Strategic Network Formation

May 19, 2017

1 Literature Review

1.1 The Context

The study of networks formation is an application of network theory, a subfield of graph theory, that is focused on the analysis of relational data. Its importance relies on the well-established correlation of socioeconomic outcomes between individuals of the same network. In spite of this increasing body of literature, economics has been somewhat oblivious to the key role of social networks until two decades ago [Choi et al., 2015]. Jackson and Wolinsky [1996] and Bala and Goyal [2000] are the seminal papers on network formation which provide different approaches to the characterization of equilibrium in models of networks and their social efficiency. Some previous studies without a systematic view of the topic include Myerson [1977], Kirman [1983] and Montgomery [1991]. Since then, network formation literature has been concentrated on theoretical developments leaving aside the validation of many useful findings through empirical research.

One of the main and recent theoretical contributions is presented in Bramoullé et al. [2014]. Through their work, a large class of games is studied leading to an striking result that unveils the importance of the so-called lowest eigenvalue[1]. This single measure is able to capture the behavior of agents towards other agents' actions. For instance, they are able to identify the behavior driven from the interaction of different firms in small markets. Through this finding, the authors are able to characterize all possible Nash equilibria in different structured games. Some applications in R&D and crime can be found in their paper.

Some of the most relevant empirical evidence is focused on the linkages between outcomes and peers as we have mentioned before. For instance, the findings of Christakis and Fowler (2007) reveal that changes in weight of individuals is a predictor of weight variations in their peers. Furthermore, Uzzi and Sprio (2005) find that network configurations are correlated with improved group performance. Specifically, in the literature of peer effects, many papers have found that educational outcomes and behavior of individuals may be defined by their classmates (Angrist and

[1] To put in context, networks can be represented as matrices where each row/column is an individual.

Lang, 2004; Carrell, Fullerton, and West, 2009). Consider also that most of the literature on peer effects have defined networks as easily measurable characteristics, e.g., being in the same class [2]. Nonetheless, there is no much work in progress dealing with the formation of networks *per se*, i.e., what leads a pair of agents to establish a link of friendship/partnership between both of them.

Due to this lack of information, it seems important to shed light on the mechanism of networks formation, so that policymakers could be able to understand how the possible manipulations can affect networks and consequently bear upon the outcomes of interest. The first models addressing this question were developed in the game theory literature and usually implied that networks formation are at least partly a result of random shocks (through randomly arising opportunities for forming links), so that established links are partially exogenous, even if individuals optimally decide to form a link when such opportunities arise (see, for example, Myerson, 1977; Auman and Myerson, 1988).

Another interesting field of interest, but scarcely studied due to data complexity, is the estimation of exogenous variations effect on networks formation. For instance, Phan and Arioldi (2015) dealt with data restriction issues using a large sample of college students' Facebook posts in order to evaluate the effect of Hurricane Ike on social structure. They find that affected individuals are more likely to strengthen interaction, but maintain the same number of friends as unaffected individuals.

Choi et al. [2015] acknowledges that the key reasons for the lack of empirical support is the complexity of causal identification and the scarcity of sound and comprehensive data. In this context, experimental research has emerged as a great tool to validate theoretical findings. Developments in this area include topics as wide as strategic communication (Hagenback and Koessler, 2010), learning in a network context [Acemoglu et al., 2011; Mueller-Frank and Neri, 2013], asymmetric information (Galeotti et al., 2013), the role of information on markets (Cassar et al., 2010), among others. However, none of these studies have applications related to the network formation in the school and its effect on educational outcomes as far as we are concerned.

1.2 Modeling

The modeling of the formation of social networks is usually motivated by a utility maximization framework, so that a relationship between individuals is established if such link is welfare-increasing for both of them. Such decision may be determined by random events as by individuals characteristics and the current state of the network (e.g. links previously established). Christakis et al. [2010] add an *ex ante* opportunity link component defined as the degree of separation between the potential friends which is a measure of the number of friends (in common) they already have.

Commonly the modeling literature of network formation can be classified in two approaches.

[2]Christakis et al. [2010] mentions that further and revealing results can be found with a sample of individuals who identify themselves as connected through friendship or other social networks.

The first approach is characterized by the design and implementation of Exponential Random Graph (ERG) models. Models in the second approach are referred to as Strategic Network Formation (SNF) models.

The former approach focuses on the distributions for the matrix of established relationships [3]. The simplest and first model developed in this strand of the literature was the Erdös-Reny model, where the probability of establishing any link is constant for all potential relationships. Notwithstanding, further extensions have been developed to relax this assumption. Despite these improvements, Christakis et al. [2010] acknowledge two unattractive features of the Erdös-Reny model and its extensions.

First, the simulation of counterfactual designs after estimation may turn convoluted since there is no clear reason to explain why the parameters of the models must remain the same after structural changes, e.g., the prediction of networks given alternative rules for students assignment between classes. Second, the estimation of these models is not completely clear once heterogeneity between nodes is included. A discussion about the accuracy of new estimation methods proposed can be found in Kolaczyk (2009).

The second approach to modeling networks is based on the Strategic Network Formation (SNF) models (Jackson, 2009) also referred to as Network Evolution Models (Toivonen et al., 2009). The key feature of these models is the recognition that the probability of establishing a link is at least partially a function of individual choices. These models assume that links between individuals are established, conditional on an opportunity for such a link arising, if both individual view these links as beneficial.

Christakis et al. [2010] develop an empirical model for strategic network formation as a process where in each period a single randomly pair of agents has the opportunity to establish a link. Conditional on this opportunity, a link will be established if both agents evaluate the link as an upgrading welfare event. Such decision is based on own characteristics, potential partner's characteristics, and on features of the current state of the network, for instance, whether the two potential partners already have friends in common. Given this structure, the objective is to recover the fundamental parameters of preference that drive the agent's decision.

Mele [2016] proposes a model of network formation that combines features from both the ERG and SNF models. Mele takes into account the fact that the individuals' utilities depend not only on payoffs from direct links, but also on link externalities, e.g., reciprocity, indirect friends, popularity. In contrast to the static equilibrium approach of SNF models, network formation is dynamic since in each period individuals decides whether not only to form a new link, but also to keep an existing or to not form any link. This process may seem more adequate in contexts of high rate entry of new individuals to the universe of players. This paper provides a feasible estimator of the structural

[3] Commonly referred to as *adjacency matrix* in the literature. Usually, papers working in this direction focus on undirected links, so that the adjacency matrix is symmetric.

parameters considering the likelihood used in standard ERG models.

1.3 Applications: Education and others

There is a limited number of applications of strategic network theory on educational environments. Notwithstanding, the available results on the literature reveal important features of the dynamics of social networks in education.

Calvó-Armengol et al. [2009] study the effect of the centrality of an individual within her network, measured as the discounted sum of the number of direct and indirect friends she has[4], on educational outcomes. In the Nash equilibrium, they demonstrate that the optimal reply function of an individual is an additively separable function that depends on her observable characteristics, the number of direct peers and their observable characteristics and peers' effort. Therefore, in equilibrium, the individuals' outcome can be separated in a idiosyncratic component and a peer effect component. With this set up, they find that a standard deviation increase in the centrality measure increases the student school performance by more than 0.07 standard deviations.

Banerjee et al. [2013] and Banerjee et al. [2016] had made an important application of network formation and diffusion of information in contexts of microfinance programs. They study how the proper information and information provided by acquaintances determine the decision of participation in the program. In that sense, they implement a measure of *diffusion centrality* which provides a prediction of the extent to which piece of information seeded to a network member diffuses in finite time.

Section 2 presents descriptive statistics of our data replicating some tables and features noted in this literature review. Section 3 presents results from the the first steps of the model implemented by Calvó-Armengol et al. [2009]. We develop a summary of the model, but further details can be found in their paper.

[4]Specifically, they use the Katz-Bonacich network centrality measure.

2 Descriptive Statistics

Table 1 presents descriptive statistics and information regarding friendship links. A link is defined as a pair of students (i, j) in the same network -in our case, each classroom is considered a network. A link between i and j is established if i lists j as her friend and viceversa[5]. Furthermore, let N_k be the number of students in classroom k. Then, the number of potential (established and non-established) links is $\binom{N_k}{2} = \frac{N_k!}{2!(N_k-2)!}$ in classroom k[6].

Table 1: Descriptive Statistics: Links

	Treatment status		
	Control	**Tracking**	**Mixing**
Number of schools and shifts	57	57	57
Number of students	5071	9085	5056
Number of classrooms	170	300	171
Mean # of potential links	928	968	919
Mean # of established links (censored)	56	58	56
Rate of established links (%, censored)	7.1	7.2	7.2
Rate of non-identified links (%)	1.5	1.4	1.5
Mean number of friends (P10)	7.4	7.5	7.5
Mean number of listed friends (P11)	2.9	2.8	2.9

Table 2 shows that there is a negligible difference in the number of friends listed between treatment arms. Since students were allowed to list three friends at most, our measure for the number of established links is potentially censored. Thus, the estimates for centrality measures and clustering coefficients might be inconsistent. Furthermore, Table 1 shows that the mean number of friends in each treatment arm is around 7.

Table 2: Number of friends listed by treatment status

	Treatment status		
	Control	**Tracking**	**Mixing**
One friend listed	4.8%	5.4%	3.9%
Two friends listed	6.0%	5.9%	5.5%
Three friends listed	89.2%	88.7%	90.6%
Total	100%	100%	100%

[5] Here we assume that links are undirected in the following fashion: if i has an established link with j, then j necessarily has an established link with j. Calvó-Armengol et al. [2009] show that there is no significant difference in the results by working with directed and undirected links data.

[6] It is important to recall that, when working with network data, the units of analysis are the links instead of individuals.

Given a link (i,j), $x_i = x_j$[7] implies that i and j share a common characteristic x. Table 3 shows the proportion of common characteristics among (i) all potential links, (ii) best friends links, (iii) friends links and (iv) non-established links.

Some of the most relevant pre-treatment differences between potential -all- and established links -best friends and friends- in a same treatment arm are explained by the primary school of origin, gender and primary GPA. However, there are no major differences in these gaps between treatment arms. Table 4 presents a similar analysis for continuous variables considering a ± 0.5 standard deviation range instead of 1 standard deviation.

Table 3: Descriptive Statistics: First Degree Links (percentages)

	Treatment status											
	Control				**Tracking**				**Mixing**			
	A	BF	F	NF	A	BF	F	NF	A	BF	F	NF
Same primary school	6	9	8	6	6	10	9	5	6	9	8	5
Same age	51	51	51	51	52	54	53	52	51	52	51	51
Same gender	50	85	79	45	51	85	79	47	50	84	79	45
Same heigth (+/- 1 SD)	57	59	59	56	57	60	59	56	57	60	59	56
Same weight (+/- 1 SD)	52	54	54	52	53	56	55	52	53	55	54	53
Same mother's education level	27	30	29	27	26	27	27	26	26	28	27	26
Both mothers have a HE degree	2	2	2	2	2	2	2	2	2	2	2	2
Same primary GPA (range)	34	38	37	34	37	40	39	37	34	39	37	34
Same IDANIS BL (+/- 1 SD)	59	60	60	59	89	89	89	88	56	60	59	56
Same IDANIS EL (+/- 1 SD)	57	61	60	57	71	74	73	71	58	64	62	57
Same GPA (+/- 1 SD)	54	64	61	53	56	66	64	55	53	64	62	52
Same Math grade (+/- 1 SD)	57	66	63	56	61	67	66	60	56	64	63	55
Same Spanish grade (+/- 1 SD)	56	65	62	55	58	66	64	57	56	66	63	54
Same # of friends (+/- 1 SD)	41	47	46	40	41	47	46	41	43	47	47	42
Same hours studying Spanish	50	51	51	49	50	52	51	49	50	54	53	50
Same hours studying Math	56	57	57	55	55	57	56	55	56	58	57	56
Same hours studying	54	57	56	54	54	57	55	54	55	58	56	55
Same risky behavior	53	60	58	53	53	60	58	52	53	60	59	52
Same smoking habits	74	77	75	74	74	77	76	74	74	77	76	74
Same alcohol consumption habits	60	63	62	60	61	64	63	60	59	63	61	59

NOTE: A: All - indicates the proportion of all potential links sharing the correspondig characteristic. BF: Best friends - indicates the proportion of established links between best friends sharing the corresponding characteristic. F: Friends - indicates the proportion of established links between individuals sharing the corresponding characteristic. NF: Not friends - indicates the proportion of non-established links between individuals sharing the corresponding characteristic.

[7] For continuous variables, i and j share a common characteristic if $x_j - \sigma_x \leq x_i \leq x_j + \sigma_x$ or if $x_i - \sigma_x \leq x_j \leq x_i + \sigma_x$, where σ_x is the standard deviation of variable x for each treatment arm.

Table 4: Descriptive Statistics: First Degree Links (percentages)

	Treatment status											
	Control				**Tracking**				**Mixing**			
	A	BF	F	NF	A	BF	F	NF	A	BF	F	NF
Same heigth (+/- 0.5 SD)	31	35	33	31	31	34	33	31	31	34	32	31
Same weight (+/- 0.5 SD)	30	30	31	30	30	32	32	30	30	32	31	30
Same IDANIS BL (+/- 0.5 SD)	32	33	33	32	79	79	79	78	35	39	37	34
Same IDANIS EL (+/- 0.5 SD)	29	32	31	29	43	46	45	43	34	38	37	33
Same GPA (+/- 0.5 SD)	29	36	34	28	31	37	36	30	29	37	35	28
Same Math grade (+/- 0.5 SD)	33	40	38	32	35	41	39	35	33	39	38	32
Same Spanish grade (+/- 0.5 SD)	32	39	38	31	33	40	39	33	32	42	39	31
Same # of friends (+/- 0.5 SD)	30	37	35	29	30	35	34	29	30	36	35	29

NOTE: A: All - indicates the proportion of all potential links sharing the correspondig characteristic. BF: Best friends - indicates the proportion of established links between best friends sharing the corresponding characteristic. F: Friends - indicates the proportion of established links between individuals sharing the corresponding characteristic. NF: Not friends - indicates the proportion of non-established links between individuals sharing the corresponding characteristic.

Tables 5, 6 and 7 show the distribution of established links by performance level of the main individual and her friends. Performance is defined by pre-treatment characteristics.

Table 5: Established links of first degree by performance (percentage)

	Control		
	Low performance	**Medium performance**	**High performance**
% friends of low performance	35.5	32.1	29.1
% friends of medium performance	37.0	37.1	37.2
% friends of high performance	27.5	30.7	33.7

Table 6: Established links of first degree by performance (percentage)

	Tracking	
	Low performance	**High performance**
% friends of low performance	99.6	0.3
% friends of high performance	0.4	99.7

Table 7: Established links of first degree by performance (percentage)

	Mixing	
	Low performance	**High performance**
% friends of low performance	44.9	38.7
% friends of high performance	55.1	61.3

Table 8 shows the common characteristics analysis (presented in previous tables) for variables constructed trough factor analysis[8].

Table 8: Descriptive Statistics: Links and Behavioral Factors (percentage)

	Treatment status											
	Control				**Tracking**				**Mixing**			
	A	BF	F	NF	A	BF	F	NF	A	BF	F	NF
Same extroversion degree (+/- 1 SD)	38	40	39	38	36	39	39	36	37	39	40	37
Same sociality level (+/- 1 SD)	43	43	43	43	43	45	45	44	42	44	44	42
Same external comparison degree (+/- 1 SD)	41	41	40	41	41	42	42	41	42	42	41	42
Same individualization level (+/- 1 SD)	36	38	37	36	37	37	37	37	36	38	37	36
Same cooperation degree (+/- 1 SD)	37	40	39	37	38	41	39	38	37	40	38	37
Same competitiveness degree (+/- 1 SD)	36	36	36	36	35	37	36	36	35	38	37	35

Note: A: All - indicates the proportion of all potential links sharing the correspondig characteristic. BF: Best friends - indicates the proportion of established links between best friends sharing the corresponding characteristic. F: Friends - indicates the proportion of established links between individuals sharing the corresponding characteristic. NF: Not friends - indicates the proportion of non-established links between individuals sharing the corresponding characteristic.

[8] For more detail of the variables considered in the factor analysis see Tables 9 and 10.

Tables 9 and 10 present the same analysis of previous tables for factor analysis items.

Table 9: Descriptive Statistics: Links, Common Relevant Attitudes (percentages)

	Treatment status											
	Control				**Tracking**				**Mixing**			
	A	BF	F	NF	A	BF	F	NF	A	BF	F	NF
Attending classes	88	90	88	88	87	89	89	86	88	91	89	88
Study for tests	85	86	86	85	85	87	87	85	87	89	88	87
Practice sports	56	61	59	55	59	62	60	58	60	62	62	59
Obtain good grades	94	96	95	94	94	96	95	94	94	95	95	94
Be popular	63	67	65	63	62	65	64	62	61	65	63	61
Have a boy/girlfriend	66	70	67	66	65	69	68	65	63	69	67	63
Gather with friends	51	52	52	51	51	53	53	51	50	55	52	50
Partying	64	67	65	64	62	65	65	62	61	65	63	61
Finish school	94	95	95	94	94	95	95	94	94	95	95	94
Get a HE degree	92	93	93	92	93	94	94	92	94	96	96	94
Earn money	71	73	72	71	70	70	70	70	71	73	71	71
Get a good job	100	100	100	100	100	100	100	100	100	100	100	100
Go to the park	50	50	50	50	50	52	51	50	50	53	51	50
Go to the movies	50	52	52	51	51	53	52	51	50	52	52	50
Go to dance (during free time)	73	73	73	74	72	74	73	72	75	76	76	75
Play videogames	49	57	54	49	50	57	55	49	49	55	55	49
Go shopping	50	53	51	50	50	54	52	50	50	50	53	50
Practice sports (during free time)	61	64	63	61	61	63	63	61	61	64	63	60
Gather with friends (during free time)	50	54	52	50	50	53	52	50	50	55	54	50
Have a date (during free time)	63	65	64	62	65	67	66	65	63	67	66	63
Connect to the internet	60	63	63	60	60	64	62	61	61	64	63	61
Watch TV	54	54	55	54	53	56	55	54	54	56	56	54
Read	52	55	54	52	51	54	53	51	52	57	55	52

NOTE: A: All - indicates the proportion of all potential links sharing the correspondig characteristic. BF: Best friends - indicates the proportion of established links between best friends sharing the corresponding characteristic. F: Friends - indicates the proportion of established links between individuals sharing the corresponding characteristic. NF: Not friends - indicates the proportion of non-established links between individuals sharing the corresponding characteristic.

Table 10: Descriptive Statistics: Links, Agreement Statements (percentage)

	Treatment status											
	Control				**Tracking**				**Mixing**			
	A	BF	F	NF	A	BF	F	NF	A	BF	F	NF
Be the life of the party	54	57	55	53	53	56	54	52	53	57	55	53
Be empathetic	81	83	83	81	83	84	84	83	83	87	86	84
Share my toys with others	75	75	74	74	73	75	74	73	73	76	75	74
Avoid to stand out among others	50	52	51	50	50	53	51	50	50	52	51	50
Help when someone is in need	76	77	76	76	78	79	79	78	79	80	80	79
Have little to say	49	50	50	49	49	50	50	49	49	50	51	50
Other people bully me	61	66	64	62	61	65	63	60	61	65	63	60
Avoid to draw attention	50	50	49	50	50	53	52	50	51	53	53	50
Need to compare my performance	51	54	53	51	51	52	52	51	50	52	51	50
Quiet when surrounded by not familiar people	54	55	56	54	54	56	55	54	54	56	55	54
Like feedback	57	59	57	57	57	60	58	57	57	58	57	57
Prefer to be alone	55	57	57	55	54	57	57	54	53	56	55	53
Feel comfortable surrounded by people	59	61	61	59	59	61	60	58	60	62	61	59
At least, I have a good friend	84	85	84	84	84	85	84	84	83	85	84	83
Start the conversations	54	58	55	55	54	57	56	55	55	58	57	55
Usually, I get on well with others	77	78	79	76	75	79	78	75	79	80	80	79
Be kind with younger people	85	87	86	84	84	86	86	84	85	88	85	85
Be social	51	54	53	51	51	53	53	51	51	55	53	51
Offer help	64	67	65	63	62	65	63	63	63	65	64	63
Like to know peers' performance	53	54	53	53	53	56	55	53	54	56	55	55
Ask peers about their performance	59	59	60	59	60	61	61	61	60	63	62	59
Talk a bit	50	55	53	50	50	52	50	50	49	52	51	50
Like feedback and compare with others	53	55	53	53	54	56	55	54	54	54	55	54
Like to stand aout	55	56	57	54	55	57	57	55	55	58	56	55
Only need to know my own grades	66	69	68	66	66	68	67	66	66	69	68	66
Get on well with adults rather than kids	53	53	54	53	52	56	55	53	52	54	53	52

NOTE: A: All - indicates the proportion of all potential links sharing the correspondig characteristic. BF: Best friends - indicates the proportion of established links between best friends sharing the corresponding characteristic. F: Friends - indicates the proportion of established links between individuals sharing the corresponding characteristic. NF: Not friends - indicates the proportion of non-established links between individuals sharing the corresponding characteristic.

Table 11 applies the analysis of common characteristics with the seating chart data. However, it is important to acknowledge that seats are not assigned and in most of the cases students are allowed to choose their own seat.

Table 11: Descriptive Statistics: Links and Seating Chart (percentage)

	Treatment status											
	Control				**Tracking**				**Mixing**			
	A	BF	F	NF	A	BF	F	NF	A	BF	F	NF
Seated together	16	36	30	14	15	34	28	14	16	33	28	14
Same column	4	8	7	4	4	8	7	4	4	9	8	4
Same row	4	14	10	3	4	13	10	3	4	12	9	3
Same diagonal	7	9	10	7	7	8	9	7	7	9	9	7

NOTE: A: All - indicates the proportion of all potential links sharing the correspondig characteristic. BF: Best friends - indicates the proportion of established links between best friends sharing the corresponding characteristic. F: Friends - indicates the proportion of established links between individuals sharing the corresponding characteristic. NF: Not friends - indicates the proportion of non-established links between individuals sharing the corresponding characteristic.

Table 12: Descriptive Statistics: Best Friends

	Treatment status		
	Control	**Tracking**	**Mixing**
Boys: mean share of boys (%)	87.6	86.6	86.2
Girls: mean share of boys (%)	7.6	8.5	9.4
# months as friends	10.3	10.9	10.7
Lives close to your house (%)	26.0	27.2	27.4
Have ever been to his/her house (%)	28.2	31.6	28.7
Has ever been to your house (%)	26.0	28.9	26.6
In the past 7 days, saw him/her away from the school (%)	38.3	37.3	37.2
In the past 7 days, talked to him electronically (%)	67.4	68.7	68.6

NOTE: Each individual identifies one best friend so that unit of analysis are individuals in this case. First row denotes the share of male best friends among boys. Second row denotes the share of male best friends among girls.

Table 13: Mixing vs Control low performers: Treatment effect on friend choices

	(1)	(2)	(3)
	Actual (A) (se)	As Random (R) (se)	Difference (P[A<R])
% friends of low performance	0.145	0.133	0.012***
	[0.012]***	[0.010]***	[0.160]
% friends of medium performance	-0.314	-0.321	0.007***
	[0.017]***	[0.006]***	[0.140]
% friends of high performance	0.170	0.188	-0.018***
	[0.013]***	[0.010]***	[0.960]
Best friend: low performance	0.147	0.132	0.016***
	[0.016]***	[0.015]***	[0.160]
Best friend: medium performance	-0.379	-0.378	-0.001
	[0.031]***	[0.015]***	[0.560]
Best friend: high performance	0.178	0.182	-0.004***
	[0.016]***	[0.014]***	[0.620]

NOTE: * significant at 10%; ** significant at 5%; *** significant at 1%. Coefficients on column (2) represent the mean of the distribution of 100 simulated treatment effects. In each simulation, 3 friends were randomly assigned to each student. Column (3) shows the difference between columns (1) and (2). The significance in column (3) is denoted by the stars next to each difference. The brackets in column (3) contain empirical p-values, which represent the number of occasions in which the simulated treatment effect was greater than the actual treatment effect divided by the number of simulations (i.e. 100). The coefficient for the first variable in column (1) indicates that the share of friends with low performance is 14.5pp greater for treated (mixing) students in comparison to control students, while the coefficient for the fourth variable in column (1) indicates that the probability of having a low performer as best friend increases by 14.7pp.

3 Network Model and Measuring Centrality

In this section, we present a replication of the model implemented by Calvó-Armengol et al. [2009].

3.1 Set-up

Let K be the number of networks in the economy. Every network g_k must satisfy two conditions. First, two agents in a network g_k are either directly linked, or indirectly linked through a sequence of agents in g_k (*connectedness* requirement). Second, two agents in different networks, g_k and $g_{k'}$ cannot be connected through any such sequence (*maximality* requirement).

Considering the Nash equilibrium characterized in Calvó-Armengol et al. [2009], we have the empirical counterpart of the best reply function for each $i = 1, \ldots, n^k$ in $k = 1, \ldots, K$[9]:

$$y_{i,k} = \sum_{m=1}^{M} \beta_m x_{i,k}^m + \frac{1}{g_{i,k}} \sum_{m=1}^{M} \sum_{j=1}^{n_k} \gamma_m g_{ij,k} x_{j,k}^m + \eta_k + \varepsilon_{i,k} \tag{1}$$

$$\varepsilon_{i,k} = \mu g_{i,k} + \phi \sum_{j=1}^{n_k} g_{ij,k} \varepsilon_{j,k} + v_{i,k} \tag{2}$$

where $y_{i,k}$ is the individual i's GPA in the network k. $x_{i,k}^m$ is a set of M control variables accounting for observable differences in individual characteristics[10], $g_{ij,k}$ is a dichotomous variable that is equal to 1 when individuals i and j in network g_k have a link established, $g_{i,k} = \sum_{j=1}^{n_k} g_{ij,k}$ is the number of direct links of i, $\frac{\sum_{j=1}^{n_k} g_{ij,k} x_{j,k}^m}{g_{i,k}}$ is the set of average values of the M controls of i's direct friends (i.e. contextual effects), η_k is an (unobserved) network-specific component which might be correlated with the regressors.

Equation 2 describes the process of $\varepsilon_{i,k}$, which is the residual of individual i's level of activity in the network g_k, that is not accounted for either by individual heterogeneity and contextual effects or by (unobserved) network-specific components. Here, $\sum_{j=1}^{n_k} g_{ij,k}\varepsilon_{j,k}$ is the spatial lag term and ϕ is the spatial autoregressive parameter.

3.2 Katz-Bonacich centrality

The Katz-Bonacich centrality measures the importance of a given node (i.e. individual) in a network, which can be understood as the discounted sum of established links.

$$katz = \sum_{p=1}^{+\infty} \phi^p g_{i,k}^p \tag{3}$$

[9] Note that $\sum_{k=1}^{K} = n$.

[10] Preliminarily, we include gender, height and weight.

where ϕ is a non-negative discount factor and $g_{i,k}^{p}$ is the number of p-links away friendships from i. For instance, $g_{i,k}$ is equal to the number of direct links of i, $g_{i,k}^{2}$ is equal to the number of two-links away friendships and so on. If ϕ is small enough, then Equation 3 converges to

$$katz = \frac{\phi g_{i,k}}{1 - \phi g_{i,k}} \tag{4}$$

Note the lower bound $katz \geq \phi g_{i,k}$. In addition, Debreu and Herstein [1953] demonstrate that the upper bound of the scalar ϕ is given by the inverse of the largest eigenvalue of the adjacency matrix.

3.3 Empirical Strategy

Preliminarily, we consider a simplified approach of the strategy presented in Calvó-Armengol et al. [2009]. First, we estimate Equation 1 in order to obtain residuals $\hat{\varepsilon}_{i,k}$, which are an input to estimate Equations 1 and 2 jointly for each network g_k. Specifically, we estimate

$$y_{i,k} = \sum_{m=1}^{M} \beta_m x_{i,k}^{m} + \frac{1}{g_{i,k}} \sum_{m=1}^{M} \sum_{j=1}^{n_k} \gamma_m g_{ij,k} x_{j,k}^{m} + \mu g_{i,k} + \phi \sum_{j=1}^{n_k} g_{ij,k} \hat{\varepsilon}_{j,k} + v_{i,k} \tag{5}$$

Thus, we obtain k different values for $\hat{\phi}_k$. Then, we calculate the largest eigenvalue for each network g_k, $w(g_k)$ and keep in the sample those networks that satisfy $\hat{\phi}_k < \frac{1}{w(g_k)}$. Then, we stack all the remaining networks to estimate Equation 5 including network fixed effects η_k.

$$y_{i,k} = \sum_{m=1}^{M} \beta_m x_{i,k}^{m} + \frac{1}{g_{i,k}} \sum_{m=1}^{M} \sum_{j=1}^{n_k} \gamma_m g_{ij,k} x_{j,k}^{m} + \eta_k + \mu g_{i,k} + \phi \sum_{j=1}^{n_k} g_{ij,k} \hat{\varepsilon}_{j,k} + v_{i,k} \tag{6}$$

Thus, we obtain one value of $\hat{\phi}^{pool}$ for all the networks in the sample. In addition, we estimate Equation 6 for each treatment arm, so that we calculate three different estimates of ϕ: $\hat{\phi}^{C}$, $\hat{\phi}^{T}$ and $\hat{\phi}^{M}$.

With the values of $\hat{\phi}_k$, $\hat{\phi}^{pool}$ and the ones estimated by each treatment arm ($\hat{\phi}^{C}$, $\hat{\phi}^{T}$ and $\hat{\phi}^{M}$), we calculate three different Katz-Bonacich centrality measures.

3.4 Clustering coefficient

The general idea of the Clustering coefficient is to consider transitive relations; that is, if node j is connected to node i and i is connected to k, then j is also connected to k. If this proposition is true, then (i, j, k) is a closed triplet. In short, the clustering coefficient is equal to the proportion of closed triplets among all possible triplets.

3.5 Communities

A community is a complex definition in Network Science. For the sake of brevity, we will understand communities as a sub-structure of a network which comprehends a cohesive group of individuals. Figure 2 shows the number of communities in each network by treatment arm and the mean size (individuals) of the community.

4 Results

Table 14 presents the estimates of the Clustering coefficient, communities, and Katz-Bonacich centrality measure with estimates for ϕ. Table 15 presents an analysis of difference in means between our measures of Katz-Bonacich centrality. Table 16 presents the same analysis excluding outliers.

Table 14: Descriptive Statistics: Networks

		Treatment status		
	Statistic	**Control**	**Tracking**	**Mixing**
Clustering coefficient (triangular)	μ	0.3284	0.3258	0.3380
	σ	0.0947	0.1081	0.0920
	median	0.3124	0.3132	0.3237
Number of communities	μ	5.0762	5.1534	5.2368
	σ	0.9225	1.0402	1.0504
	median	5.0000	5.0000	5.0000
Mean size of communities	μ	6.6866	6.8285	6.5336
	σ	1.2129	1.3460	1.1209
	median	6.6000	6.6667	6.5000
Katz-Bonacich using $\hat{\phi}_k$	μ	4.4291	5.4209	5.0193
	σ	6.5368	10.659	9.8542
	median	2.3796	2.3256	2.4789
Katz-Bonacich using $\hat{\phi}^{pool}$	μ	2.4131	2.5029	2.4187
	σ	0.6560	0.8415	0.6397
	median	2.2938	2.3167	2.3196
Katz-Bonacich using $\hat{\phi}^C$, $\hat{\phi}^T$ and $\hat{\phi}^M$	μ	2.4530	1.9091	2.7496
	σ	0.6811	0.4191	0.8487
	median	2.3275	1.8344	2.6054

Note: Katz-Bonacich statistics exclude observations between the 99th and 100th percentile of each distribution. For more detail on the construction each measure, see Miura [2012].

Figure 1: Clustering coefficients, PDF and CDF

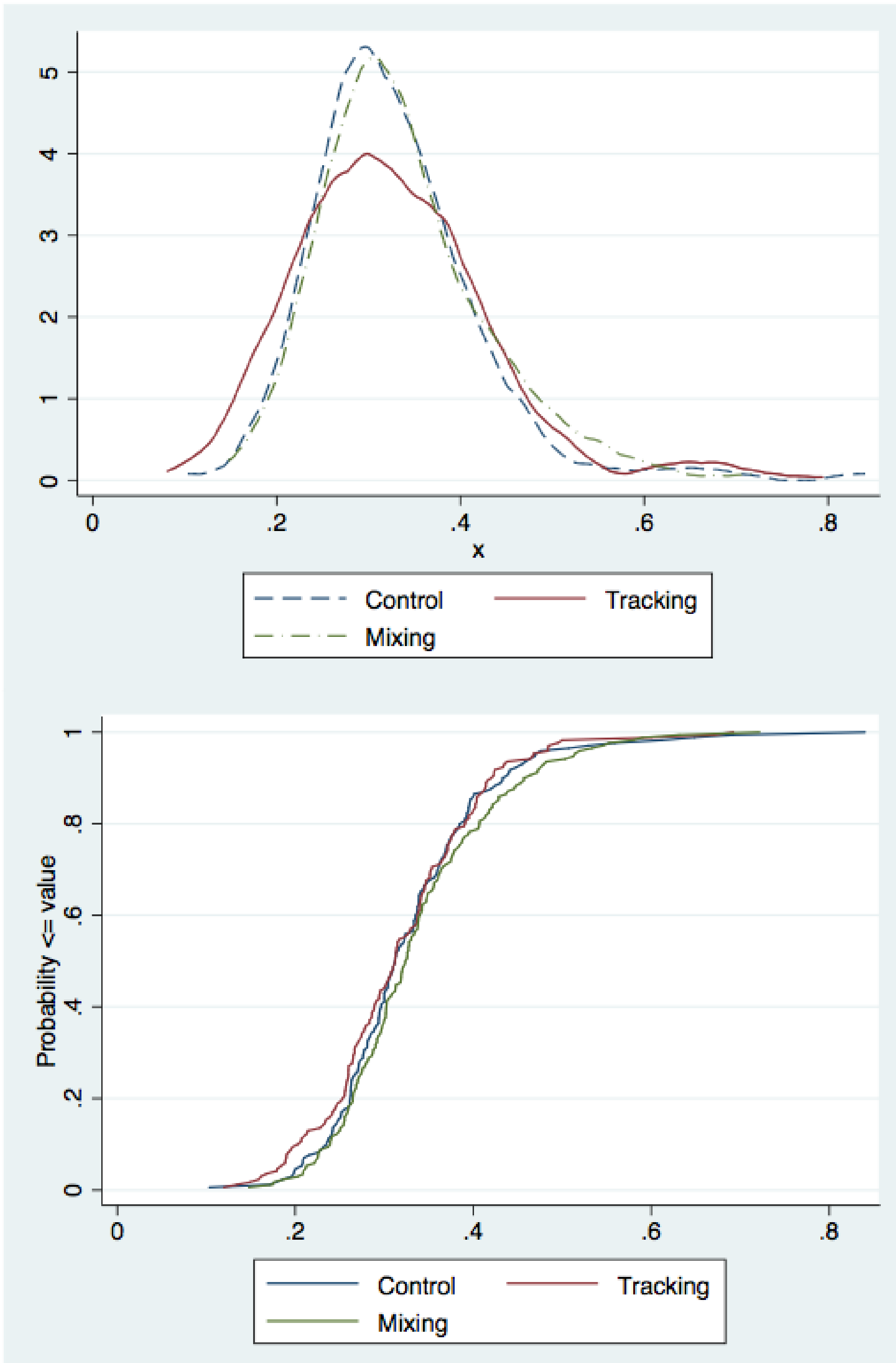


Figure 2: Number of communities and mean size, PDF

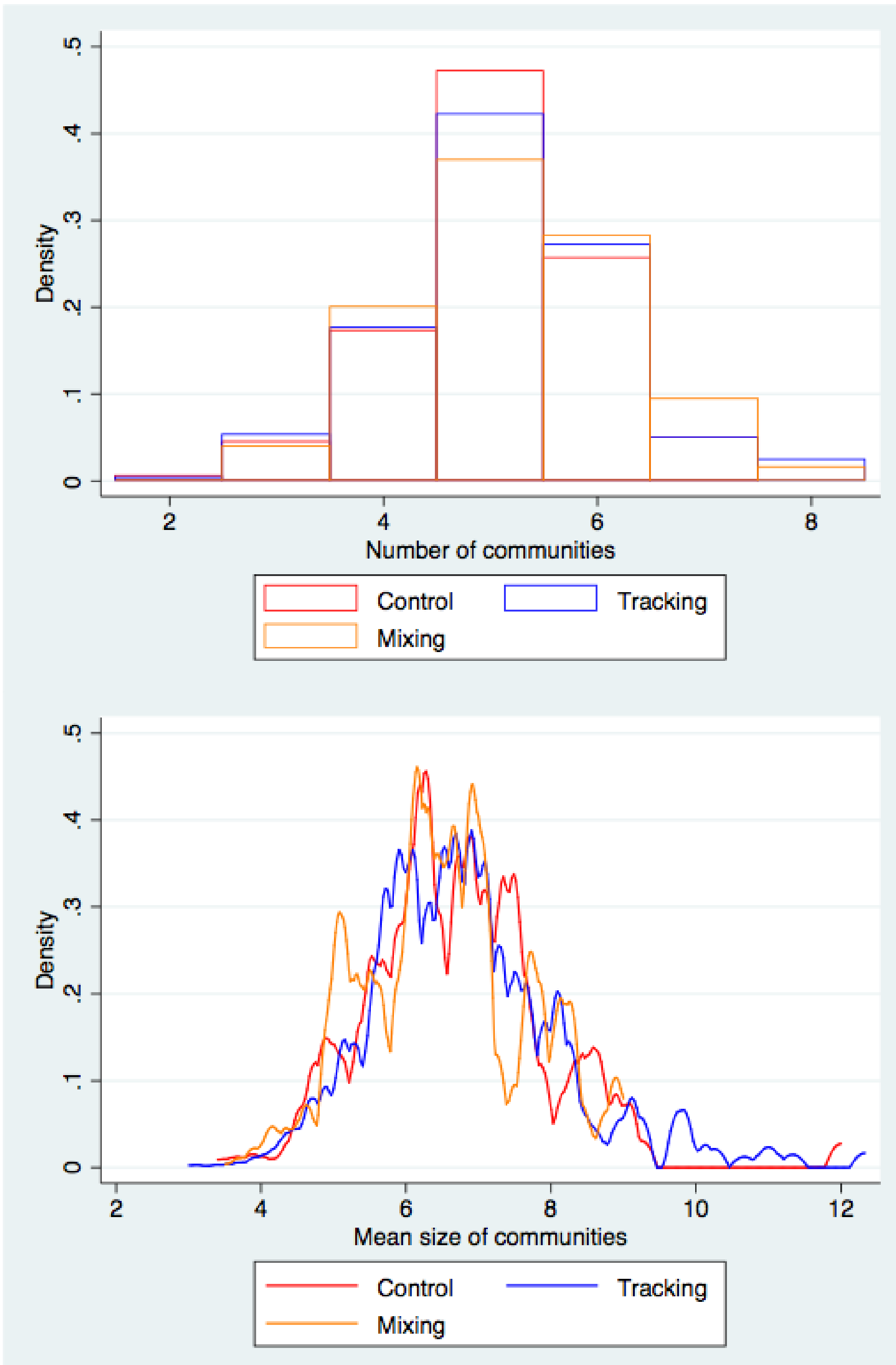
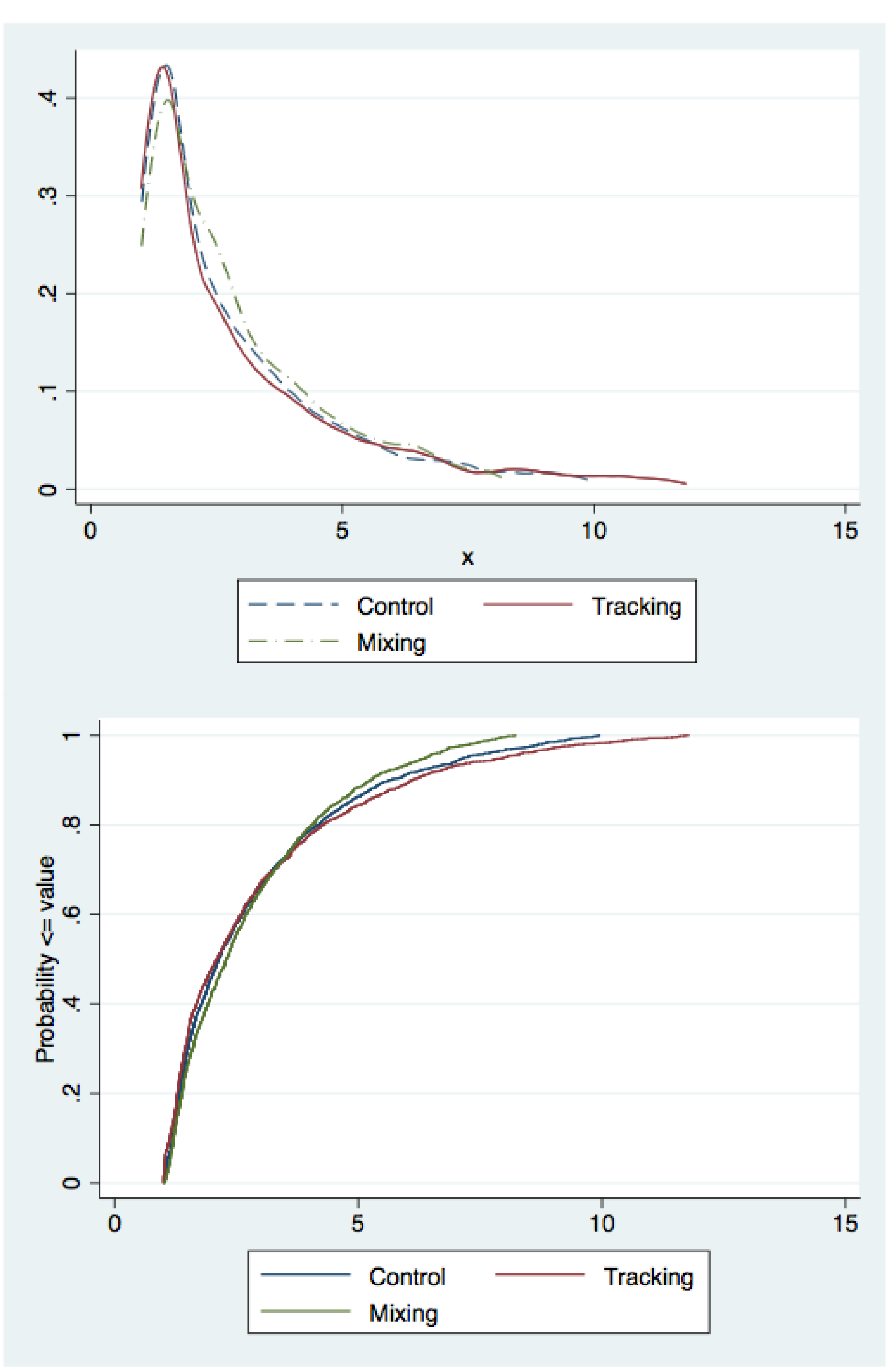
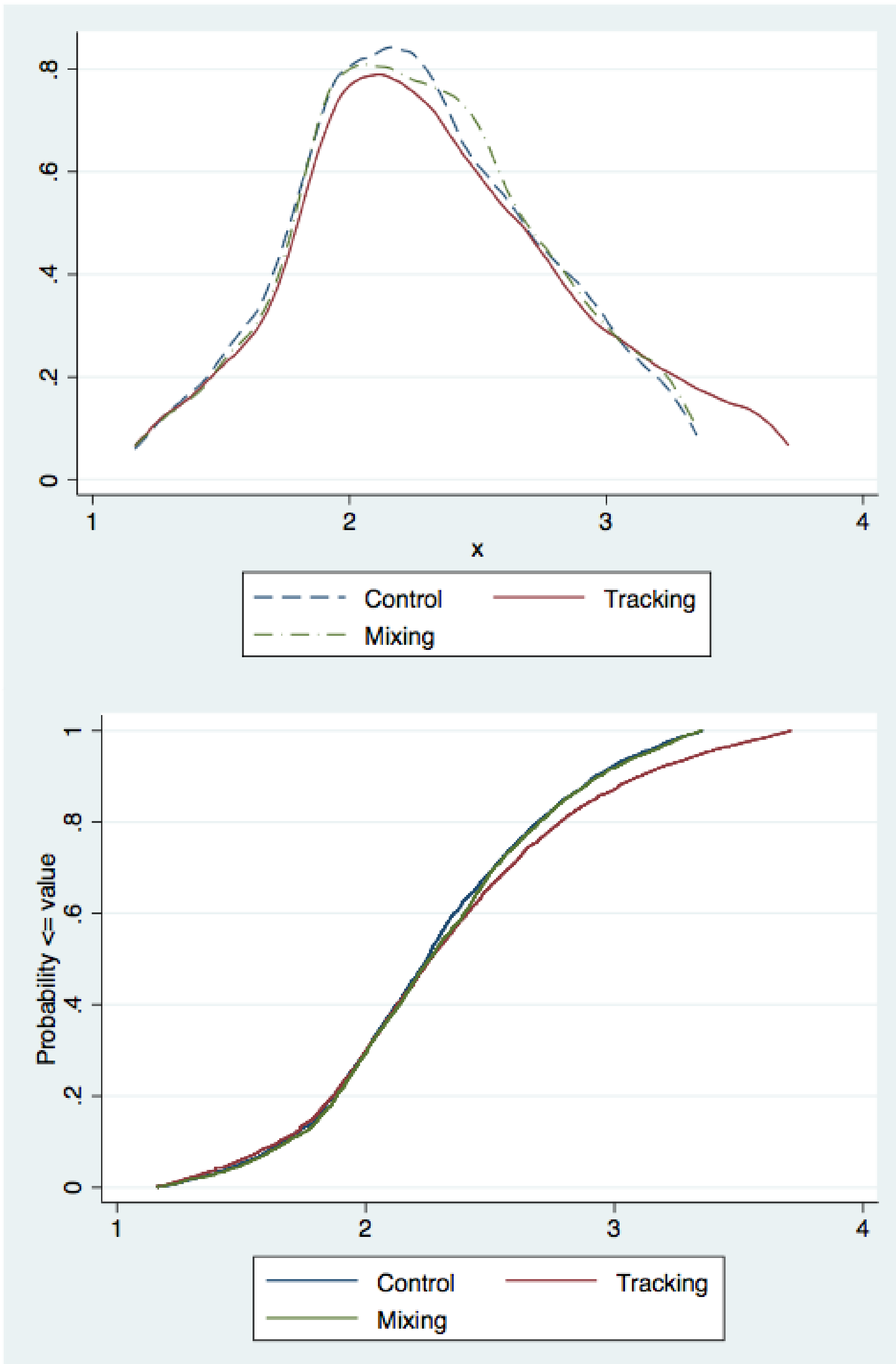


Figure 3: Katz-Bonacich with alpha calculated per classroom, PDF and CDF



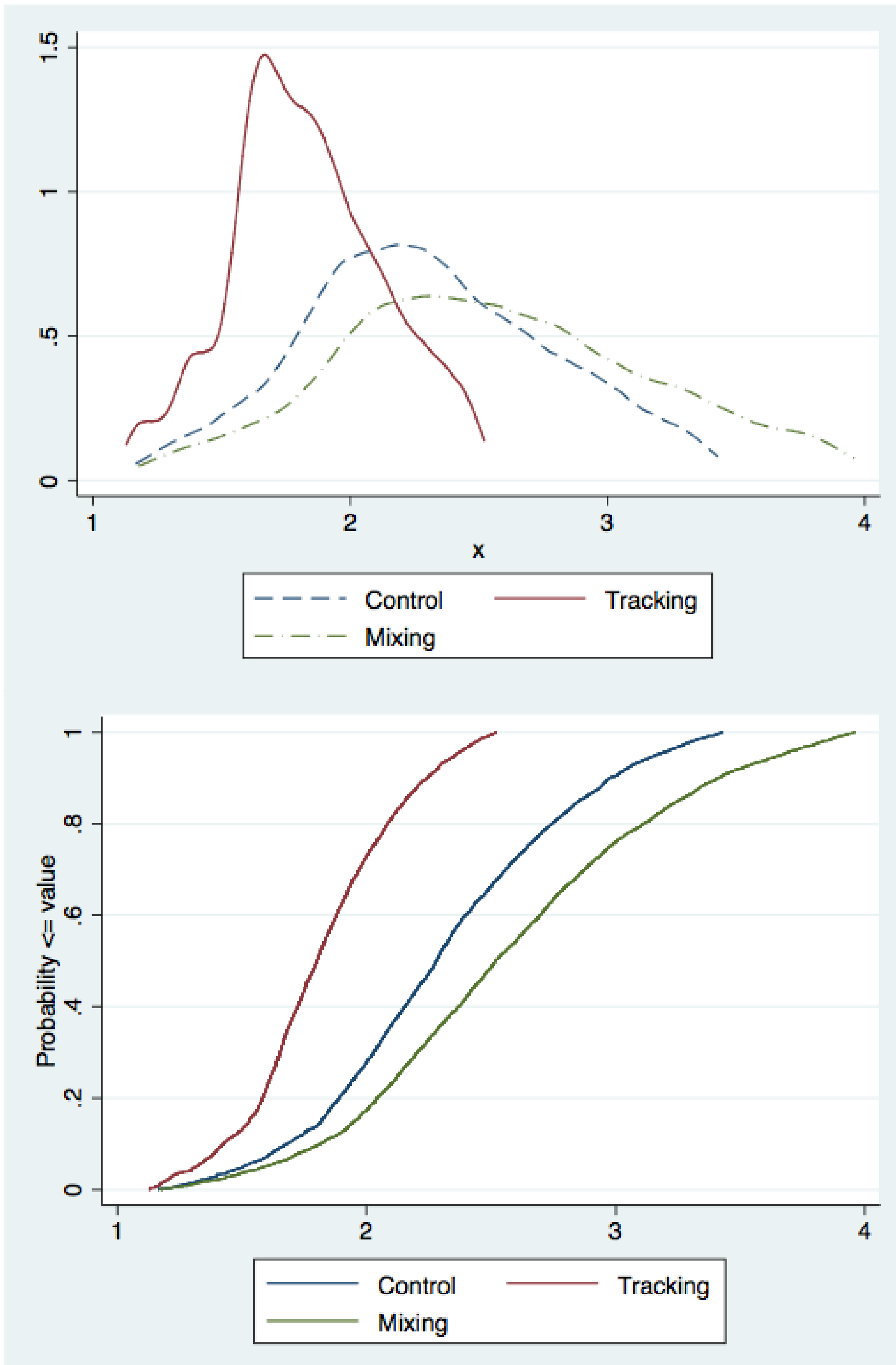
NOTE: Figures exclude observations between the 91th and 100th percentile of each distribution.

Figure 4: Katz-Bonacich with pooled alpha, PDF and CDF



Note: Figures exclude observations between the 91th and 100th percentile of each distribution.

Figure 5: Katz-Bonacich with pooled alpha by treatment status, PDF and CDF



NOTE: Figures exclude observations between the 91th and 100th percentile of each distribution.

Table 15: Mean Difference in Coefficients

Variable	Mean	β
Clustering coefficient: Control vs Tracking	0.328	-0.003
	[0.095]	[0.010]
Clustering coefficient: Control vs Mixing	0.328	0.010
	[0.095]	[0.010]
Clustering coefficient: Tracking vs Mixing	0.326	0.012
	[0.108]	[0.009]
Number of communities: Control vs Tracking	5.076	0.077
	[0.922]	[0.004]***
Number of communities: Control vs Mixing	5.076	0.161
	[0.922]	[0.005]***
Number of communities: Tracking vs Mixing	5.153	0.083
	[1.040]	[0.005]***
$\hat{\phi}_k$: Control vs Tracking	6.045	0.851
	[20.238]	[0.484]*
$\hat{\phi}_k$: Control vs Mixing	6.045	1.304
	[20.238]	[0.679]*
$\hat{\phi}_k$: Tracking vs Mixing	6.896	0.453
	[18.059]	[0.617]
$\hat{\phi}^{pool}$: Control vs Tracking	2.448	0.163
	[0.756]	[0.016]***
$\hat{\phi}^{pool}$: Control vs Mixing	2.448	0.006
	[0.756]	[0.015]
$\hat{\phi}^{pool}$: Tracking vs Mixing	2.612	-0.157
	[1.080]	[0.015]***
$\hat{\phi}^C$, $\hat{\phi}^T$ and $\hat{\phi}^M$: Control vs Tracking	2.490	-0.542
	[0.788]	[0.012]***
$\hat{\phi}^C$, $\hat{\phi}^T$ and $\hat{\phi}^M$: Control vs Mixing	2.490	0.317
	[0.788]	[0.018]***
$\hat{\phi}^C$, $\hat{\phi}^T$ and $\hat{\phi}^M$: Tracking vs Mixing	1.948	0.859
	[0.472]	[0.016]***

NOTE: * significant at 10%; ** significant at 5%; *** significant at 1%. Note: β estimated as the mean difference across treatment arms from an OLS regression. Standard error(deviation) of β(means) are in brackets.

Table 16: Mean Difference in Katz-Bonacich without outliers

Variable	Mean	β
$\hat{\phi}_k$: Control vs Tracking	4.425	1.070
	[6.531]	[0.209]***
$\hat{\phi}_k$: Control vs Mixing	4.425	0.480
	[6.531]	[0.223]**
$\hat{\phi}_k$: Tracking vs Mixing	5.496	-0.590
	[10.720]	[0.245]**
$\hat{\phi}^{pool}$: Control vs Tracking	2.448	0.163
	[0.756]	[0.016]***
$\hat{\phi}^{pool}$: Control vs Mixing	2.448	0.006
	[0.756]	[0.015]
$\hat{\phi}^{pool}$: Tracking vs Mixing	2.612	-0.157
	[1.080]	[0.015]***
$\hat{\phi}^C$, $\hat{\phi}^T$ and $\hat{\phi}^M$: Control vs Tracking	2.490	-0.542
	[0.788]	[0.012]***
$\hat{\phi}^C$, $\hat{\phi}^T$ and $\hat{\phi}^M$: Control vs Mixing	2.490	0.317
	[0.788]	[0.018]***
$\hat{\phi}^C$, $\hat{\phi}^T$ and $\hat{\phi}^M$: Tracking vs Mixing	1.948	0.859
	[0.472]	[0.016]***

NOTE: * significant at 10%; ** significant at 5%; *** significant at 1%. Note: β estimated as the mean difference across treatment arms from an OLS regression. Standard error(deviation) of β(means) are in brackets. Observations between percentile 99th and 100th are excluded from the sample.

5 Appendix

Figure 6: Network graph (random classroom): Control, friends

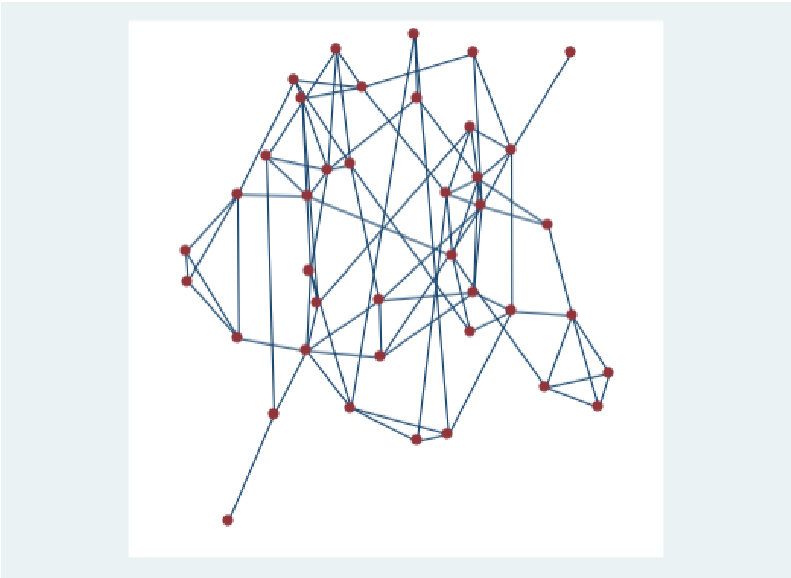


Figure 7: Network graph (random classroom): Control, best friends

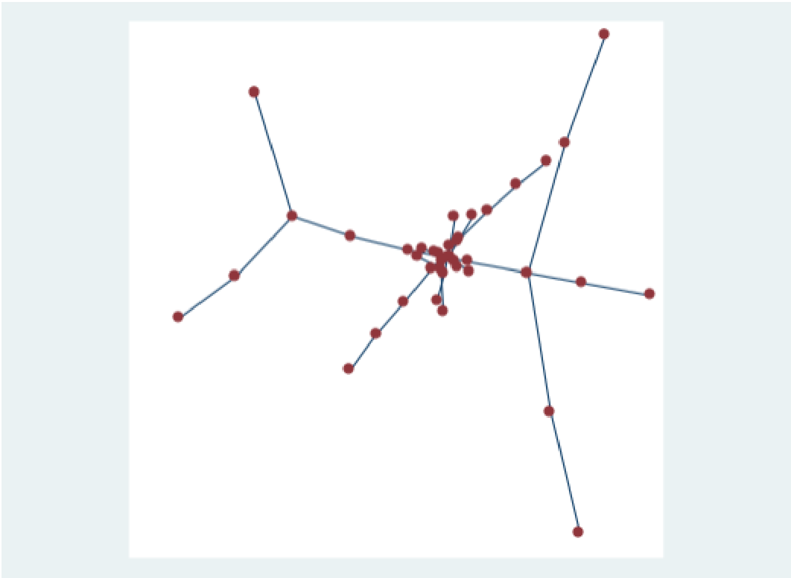


Figure 8: Network graph (random classroom): Tracking, friends

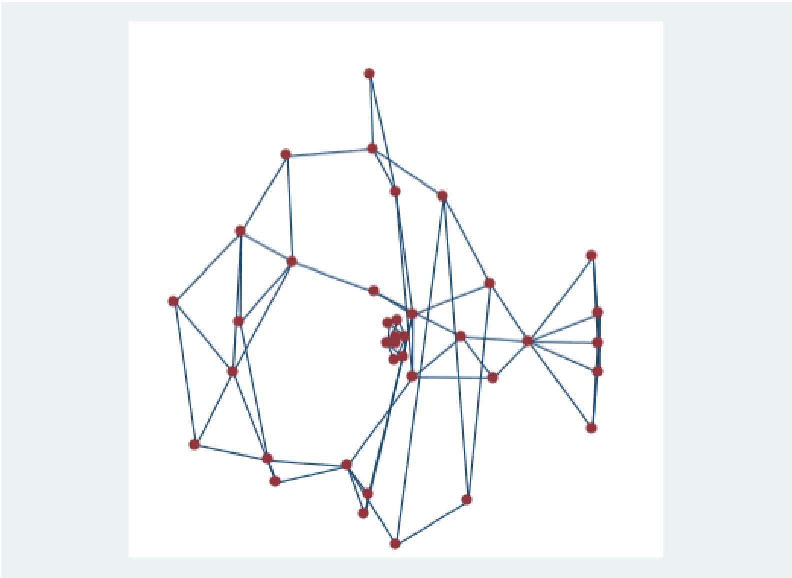


Figure 9: Network graph (random classroom): Tracking, best friends

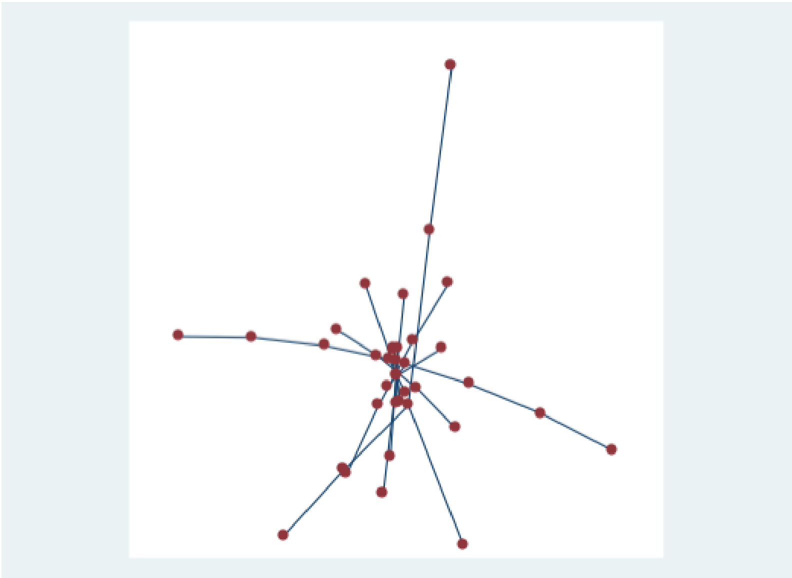


Figure 10: Network graph (random classroom): Mixing, friends

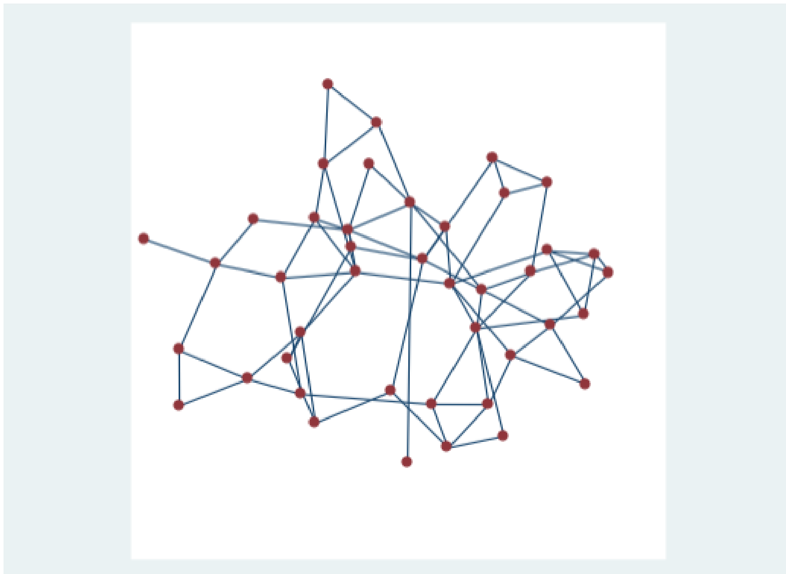


Figure 11: Network graph (random classroom): Mixing, best friends

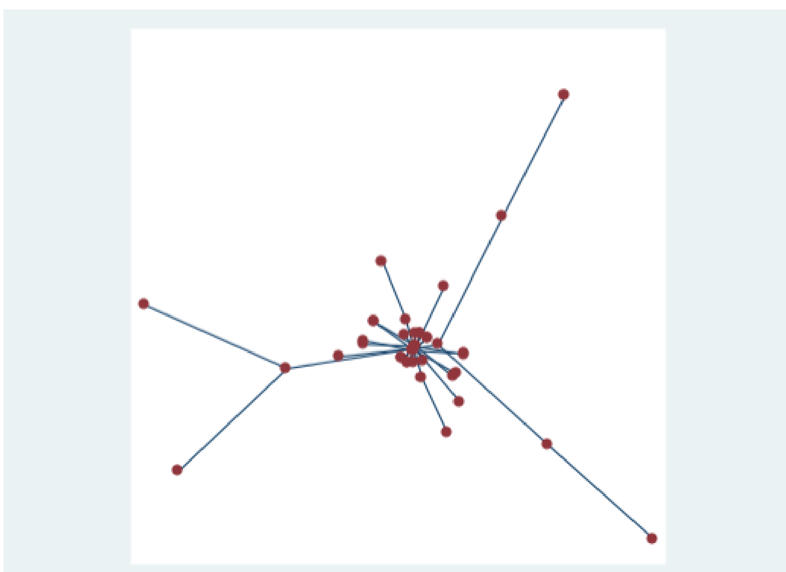


Table 17 presents a different analysis of common characteristics for factor variables. For instance, the column BF (F) shows the proportion of best friends (friends) links established among all potential links between individuals that share the same characteristic.

Table 17: Descriptive Statistics: Links and Behavioral Factors (percentage)

	Treatment status								
	Control			**Tracking**			**Mixing**		
	A	BF	F	A	BF	F	A	BF	F
Same extroversion degree (+/- 1 SD)	39	41	15	37	40	15	39	39	15
Same sociality level (+/- 1 SD)	44	40	15	43	40	15	44	41	15
Same external comparison degree (+/- 1 SD)	41	40	14	41	40	15	42	40	15
Same individualization level (+/- 1 SD)	37	40	15	38	40	15	38	40	14
Same cooperation degree (+/- 1 SD)	38	40	15	39	40	15	39	39	14
Same competitiveness degree (+/- 1 SD)	37	39	15	37	40	15	38	39	16

NOTE: A: All - indicates the proportion of all potential links sharing the correspondig characteristic. BF: Best friends - indicates the proportion of best friends established links among all potential links that share the corresponding characteristic. F: Friends - indicates the proportion of established links among all potential links that share the corresponding characteristic.

Table 18 only considers links between males, while Table 19 only considers links between females.

Table 18: Descriptive Statistics: Boys Links, Common Relevant Attitudes (percentages)

	Treatment status											
	Control				**Tracking**				**Mixing**			
	A	BF	F	NF	A	BF	F	NF	A	BF	F	NF
Attending classes	86	90	90	86	86	90	88	86	86	91	89	85
Study for tests	83	88	83	84	83	86	83	84	86	89	86	85
Practice sports	58	64	62	58	62	67	65	62	62	67	66	61
Obtain good grades	100	100	100	100	100	100	100	100	100	100	100	100
Be popular	59	65	63	59	56	62	58	56	57	62	57	57
Have a boy/girlfriend	58	61	60	58	57	62	59	56	56	61	59	56
Gather with friends	49	50	50	50	49	50	50	50	49	54	50	49
Partying	58	62	59	58	57	62	60	57	56	60	59	57
Finish school	100	100	100	100	96	100	100	93	100	100	100	100
Get a HE degree	90	97	92	91	90	100	95	90	100	100	100	100
Earn money	74	75	76	74	75	77	75	74	74	79	76	74
Get a good job	100	100	100	100	100	100	100	100	100	100	100	100
Go to the park	49	50	48	49	49	54	50	50	49	50	50	49
Go to the movies	49	54	51	50	51	56	52	51	49	56	52	50
Go to dance (during free time)	76	80	79	77	75	77	75	75	78	83	81	77
Play videogames	53	60	56	52	52	57	55	52	50	57	55	50
Go shopping	50	54	53	51	51	55	52	51	50	50	51	52
Practice sports (during free time)	62	70	67	62	62	67	67	63	62	70	68	63
Gather with friends (during free time)	50	54	52	50	49	50	52	50	49	55	52	50
Have a date (during free time)	56	62	57	57	59	62	61	59	56	64	60	57
Connect to the internet	58	67	62	59	58	62	59	58	57	62	62	57
Watch TV	53	56	55	54	53	57	55	54	52	55	53	53
Read	49	50	50	49	49	50	50	50	49	56	53	50

NOTE: A: All - indicates the proportion of all potential links sharing the correspondig characteristic. BF: Best friends - indicates the proportion of established links between best friends sharing the corresponding characteristic. F: Friends - indicates the proportion of established links between individuals sharing the corresponding characteristic. NF: Not friends - indicates the proportion of non-established links between individuals sharing the corresponding characteristic.

Table 19: Descriptive Statistics: Girls Links, Common Relevant Attitudes (percentages)

	Treatment status											
	Control				**Tracking**				**Mixing**			
	A	BF	F	NF	A	BF	F	NF	A	BF	F	NF
Attending classes	89	93	91	89	89	92	92	89	90	100	96	91
Study for tests	88	91	89	88	89	92	90	89	88	94	93	89
Practice sports	52	62	57	52	52	56	55	53	54	58	56	55
Obtain good grades	100	100	100	100	100	100	100	100	100	100	100	100
Be popular	67	71	67	67	65	67	67	65	64	71	68	65
Have a boy/girlfriend	75	80	75	74	75	78	76	74	71	75	74	72
Gather with friends	51	57	53	52	51	56	53	51	50	57	53	51
Partying	68	70	68	68	66	71	69	65	66	71	70	65
Finish school	100	100	100	100	100	100	100	100	100	100	100	100
Get a HE degree	100	100	100	100	100	100	100	100	100	100	100	100
Earn money	66	70	68	66	64	67	65	65	66	67	67	67
Get a good job	100	100	100	100	100	100	100	100	100	100	100	100
Go to the park	49	50	50	50	49	54	50	50	49	54	53	50
Go to the movies	50	50	50	51	49	54	52	50	49	50	53	50
Go to dance (during free time)	68	71	69	69	71	75	73	72	71	73	72	72
Play videogames	54	57	55	56	53	60	56	53	56	60	57	56
Go shopping	49	54	49	50	49	53	52	50	49	50	51	49
Practice sports (during free time)	57	62	63	56	56	62	60	58	57	62	60	56
Gather with friends (during free time)	49	54	52	49	49	53	51	50	49	57	51	50
Have a date (during free time)	69	70	70	70	68	75	71	69	73	77	75	74
Connect to the internet	59	63	61	59	62	67	65	61	62	67	66	62
Watch TV	51	54	54	52	51	56	54	52	53	57	56	53
Read	56	60	57	56	52	56	55	52	54	58	56	54

NOTE: A: All - indicates the proportion of all potential links sharing the correspondig characteristic. BF: Best friends - indicates the proportion of established links between best friends sharing the corresponding characteristic. F: Friends - indicates the proportion of established links between individuals sharing the corresponding characteristic. NF: Not friends - indicates the proportion of non-established links between individuals sharing the corresponding characteristic.

References

Acemoglu, D., M. A. Dahleh, I. Lobel, and A. Ozdaglar (2011). Bayesian learning in social networks. *Review of Economic Studies 78*(1), 1201–1236.

Bala, V. and S. Goyal (2000). A noncooperative model of network formation. *Econometrica 68*(5), 1181–1229.

Banerjee, A., A. Chandrasekhar, E. Duflo, and M. Jackson (2013). The difussion of microfinance. *Science 341*(26).

Banerjee, A., A. Chandrasekhar, E. Duflo, and M. Jackson (2016). Gossip: Identyfing central individuals in a social network.

Bramoullé, Y., R. Kranton, and M. D'Amours (2014). Strategic interaction and networks. *The American Economic Review 104*(3), 898?930.

Calvó-Armengol, A., E. Pattachini, and Y. Zenou (2009). Peer effects and social networks in education. *Review of Economic Studies 76*(1), 1239–1267.

Choi, S., E. Gallo, and S. Kariv (2015). Networks in the laboratory. mimeo.

Christakis, N., J. Fowler, G. W. Imbens, and K. Kalyanaraman (2010). An empirical model for strategic network formation. NBER Working Paper No. 16039.

Debreu, G. and I. Herstein (1953). Nonnegative square matrices. *Econometrica 21*, 597–607.

Jackson, M. and A. Wolinsky (1996). A strategic model of social and economic networks. *Journal of Economic Theory 71*(1), 44–74.

Kirman, A. (1983). Communication in markets: A suggested approach. *Economics letters 12*(1), 1–5.

Mele, A. (2016). A structural model of dense network formation. Econometrica, forthcoming.

Montgomery, J. (1991). Social networks and labor market outcomes. *The American Economic Review 81*(1), 1408–1418.

Mueller-Frank, M. and C. Neri (2013). Social learning in networks: Theory and experiments. Available at SSRN: https://ssrn.com/abstract=2328281.

Myerson, R. (1977). Graphs and cooperation in games math. *Operations Research 2*(1), 225–229.