Homophily in preferences or meetings?

Identifying and estimating an iterative network formation model

Luis Alvarez[1] Cristine Pinto[2] Vladimir Ponczek[2]

[1]University of São Paulo

[2]São Paulo School of Economics, FGV

World Congress, August 2020

Overview

- Homophily, the observed tendency of individuals with similar characteristics maintaining a relationship, is a salient feature of social and economic networks.
- Appropriate modelling of homophily has been a consistent topic in the recent push for estimable econometric models of network formation (Goldsmith-Pinkham and Imbens, 2013; Chandrasekhar and Jackson, 2016; Mele, 2017; Graham, 2016, 2017).
- Strand of the literature distinguishes between homophily due to choice (**preferences**) and homophily that is due to opportunity (**meetings**) (Jackson, 2010).
  - Potentially nonnegligible policy implications.
  - Theoretical models on the distinction can be found in Currarini et al. (2009) and Bramoullé et al. (2012).

Overview

- These models have some important drawbacks:
  - they analyse long-run or steady state behaviour. No role for transitional dynamics;
  - they are either purely probabilistic or allow for only a limited set of pay-offs.
- These features limit the empirical usefulness of these models.

This paper

1. study identification and estimation of an iterative network formation model which incorporates both "homophilies";
   - network formation process is related to the theoretical literature on network formation (Jackson and Watts, 2002);
   - model builds upon iterative network formation algorithm found in Mele (2017):
     1. more general payoffs and meeting processes;
     2. study identification of meeting parameters;
     3. analyse off-stationary equilibrium behaviour.
2. apply this model in the context of intraclassroom network data in Northeastern Brazil.
   - model allows us to conduct counterfactual exercises.

Network game: setup I

- Finite set of agents $\mathcal{I} := \{1, 2 \dots N\}$.
- Agent $i \in \mathcal{I}$ endowed with $k \times 1$ vector of exogenous characteristics W_i.
- Agents' vectors are stacked on matrix $X := \begin{bmatrix} W_1 & W_2 & \cdots & W_N \end{bmatrix}'$.
- Exogenous characteristics are drawn according to law $\mathbb{P}_X$ before game starts and remain *fixed* throughout.
- Support of X is $\mathcal{X}$. Realisation of X denoted by element $x \in \mathcal{X}$.
- Time is discrete.

Network game: setup II

- At each round $t \in \mathbb{N}$, agents' relations are described by a *directed network*.
- Information on the network is stored on a $N \times N$ *adjacency matrix* g.
  - $g_{ij} = 1$ if i nominates j as a friend and 0 otherwise. $g_{ii} = 0$ by assumption.
- $\mathcal{G}$: set of all $2^{N(N-1)}$ possible adjacency matrices.

Network game: setup III

- $u_i : \mathcal{G} \times \mathcal{X} \to \mathbb{R}$, $u_i(g, x)$: utility i derives under g and when covariates are $X = x$.
- individual utility may possibly depend on the entire matrix of covariates X.

Network game: timing I

- Agents are **myopic**, i.e. they form, maintain or sever relationships based on the current utility these bring.
- At each round, a *matching process* m^t selects a **pair** of agents (i, j).
- m^t: stochastic process $\{m^t : t \in \mathbb{N}\}$ over $\mathcal{M} := \{(i, j) \in \mathcal{I} \times \mathcal{I} : i \neq j\}$.
- After the matching process selects a pair of agents, a pair of choice-specific idiosyncratic shocks $(\epsilon_{ij,t}(0), \epsilon_{ij,t}(1))$ are drawn.
- $\epsilon_{ij,t}(1)$: taste shock of i in forming/maintaining a relationship with j at time t.
- Shocks are unobserved by the econometrician.

Network game: timing II

- After a pair (i, j) is selected and shocks are observed, i gets to choose whether to form/maintain or not form/sever a relationship with j.
- Given that choice is myopic, agent i forms/maintain a relation with j iff:

$$u_i([1, g_{-ij}], X) + \epsilon_{ij,t}(1) \geq u_i([0, g_{-ij}], X) + \epsilon_{ij,t}(0)$$

where $[a, g_{-ij}]$ is the network equal to g except at ij, where it equals a.

Network game: assumptions

The next assumptions constrain the matching process and distribution of shocks.

Assumption 1

The meeting process $\{m^t : t \in \mathbb{N}\}$ is described by a *time-invariant* matching function $\rho : \mathcal{M} \times \mathcal{G} \times \mathcal{X} \mapsto [0,1]$, where $\rho((i,j), g, x)$ is the probability that (i,j) is selected when covariates are $X = x$ and *the previous-round network was g*. Moreover, for all $g \in \mathcal{G}$, $x \in \mathcal{X}$, $(i,j) \in \mathcal{M}$, $\rho((i,j), g, x) > 0$.

Assumption 2

Shocks are drawn iid across pairs and time from a *known* distribution $(\epsilon(0), \epsilon(1))' \sim F_\epsilon$ which is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^2$ and has positive density almost everywhere.

Network game: transition matrix

- Conditional on $X = x$, network game + initial distribution induce a Markov chain $\{g^t : t \in \mathbb{N} \cup \{0\}\}$ on the set of netwotks $\mathcal{G}$.
- $\Pi(x)$: $2^{N(N-1)} \times 2^{N(N-1)}$ transition matrix on $\mathcal{G}$ (conditional on $X = x$).
- $\Pi(x)_{gw}$, $g, w \in \mathcal{G}$, is the probability of transitioning to w given the current period network g.
- Neighbourhood of network g:
  $N(g) \coloneqq \{w \in \mathcal{G} \setminus \{g\} : \exists! \ (i,j) \in \mathcal{M}, g_{ij} \neq w_{ij}\}$

Identification I

- Identification and estimation under many-network asymptotics.
- We have access to a random sample $\{G_c^{T_0}, G_c^{T_1}, X_c\}_{c=1}^{C}$.
- $G_c^{T_0}$ and $G_c^{T_1}$: observations of network c over two (possibly nonconsecutive) periods (labelled *first* and *second*).
- Denote by $\Pi(x; \theta_0)$ the transition matrix when covariates are $X = x$, and $\theta_0 := ((u_i)_{i=1}^{N}, \rho)$ the "true" parameters (functions).
- τ_0: number of *rounds* taken place between the first and second period. Potentially unknown.

Identification II

- We are able to identify $\Pi(X;\theta_0)^{\tau_0}$, provided that the first-period conditional distribution, $\pi_0(X)$, is such that $\pi_0(X) >> 0$.
- τ_0: number of *rounds* taken place between the first and second period.

Example

If X is empty, we can consistently estimate $(\Pi^{\tau_0})_{gw}$, $g, w \in \mathcal{G}$, by $(\widehat{\Pi^{\tau_0}})_{gw} = \sum_{c=1}^{C} \mathbb{1}\{G_c^{T_0} = g, G_c^{T_1} = w\} / \sum_{c=1}^{C} \mathbb{1}\{G_c^{T_0} = g\}$, provided $\mathbb{P}[G_c^{T_0} = g] > 0$.

Assumption 3 (Full support)

$\pi_0(g|X) > 0$ for all $g \in \mathcal{G}$ $\mathbb{P}_X$-a.s.

- **Main idea:** identify τ_0 from $\Pi(X;\theta_0)^{\tau_0}$; then identify θ_0.

Identification III

- If $\tau_0 \leq N(N-1)$, we can identify τ_0 by "counting" the number of positive entries of $\Pi(X;\theta_0)^{\tau_0}$.

Assumption 4 (upper bound on τ_0)

The number of rounds which took place in the network formation game between the first and second period (τ_0) is smaller than or equal to $N(N-1)$.

- Similar assumption found in Christakis et al. (2010).
- We note our main identification result holds irrespective of the bound.

Identification IV

- We'll follow an identification at infinity approach in order to identify θ (Tamer, 2003; Bajari et al., 2010).
- $X_i^u(g)$: covariates that enter the utility of agent i under g
    - $u_i(g, X) = u_i(g, X_i^u(g))$
- $X^m(g)$: covariates that enter the matching function under g
    - $\rho((i,j), g, X) = \rho((i,j), g, X^m(g))$
- $u_i(w, X_i^u(w)) - u_i(g, X_i^u(g)) =: \delta_{ij}(g, w, X_{ij}^u(g, w))$, the marginal gain from each choice, for all $g \in \mathcal{G}$, $w \in N(g)$, $g_{ij} \neq w_{ij}$
    - $X_{ij}^u(g, w)$: subvector of $[X_i^u(g), X_i^u(w)]$ with the covariates relevant in the marginal gain.

Identification V

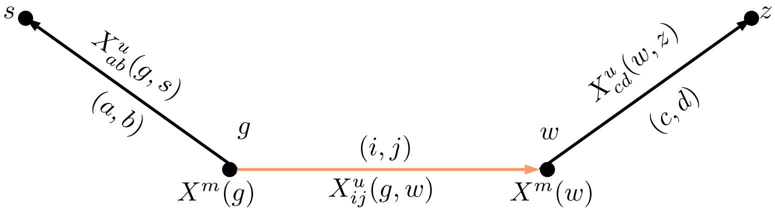
Assumption 5 (Location normalisation)

There exists some $g_0 \in \mathcal{G}$, $u_i(g_0, X_i^u(g_0)) = 0$ for all $i \in \mathcal{I}$.

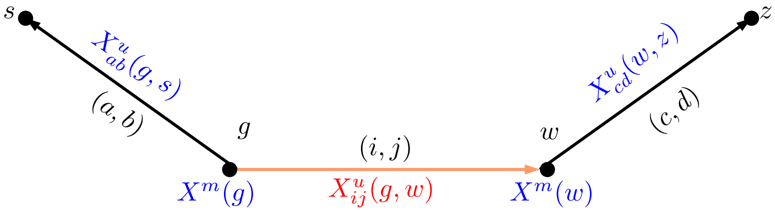
Assumption 6 (Large support exclusion restriction)

Details

Identification VI



Identification VII



$$(\Pi^{\tau_0})_{gg} = \sum_{m \in N(g) \cup \{g\}} (\Pi^{\tau_0 - 1})_{gm} \Pi_{mg}$$

Identification VIII

Theorem

Suppose Assumptions 1, 2, 3, 5 and 6 hold. Then the model is identified for any τ_0 known or identified.

Similar results established for exclusion restriction in utility.

Frequentist Estimation I

- We consider a sample $\{G_c^{T_0}, G_c^{T_0}, X_c\}_{c=1}^{C}$.
- Estimate τ_0 by "counting" the number of differing edges in each network between periods; and then taking the maximum.

$$\hat{\tau} = \max_c \{ \| G_c^{T_1} - G_c^{T_0} \|_1 \}$$

- Under iid sampling, Assumptions 1 and 2, and the bound in Assumption 4, $\hat{\tau} \overset{\text{a.s.}}{\to} \tau_0$ as $C \to \infty$.

Frequentist Estimation II

- Let vector $\beta_0 \in \mathbb{B} \subseteq \mathbb{R}^l$ encompass a parametrisation of peferences and meetings, i.e. $u_i(g, X) = u_i(g, X; \beta_0)$ and $\rho_{ij}(g, X) = \rho_{ij}(g, X; \beta_0)$ for all $(i, j) \in \mathcal{M}$, $g \in \mathcal{G}$.
- Once $\hat{\tau}$ is estimated, we can try to estimate β_0 via MLE in the second-step.
- **Difficulty:** in evaluating model likelihood, have to compute all walks between $G_c^{T_0}$ and $G_c^{T_1}$ for all c. Gruesome for even moderate values of $\hat{\tau}$.
- Simulated GMM/Indirect Inference approach also difficult: unknown low-dimensional sufficient statistics.
- We opt for a Bayesian approach known as Approximate Bayesian Computation (ABC).

Approximate Bayesian Computation (ABC) I

- Close correspondence with Nonparametric (Frequentist) Econometrics (Blum, 2010) and Indirect Inference (Frazier et al., 2018).
- **Idea:** draw parameters from prior p_0. Generate artificial sample $\tilde{\mathbf{Y}}_C := \{\tilde{G}_c^{T_1}\}_{c=1}^{C}$. Compare it with data $\mathbf{Y}_c := \{\tilde{G}_c^{T_1}\}_{c=1}^{C}$. Accept if $\|\tilde{\mathbf{Y}}_C - \mathbf{Y}_c\| \leq \epsilon$, where ϵ is tolerance.
- May choose tolerance so algorithm provides "reasonable" acceptance rate (Li and Fearnhead, 2018).
- We improve base algorithm by resorting to:
  - Importance sampling (draw from q_0 s.t. supp $p_0 \subseteq$ supp q_0). Use algorithm in Li and Fearnhead (2018) to approximate for "optimal" proposal.
  - Accept draws with probability $K(\tilde{\mathbf{Y}}_C, \mathbf{Y}_C; \epsilon)$, where K is a rescaled kernel. "Smooth" rejection rule.

Approximate Bayesian Computation (ABC) II

- In our base estimation, we use the **entire** vector of second-period network data to assess the quality of a simulated draw.
- This may lead to poor approximation of posterior moments (curse of dimensionality).
- We consider two alternative approaches that may mitigate this problem.
  - Semi-automatic construction of low-dimensional statistics as in Fearnhead and Prangle (2012).
  - Leverage independence across classrooms to implement the EP algorithm of Barthelmé and Chopin (2014). Assumes Gaussian approximation to posterior. Ameliorates curse of dimensionality by considering each classroom sequentially.
- Results are mostly similar under our three approaches

Data

- In our application, we use data from Pinto and Ponczek (2017).
- Data on 3rd and 5th grade students from 30 schools.
- Two rounds of data were collected: early and late 2014.
- Data on (directed) intraclassroom friendship networks: students were asked to nominate up to 8 friends within their classroom. ▸ Plot ▸ Summary stats
- Working sample: 161 networks (classrooms); 1,589 students.
- Strong evidence of homophily: same-sex classmates are, on average, 19.8 pp more likely to be friends, relatively to boy-and-girl pairs, conditional on age, cognitive and noncognitive skill controls. ▸ Details

Estimation I

- We use a two-step approach in order to estimate our model.
  1. Use frequentist estimator to estimate number of rounds.
  2. Apply ABC algorithm to estimate matching and utility parameters, taking the number of rounds as given.
- In the first step, we get $\hat{\tau} = 76$.
- Parametrise the meeting process as, for i and j in classroom c:

$$\rho((i,j), g, X_c) \propto \exp\left(\beta_m' W_{ij,c,0} + \delta_0 g_{ij} + \delta_1 (1 - g_{ij}) Z_{ij,c,0}\right)$$

- $W_{ij,c,0}$: pairwise distances in gender, age (in years) and measures of cognitive and non-cognitive skills between i and j at the *baseline*.
- $Z_{ij,c,0}$: distance between i and j in the alphabetically-ordered classlist. Excluded from utilities ("instrument").

Estimation II

- Specification of utilities is a linear parametrisation of Mele (2017), where $\beta_{un} = \beta_{up}$:

$$u_i(g, X_c) = \underbrace{\sum_{k \neq i} \beta_{ud}' \begin{pmatrix} 1 \\ W_{ik,c,0} \end{pmatrix} g_{ik}}_{\text{direct links}} + \underbrace{\sum_{k \neq i} \beta_{ur}' \begin{pmatrix} 1 \\ W_{ik,c,0} \end{pmatrix} g_{ki}}_{\text{mutual links}} +$$

$$+ \underbrace{\sum_{k \neq i} g_{ik} \sum_{\substack{l \neq i \\ l \neq k}} \beta_{un}' \begin{pmatrix} 1 \\ W_{il,c,0} \end{pmatrix} g_{kl}}_{\text{indirect links}} + \underbrace{\sum_{k \neq i} g_{ik} \sum_{\substack{l \neq i \\ l \neq k}} \beta_{up}' \begin{pmatrix} 1 \\ W_{kl,c,0} \end{pmatrix} g_{li}}_{\text{popularity}}$$

- Use independent zero-mean normals with $1/9$ sd as priors.
- Aim for an acceptance rate of 1%. Smooth rejection rule using rescaled normal pdf. No. of simulations is set to 100,000 (1,000 accepted).

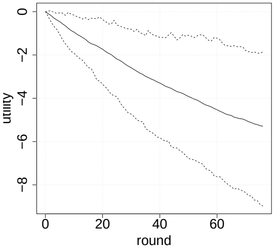
Counterfactual scenarios I

- Once we estimate our model, we consider the evolution of networks under four different sequences of matching parameters:
  1. when these are kept at their estimated value (*base* case);
  2. when random unbiased matching is imposed across networks;
  3. when, keeping grade and classroom sizes in schools fixed, we track students according to their cognitive skills (*tracking* case).
  4. when, upon meeting, friendships are formed at random w/ probability 1/2.

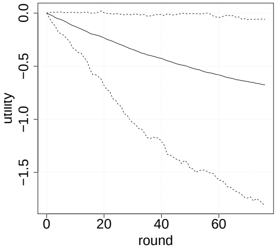
Counterfactual scenarios II

- We analyse the evolution of the aggregate utility index $\sum_{c=1}^{C}\sum_{i=1}^{N_c} u_i(g_c^t, X_c)$, from the baseline until the followup period.
- Compare aggregate welfare in counterfactual scenario minus base case.
- Rescale welfare by abs. value of avg. coefficient of homophily in age on direct links times number of students. Interpret quantities as required change in age distance of direct links in base case so students are indifferent between policies.

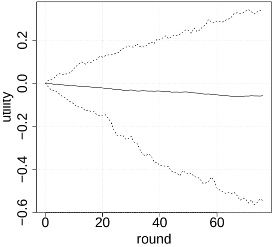
Aggregate welfare: random matching vs base case



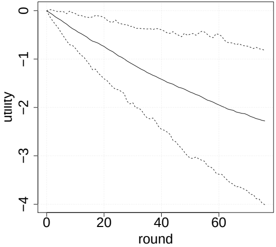
(a) Total



(b) Direct

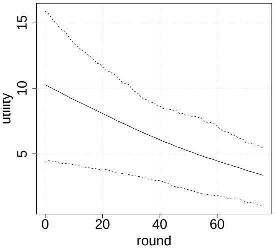


(c) Mutual

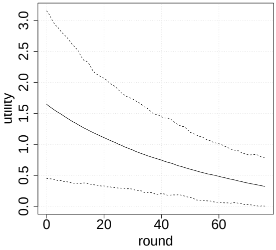


(d) Indirect/Popularity

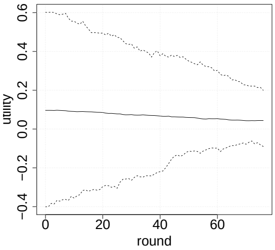
Aggregate welfare: tracking vs base case



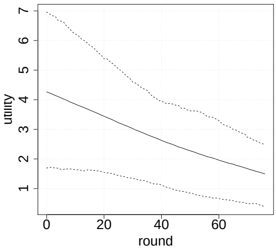
(a) Total



(b) Direct

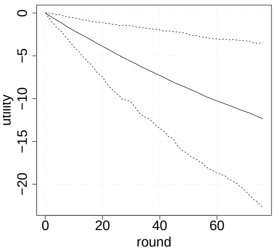


(c) Mutual

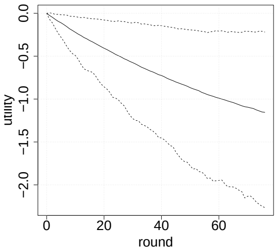


(d) Indirect/Popularity

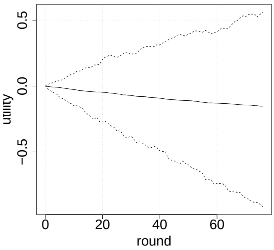
Aggregate welfare: random friendship vs base case



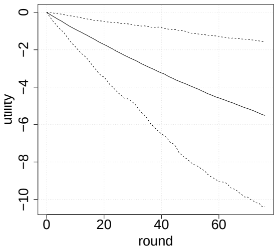
(a) Total



(b) Direct



(c) Mutual



(d) Indirect/Popularity

Concluding remarks

- In this paper, we study identification and estimation of an iterative network formation algorithm that distinguishes between homophily in preferences and homophily in meetings.
  - Showed how to identify both preference- and meeting-related parameters.
  - Estimation using a likelihood-free Bayesian approach.
- We applied our methodology to classroom network data in Northeastern Brazil.
  - Important to take meeting technology into account, as it has welfare implications over the school year.
- Previous version of paper (w/o alternative estimation methods): `www.ime.usp.br/~alvarez`.

References I

Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press.

Bajari, P., H. Hong, and S. P. Ryan (2010). Identification and estimation of a discrete game of complete information. *Econometrica 78*(5), 1529–1568.

Barthelmé, S. and N. Chopin (2014). Expectation propagation for likelihood-free inference. *Journal of the American Statistical Association 109*(505), 315–333.

Blum, M. G. B. (2010). Approximate bayesian computation: A nonparametric perspective. *Journal of the American Statistical Association 105*(491), 1178–1187.

Bramoullé, Y., S. Currarini, M. O. Jackson, P. Pin, and B. W. Rogers (2012). Homophily and long-run integration in social networks. *Journal of Economic Theory 147*(5), 1754 – 1786.

Chandrasekhar, A. and M. O. Jackson (2016). A network formation model based on subgraphs.

References II

Christakis, N. A., J. H. Fowler, G. W. Imbens, and K. Kalyanaraman (2010, May). An empirical model for strategic network formation. Working Paper 16039, National Bureau of Economic Research.

Currarini, S., M. O. Jackson, and P. Pin (2009). An economic model of friendship: Homophily, minorities, and segregation. *Econometrica 77*(4), 1003–1045.

Fearnhead, P. and D. Prangle (2012). Constructing summary statistics for approximate bayesian computation: semi-automatic approximate bayesian computation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 74*(3), 419–474.

Frazier, D. T., G. M. Martin, C. P. Robert, and J. Rousseau (2018). Asymptotic properties of approximate bayesian computation. *Biometrika 105*(3), 593–607.

References III

Goldsmith-Pinkham, P. and G. W. Imbens (2013). Social networks and the identification of peer effects. *Journal of Business & Economic Statistics 31*(3), 253–264.

Graham, B. S. (2016, April). Homophily and transitivity in dynamic network formation. Working Paper 22186, National Bureau of Economic Research.

Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica 85*(4), 1033–1063.

Jackson, M. (2010). *Social and Economic Networks*. Princeton University Press.

Jackson, M. O. and A. Watts (2002). The evolution of social and economic networks. *Journal of Economic Theory 106*(2), 265 – 295.

Li, W. and P. Fearnhead (2018). On the asymptotic efficiency of approximate bayesian computation estimators. *Biometrika 105*(2), 285–299.

References IV

McFadden, D. (1973). Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (Ed.), *Frontiers in Econometrics*, pp. 105–142. New York, NY, USA: Academic Press New York.

Mele, A. (2017). A structural model of dense network formation. *Econometrica 85*(3), 825–850.

Pinto, C. C. d. X. and V. P. Ponczek (2017). The Building Blocks of Skill Development. Unpublished.

Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *The Review of Economic Studies 70*(1), 147–165.

Assumption 6

Notation: write $A \setminus B$ for the subvector of A such that $A = [A \setminus B, B]$.

Assumption 6 (Large support exclusion restriction)

For all $g \in \mathcal{G}$, $w \in N(g)$, $g_{ij} \neq w_{ij}$, there exists a $m \times 1$ subvector $Z^u_{ij}(g, w)$ of $X^u_{ij}(g, w)$, i.e. $X^u_{ij}(g, w) = [Z^u_{ij}(g, w), \tilde{X}^u_{ij}(g, w)]$, such that no covariate in $Z^u_{ij}(g, w)$ is an element of $[X^m(g), X^m(w), (X^u_{kl}(g, [1 - g_{kl}, g_{-kl}]), X^u_{kl}(w, [1 - w_{kl}, w_{-kl}]))_{(k,l) \neq (i,j)}]$. Moreover, $Z^u_{ij}(g, w)$ admits a conditional Lebesgue density $f(Z^u_{ij}(g, w) | X \setminus Z^u_{ij}(g, w))$ that is positive a.e. (for almost all realisations of $[X \setminus Z^u_{ij}(g, w)]$); and there exists $\vec{r} \in \mathbb{R}^m$ s.t. $\lim_{t \to \infty} \delta_{ij}(g, w, [Z^u_{ij}(g, w) = t\vec{r}, \tilde{X}^u_{ij}(g, w)]) = \infty$.

Return

Possible parametrisation

Example (Possible Parametrisation)

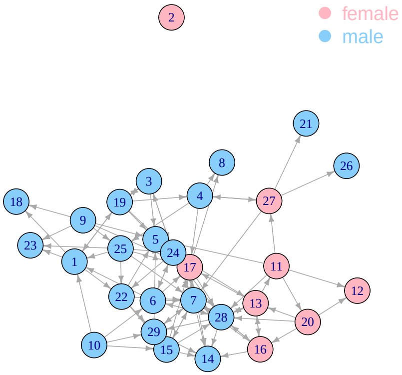
If we take

$$\rho((i,j), X^m(g)) = \frac{\exp(\alpha_g' X_{ij}^m(g))}{\sum_{(k,l)\in\mathcal{M}} \exp(\alpha_g' X_{kl}^m(g))}$$

$$u_i(g, X_i^u(g)) = \beta_g' X_i^u(g)$$

where $X_i^u(g)$ may include other individuals' characteristics, then the model is identified under the previous restrictions, provided the usual rank conditions hold (cf. Amemiya (1985, p.286-292); also McFadden (1973))

Data: visualisation



Return

Data: summary statistics

Table: Pairwise distance in covariates – Summary statistics

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
edge (baseline)	17,736	0.172	0.377	0	0	0	1
edge (followup)	17,736	0.179	0.383	0	0	0	1
distance in classlist (baseline)	17,736	8.009	5.553	1	3	12	30
distance in age (years) (baseline)	17,736	0.827	0.908	0.000	0.249	1.041	6.633
distance in gender (baseline)	17,736	0.502	0.500	0	0	1	1
distance in cognitive skills (baseline)	17,736	0.099	0.081	0.000	0.035	0.144	0.530
distance in conscientiousness (baseline)	17,736	0.568	0.485	0.000	0.207	0.798	3.275
distance in neuroticism (baseline)	17,736	0.676	0.538	0.000	0.249	0.976	3.491

Return

Dyadic regressions I

- Is there evidence of homophily in our data?
- Run regressions of the type:

$$g_{ij,c,1} = \beta' W_{ij,c,0} + \alpha_i + \gamma_j + \epsilon_{ij,c}$$

- $g_{ij,c,1} = 1$ if, in classroom c, individual i nominates j as a friend *at the followup period*.
- $W_{ij,c,0}$: pairwise distances in gender, age (in years) and measures of cognitive and non-cognitive skills between i and j at the *baseline*.

Dyadic regressions II

	Dependent variable:	
	edge	
	(1)	(2)
distance in classlist		-0.002^{***}
		(0.0005)
distance in age	-0.018^{***}	0.002
	(0.006)	(0.003)
distance in gender	-0.198^{***}	-0.186^{***}
	(0.006)	(0.006)
distance in cognitive skills	-0.254^{***}	-0.127^{***}
	(0.049)	(0.034)
distance in conscientiousness	-0.031^{***}	-0.019^{***}
	(0.009)	(0.006)
distance in neuroticism	-0.017^{**}	-0.023^{***}
	(0.008)	(0.005)
Sender fixed effects?	Yes	No
Receiver fixed effects?	Yes	No
Time effect?	Yes	No
Observations	17,736	17,736

Note: *p<0.1; **p<0.05; ***p<0.01
Standard errors clustered at the classroom-level in parentheses.

Back to data

Homophily patterns under base and counterfactual policies

Table: Homophily measures under base counterfactual policies

	Base case	Random matching	Tracking	Random friendship
	Regression coefficients			
distance in classlist	-0.0068	-0.0016	-0.0124	-0.0096
	[-0.0097;-0.0035]	[-0.0024;-6e-04]	[-0.0156;-0.0086]	[-0.0122;-0.0054]
distance in age	-0.0146	-0.0161	-0.0152	0.0011
	[-0.0196;-0.0085]	[-0.0219;-0.0103]	[-0.024;-0.0053]	[-0.0047;0.0069]
distance in gender	-0.0921	-0.1077	-0.0421	-0.0938
	[-0.1068;-0.0775]	[-0.1197;-0.0958]	[-0.0532;-0.0309]	[-0.1055;-0.0849]
distance in cognitive skills	-0.0955	-0.1342	0.0322	-0.1096
	[-0.1485;-0.0446]	[-0.1982;-0.0675]	[-0.0468;0.1613]	[-0.1557;-0.0529]
distance in conscientiousness	-0.0122	-0.0117	-0.0079	-0.0112
	[-0.0259;0.0027]	[-0.0267;0.0025]	[-0.0249;0.0053]	[-0.0215;0.0013]
distance in neuroticism	-0.0111	-0.0106	-0.0102	-0.0086
	[-0.0201;-0.0039]	[-0.0215;-0.0023]	[-0.0205;0.005]	[-0.0251;0.0046]
	Degree summary statistics			
Avg. degree	0.1935	0.23	0.1709	0.2479
	[0.1687;0.2241]	[0.207;0.261]	[0.1515;0.1967]	[0.2136;0.2802]
Var. degree	0.1559	0.177	0.1416	0.1862
	[0.1402;0.1739]	[0.1641;0.1929]	[0.1285;0.158]	[0.168;0.2017]