s_i = e_i^{1/\widehat{\beta}}$, so $\log(s/k) = \log\left[\widehat{\delta}/\widehat{\beta}\right]$.
- Since $k = a(\mathbf{b})s^2$, and $s_i = b_{\tau(i)}s$, $s_i = \frac{b_{\tau(i)}}{a(\mathbf{b})}\frac{k}{s}$.

- We do an ML fit of: $s_i = \frac{b_{\tau(i)}}{a(\mathbf{b})} \exp[-\widehat{\delta}/\widehat{\beta}] + \nu_i$ conditional on $a(\mathbf{b}) < 2/3\sqrt{3}$

- First only four different types $(b_1, ..., b_4)$ - agents to types by quartiles.

- Then ten parameters $(b_1, ..., b_{10})$, in deciles.

- We obtain:

$$
\begin{aligned}
(a; b_1, ..., b_4) &= (0.1857; 1.75, 1.87, 1.98, 2.11) \\
(a; b_1, ..., b_{10}) &= (0.2097; 1.21, 1.33, 1.42, 1.55, 1.61, 1.76, 1.85, 1.90, 1.99, 2.
\end{aligned}
$$

- Easy to check, that individuals rank partners in decreasing value of their type for the high equilibrium(the opposite order for the low equilibrium).

- In this particular case, the only stable pairwise matching groups types 1 with types 2, and types 3 with types 4.

- This stable matching does not maximize social welfare at the high equilibrium.

- Allowing groups with more than two types, the ordering for type 1 at high equilibrium is: $(1,3,4)$, $(1,4)$, $(1,2,3,4)$, $(1,2,4)$, $(1,3)$, $(1,2,3)$, $(1,2)$.

We can also use the estimated types to illustrate comparative statics of a mean preserving spread.

- Divide the 107 nodes into 27 agents of each type.
- Then, let x individuals type 1 and 4, and $54 - x$ individuals for type 2 and 3, and vary x from 1 to 53.
- We observe numerically the monotonicity of in prop. 6.
- And, for this parameters, the utility of types 3 and 4 (not covered in prop 6) changes in the same direction as the others.

The case of homogeneous populations *(1/5)*

- Remark 7 implies that individuals of the highest type prefer to segregate.

- A simple induction argument justifies that the highest types of any heterogeneous subgroup would want to segregate.

- One would expect that some homogeneous groups to exists in a given society.

- We can also conduct some robustness checks on the technology and further insights on topology.

Player i's utility is:

$$u_i(\mathbf{s},\mathbf{k}) = bs_i + a\sum_{j=1}^{n} p_{ij}s_js_i - \frac{1}{c+1}s_i^{c+1} - \frac{1}{c+1}k_i^{c+1}, \tag{9}$$

where $a, b \geq 0$ and $c \geq 1$. The case $c = 1$ corresponds to quadratic costs. As c increases, the cost function becomes steeper.

We introduce $\phi : \mathbb{R} \to \mathbb{R}$ given by:

$$\phi\left(x\right) = c^{\frac{1}{c+1}}\left[x^{c+1} - b^{1+\frac{1}{c}}\right]^{\frac{1}{c+1}}.$$

- For large populations, up to two *interior symmetric equilibria* solving:

$$\begin{cases} k^c = as^2 \\ s^c\left[1 - ak\right] = b \end{cases}, \tag{10}$$

with added condition $k^* \leq \phi\left(s^*\right)$. This is equivalent to:

$$u_i\left(s^*, k^*\right) = \frac{1}{c+1}\left[cs^{*c+1} - k^{*c+1}\right] \geq \frac{c}{c+1} b^{1+\frac{1}{c}} = u_i\left(b^{1/c}, 0\right). \tag{11}$$

- The condition $k^* < \phi\left(s^*\right)$ guarantees that $\left(b^{1/c}, 0\right)$ is not a strict best-response player i to the rest playing $\left(s^*, k^*\right)$.

Proposition 8 *Suppose that (10) has three different solutions. Let* $\left(s^*, k^*\right) \geq \left(s^{**}, k^{**}\right)$ *be the two ranked interior symmetric approximate equilibria for large population. When* a *increases,* s^{**} *and* k^{**} *increase, while* s^* *and* k^* *decrease. In both cases, the percentage change in* k *is higher than that of* s.

The topology of Erdös-Rényi equilibrium networks

- In the Erdös-Rényi (Bernoulli) random networks that correspond to the interior equilibria expected number of links is k

- Network connectivity (or degree) is not correlated across different nodes.

- When $k^* < 1$, the networks is composed of a huge number of disjoint small trees.

- When $k^* > 1$, a single giant component that encompasses a high fraction of all the network nodes emerges.

The case of homogeneous populations *(5/5)*

Proposition 9 *When* $a \geq 1$*, no equilibrium network has a giant component. Suppose that* $a < 1$ *and that there are two non-empty equilibrium networks. Then, the two equilibrium networks display different topological characteristics (one network with a giant component, one without) if and only if* $ab^{2/c} < (1-a)^{2/c}$*. If, instead,* $ab^{2/c} > (\frac{c}{a+c})^2$*.*

Building socio-economic networks: how many conferences should you attend?

Antonio Cabrales, Antoni Calvó-Armengol, Yves Zenou

October 2006