A Model of Peer Effects in Schools

Ugo Bolletta[1]

[1]Aix-Marseille School of Economics

September, 2018

Abstract

When designing interventions aiming to foster peer effects in schools, knowledge on endogenous sorting of individuals into groups is key. We propose a theoretical model where agents form groups endogenously, and their outcomes are affected accordingly. Using a popular payoff structure, we show that equilibrium outcomes have a direct correspondence with the linear-in-means model, used to study empirically peer effects and we characterize the set of stable networks. The model can be fit with real data, providing the tools to i) infer the network in case we do not have this information, ii) match theoretical outcomes with real ones, to infer the motives behind network formation and iii) simulate effects of interventions that manipulate the composition of classrooms. Simulating data we show that the model can consistently reproduce the results of Carrell et al. (2013).

Keywords: endogenous network, peer effects, segregation, policy interventions, homophily

JEL Codes: D83, D85, L14, Z13

1 Introduction

Recently, some importance has been attributed to policy interventions that aim at improving outcomes for some groups of targeted individuals, by manipulating the composition of their neighbors and exploiting the channel of peer effects. Remarkable examples are schools and neighborhoods, where social influence is known as being one of the main drivers of individual outcomes. A deep understanding of this kind of interventions is relevant because they are costless and could foster beneficial relationships among individuals, minimizing detrimental encounters. The main obstacle is that once we manipulate the pool of individuals in an environment, we have no control on the endogenous responses of individuals to the said manipulation, because social relationships are endogenous and chosen by individuals.

In this paper we propose a theory that takes into account endogenous choices of peers and peer effects. We model the society through a network, where individuals are nodes and relationships are represented through links. A specific network configuration then impacts the influence that individuals mutually exercise to each other, the so called peer effects. Agents choose strategically links, whose final configuration determines behaviors, in this context interpreted as effort provision. The model provides the tools to understand the reactions of individuals to interventions. Our theory is tightly connected to the empirical model known as linear-in-means model, introduced by Manski (1993). Therefore, we show that the model could be thought as the data generating process and therefore be corroborated with a structural empirical analysis. In particular, depending on the structure of data, we could infer the network structure, predict peer effects, run hypothetical policy experiments or any combination of these.

One important application of compositional interventions was implemented in the paper by Carrell et al. (2013), among others. In that paper, authors worked with a school and had the opportunity to manipulate classrooms. During an initial period of observation where the authors measured peer effects exploiting random assignment of students into classrooms they found heterogeneous effect for *low achievement students* (from now on, LAS), which at the margin were responding positively in terms of schooling outcome to an incremental exposition of *high achievement students* (HAS). The evidence motivated an intervention where LAS and HAS where pooled together into classroom with the intent of improving outcomes for LAS. The result of that was that students exposed to the intervention performed worse than the students with similar abilities in randomized classrooms. The authors in the concluding section say:

"*We believe that endogenous responses to large policy interventions such as the ones we observe are a major obstacle to foreseeing the effects of manipulating peer groups.[..] We conclude that social processes are so rich and complex that one needs a deep understanding of their formation before one can formulate optimal policy.*"[1]

We show that the results found in Carrell et al. (2013) can be rationalized with high levels of segregation, which jeopardizes the potentially beneficial transmission of peer effects. In a numerical example, we simulate data and replicate the conditions that where implemented by the authors in a real context. An empirical analysis of the simulated data reveal that similar estimates of the effects in the real intervention are implied by a segregated network where no interactions between HAS and LAS occur.

The main theoretical results are on the characterization of stable networks. The equilibrium concept is Pairwise Nash stability, which requires all agents best responding to others' strategies, but links must be mutually accepted. The model is solved through a sequential algorithm following Watts (2001). There is an iterative process that selects a couple of agents, and lets them evaluate whether the link is mutually beneficial. In case it is not, any of the two agents can sever any number of already existing links, and evaluate again the selected link.

In a first version of the model where only schooling outcomes matter for the link formation, we show that the set of stable networks is characterized by local completeness. This means that agents form overlapping groups, and within each group all agents are friend with each other. This structure presents characteristics very well prevalent in real networks, such as clustering (a friend of my friend is also my friend) and homophily.[2]

To take into account further dimensions of heterogeneity, we propose an enriched version of the model where other characteristics not correlated with outcome, matter for the link formation process. In this way we could rationalize every observed network structure. We believe this is relevant, because we could let the theory pre-

[1] Carrell et al. (2013), pag. 881.

[2] See Jackson and Rogers (2007) for an extensive discussion of the main features of real world networks.

dicts the links in an observed sample, and then unveil the underlying mechanisms and preferences of individuals behind network formation processes. We recommend a structure of data that allows for this analysis.

From the perspective of policy recommendations, we conclude that segregation in classrooms jeopardizes peer effects and should therefore be prevented. However, a careful scheme of incentives should be considered in order to favor interactions that will lead to a significant improvement for a targeted group of students. To do that, it is key to have a clear picture of the environment and specific context and therefore it is important to collect data that unveil the underlying mechanisms behind friendships.

Among the other implications of the model, it emerges the importance of having an understanding on the motives behind link choices. Such analysis could be done, for instance, following Marmaros and Sacerdote (2006), and then find an "optimal" mix of students such that social relations in classrooms are nudged properly. This could be even more relevant in developing context where schools tend to be more resource constrained.

The rest of the paper is organized as follows. Section 2 discuss the related literature. It follows the model in section 3, where we introduce the setting and notation. Section 4 discusses a more general version of the paper, and its possible applications to experiments and empirical works. Thus the model is brought into a numerical framework in section 5 where we replicate the exercise of Carrell et al. (2013) in a simulated environment. Section 6 concludes. Proofs are in Appendix A.

2 Related Literature

As an applied theory paper, the related literature is around two different streams of work, a theoretical one and an applied one. The first one is linked to the literature on network formation and network games, while the second is the literature on peer effects.

The latest works on peer effects in the classroom were focused on the composition of classrooms. Keeping randomized classrooms as benchmark, the treatment consists of specific compositions of students' body that are expected to enhance transmission of peer effects. Nevertheless, there is no theory behind the generation of those treatments, and we argue that the lack of it is an explanation of why they have not delivered the expected results in some cases. For instance the work by Carrell et al.

(2013) identified "bimodal distribution" of students as those more likely to foster peer effects, targeted on low achievement students. The results showed that students reacted endogenously to this treatment in terms of group composition, and the two group of students (low achievement and high achievement) remained segregated. In this paper we propose a theoretical framework that could be exploited in that direction, and therefore could guide further experiments having clearcut predictions as a benchmark. Other examples in this literature are Li et al. (2014) which to favor interaction among high achievement and low achievement students through monetary transfers. This is interpretable in the framework of this paper as an increase on the benefits deriving from the interactions.

Other work on peer effects in the classroom is provided by De Giorgi and Pellizzari (2014). The purpose of this work is to detect the determinants of peer effects. The authors find that social interactions that drive peer effects are formed because they provides mutually insurance, and so it seems that the channel of risk sharing is moving the subjects of their study. Students there interact to avoid the risk of bad shocks, as they are mostly individual, and they tend to vanish at the group level. This may be relevant in the following analysis as we assume that multiple factors could drive social interactions. Our model could encompass this result in two possible ways: if the risk is mostly associated to schooling outcome they will affect the cost of interactions, or otherwise they will affect the benefit. In the latter case we talk about role models. In the numerical example we will assume specifically that link formation is mainly driven by better quality agents. That will lead to homophily in grades, which is documented in Barnes et al. (2014).

Although schools are an important matter, peer effects are known to be relevant in other contexts. Kling et al. (2007), Ludwig et al. (2012) and Chetty et al. (2016) focus on neighborhood effects analyzing the effects of Saia (2018) has studied the context of parliament. Exploiting random assignment of seats in the Icelandic parliament, the author shows that being seated next to members of the opposing party increases the likelihood that a parliamentary votes against its own party proposals.

Moving to the theory, a key reference is given by Calvó-Armengol et al. (2009). This paper studies a game played on the network, or in other words where the outcomes depend on the network structure, focusing on the case of peer effects in schools. They find that the equilibrium outcomes are proportional to the Bonacich centrality. Thus agents with higher centrality exhibits higher schooling outcomes and peer effects. This is because in their model the game is represented by pure complementarities. Therefore more peer effects imply higher outcomes. While our payoff structure is inspired by this paper, our model embeds idiosyncratic reference points which are interpreted as individual optimal levels of effort under isolation. Therefore both positive and negative influence is considered. The major contribution is due to the strategic network formation component that we introduce here. In that paper, in fact, the network is assumed to be exogenous. We believe this is an important step forward in this literature.

However, few attempts have been made to endogenize network formation in a game with peer effects. Remarkable examples closely related to this work are Boucher (2015) and Boucher (2016). Let us discuss these two separately. Boucher (2015) proposes a model that aims to combine network formation and peer effects. The agents are assumed to have preferences for homophily. Payoff are characterized by a distance function that agents want to minimize. The model produces a unique equilibrium network and the result is then used to estimate the parameters that measure preferences for homophily on observed dimensions, such as race, age, gender and socio-economic background. The main trade off in the model is between preferences for homophily and capacity constraints on links. We improve on this work by proposing a more general model inspired by the empirical literature, where homophily emerges as a result of the interplay between choice of peers and peer effects. This allows to perform a similar analysis on the preferences for homophily, while not trading-off for a more general model with peer effects.

Boucher (2016) builds a similar model to the one proposed here. The author studies a framework where agents simultaneously choose the network and peer effects are determined accordingly. While a full characterization is not achieved here, because the equilibrium is solved through the potential function which is not concave in this case, the results can be ranked in terms of overall variance of behaviors. Under the assumption that extreme behaviors are detrimental, the author suggests that denser networks, i.e. that exhibit higher levels of conformism, are desirable. These implications are tested on the data. We contribute proposing a framework that captures the high degree of flexibility in terms of accounted heterogeneity, but we propose a routine that allows for the full resolution of the model. Therefore, I show how the model can be brought to the data and be used for policy experiments.

In terms of methodology this work is also related to some papers that model separately network formation and peer effects. Among others, Badev (2013) and Goldsmith-Pinkham and Imbens (2013). These works takes into account peer effects and endogenous network formation. However the model here differ substantially. With respect to the former we let the outcome of the game being defined on a continuous set, while there, the outcome and the modeling strategy rely strongly on the fact that actions are binary. Thus that works mainly focuses on smoking behavior or similar binary strategies. Peer effects in the classroom are by definition an intensive margin measure, and therefore we believe it is more appropriate to model actions in a continuous way.

More broadly, this paper is related to the literature on strategic network formation. In the first place the paper by Jackson and Wolinsky (1996) proposed a framework to work with, and introduced the notion of pairwise stability. From that paper we took inspiration from the so called *connections model*, which elicits strategic behaviors through an extremely reduced form of benefit and costs deriving from interactions. Here we go beyond this reduced form and we let the costs deriving from interactions be determined by an underlying coordination game. The structure of the problem characterizes a network game. The usual practice in the literature was to study games in this category keeping the network exogenous. For instance Ballester et al. (2006) studies a game of strategic complements, while Bramoullé and Kranton (2007) and Bramoullé et al. (2014) focus on strategic substitutes. Making the network endogenous is instead relatively new, and to the best of our knowledge few works attempt to do so. Among these there is the paper by Galeotti and Goyal (2010) which is the first contribution into the literature in this direction. More recently there is Kinateder and Merlino (2014). Both these works analyze public good games, and therefore games with strategic substitutes. Finally Bolletta and Pin (2018) consider a game that exhibits the same equilibrium best-response in a context of evolution of opinions, where agents form the network strategically. Results show that polarization of opinions can arise and persist under disconnected networks, but also during the transition to consensus, where opinions tend to condensate around two poles, but will slowly converge over time.

3 The model

In this section we present the main ingredients of our theoretical framework. After setting the notation we will define the preferences and discuss their properties.

3.1 Notation

There are N agents characterized by a type $\theta \in \mathbb{R}^+$. Agent's objective function is to maximize their utility $U_i = u(y_i, \theta_j, \theta_i)$. Utility depends on their effort, the characteristics of their peers, and their own characteristics. Peers are represented through links of a network. It is therefore assumed that if two agents are not directly linked, no direct peer influence take place between such two agents. The network G is a collection of all the links. The reference group of an agent i in a network configuration g is denoted by $g_i = \{j|g_{ij} = 1\}$. The number of peers, i.e. the degree of an agent i, is denoted by $|g_i| = d_i(g)$. Finally we denote with $g_i(\emptyset)$ the situation in which the set of links for agent i is empty. Similarly, $g(\emptyset)$ represent the case where the whole network is empty.

3.2 Preferences

We imagine that agents are endowed with a set of characteristics that we bundle together in the type θ_i. Through this term we want to indicate the potential schooling outcome of an agent if in isolation, or in other words, cleaned from any peer effects. In reduced form this would be something depending on observable and non-observable characteristics, as well as contextual effects that interact globally with every student. If isolated their schooling potential would depend only on their exogenous characteristics. We do not represent the student's production function, but it is possible to extend this framework and obtain the types as a result of a more complex and detailed model. A very simple way of representing types is the following:

$$\theta_i = \sum_{m}^{M} \beta_m x_i^m \tag{1}$$

Equation 1 An econometrician may want to distinguish between observable and non-observable characteristics, but in this theoretical framework we do not do that. The type θ here functions as a reference point. Effort is costly on one hand, but there is a complementarity term between effort and type. The higher the type, the higher the utility for any level of effort. This results in higher levels of exerted effort for higher types, representing the idea that high types are more capable and therefore work more than lower types.

$$U_i(g_i(\emptyset)) = -\frac{1}{2}y_i^2 + y_i\theta_i \tag{2}$$

Agents' production fully depends on their characteristics. The optimal effort provision is indeed

$$y_i^*(g_i(\emptyset)) = \theta_i. \tag{3}$$

Here there is an important assumption of the model. Agents behaviors are interpreted as effort. We need therefore a further assumption that effort maps into grades and performances homogeneously across individuals. The source of heterogeneity is introduced in the model through exogenously given types and the network that will be determined endogenously. Agents derive utility directly from exerting effort, and they will differ in terms of their optimal levels which are shaped by their type. While on one hand this may seem extreme, because students' utility may depend on several factors other than grades at school, we are just saying that agents are always better off from having good grades, everything else equal. In fact, we will extend the model to more general preferences in section 4. Until then, we will focus on these preferences as the preferences for agents under no social interactions. Therefore, as long as peer effects do not exist, we will assume this simple structure to describe agents' behaviors.

Let us now add up a component that captures the effects of social interactions.

$$U_i(y_i, y_{-i}, \theta_i, g) = -\frac{1}{2}y_i^2 + \alpha\frac{1}{d_i}\sum_{j\in N(i)} y_iy_j + (1-\alpha)y_i\theta_i \tag{4}$$

If an agent degree is positive in a network g, the production of effort depends both on own characteristics and others' levels of effort. The term $\frac{1}{d_i}$ weights each social interaction. This implies that the more the connections an individual have, the less the influence that each agent exerts on her/him. Moreover, it is assumed that influence is homogeneous across connections. Thus, one could have more friends, and there is no friend that is more prominent than others.[3] Peer effects are represented through a complementarity term. Therefore, given any idiosyncratic focal point defined through θ_i, total effort for an agent i increases with the outcome of other students in her group. This is well evident considering optimal levels of effort provision:

$$y_i^*(g) = \alpha \bar{y}_i + (1-\alpha)\theta_i \tag{5}$$

where $\bar{y}_i$ denotes the average of characteristics among the agents in the reference group of i, $\bar{y}_i = \frac{1}{d_i(g)} \sum_{j \in g_i} y_j$. Then if we compare both the utility and the optimal level of effort we have the following relationships. If an agent is connected with agents on average better than him, he would get higher utility and exert a higher level of effort. The opposite if connected with agents on average characterized by lower type.

However, before formalizing this result, we report our first result about the equilibrium in the sub-game of effort choice in presence of social interactions (the second stage), assuming that the network is exogenous. In particular we want that no indeterminacy arises here, so that we could move on the network formation stage, taking for granted that behaviors will adjust to the stable network accordingly. The proof of this result, as of all the other results are in Appendix A.

Lemma **1.** *For all $0 \leq \alpha < 1$ there exist a unique Nash equilibrium of the coordination game with $y_i^* \in [\underline{\theta}, \tilde{\theta}]$ for all $i \in N$.*

This result basically states, that for every possible exogenous network, the vector of equilibrium behaviors is unique. There are several ways to prove this result. An alternative way would have been through a matrix representation of the system of best responses. Once doing so we would have had the following result.

$$\mathbf{y} = (I - \alpha \mathbf{W})^{-1} (1-\alpha) \boldsymbol{\theta} \tag{6}$$

Equation 6 is of great importance. In fact we first notice that for the equation to exhibit unique solutions we need to ensure that the matrix $(I - \alpha \mathbf{W})$ is invertible. Formally, we need that the spectral radius of the matrix $\alpha \mathbf{W}$ to be lower than 1. Given that the spectral radius of any row-stochastic matrix is 1, the condition holds

[3] This assumption could be relaxed. This would complicate dramatically the derivation of analytical results, but in a numerical simulation it is possible to account for heterogeneity in the importance of each connection for each individual. However it is crucial that the sum of all weights is equal to 1.

for any $\alpha < 1$. Thus, adding the assumption of $\mathbf{W}$ being row-stochastic, which follows from the payoff structure, we proved our statement. Another remark is that the term $(I - \alpha\mathbf{W})^{-1}$, represents the Bonacich centrality vector. An interesting additional analysis would concern the behavior of equation 6 as α tends to one in the limit. It is possible indeed to show that in such a case, the term $(I - \alpha\mathbf{W})^{-1}$ captures the eigenvector centrality. Besides this model being extremely nice in terms of properties and informativeness about how the network structure matters, it has also a strong interpretative power. The next remark shows the connections with an empirical model to study peer effects.

Remark **1.** *The best response given by equation 5 corresponds to the linear-in-means model as in Manski (1993), with a restriction on the parameters given by* $\beta_1 = \frac{1-\alpha}{1-\alpha^2}$ *and* $\beta_2 = \frac{\alpha}{1+\alpha}$*, if and only if the underlying network is* k*-regular.*

It is worth considering that the in general the network structure is not k-regular. The most natural example of such configuration is the complete network. From this we learn that if the groups are of different sizes, running a simple linear-in-means model, misses potentially relevant heterogeneous effects in the population, and the coefficients are not identified. This is a similar result to the one found in Bramoullé et al. (2009), in Proposition 2. However we show this result through our system of best replies. The other condition stated in Remark 1 is about a restriction on the parameters to be estimated. We keep this restriction throughout the model to derive cleaner results. However, if one would follow a more structural approach, bringing the model directly to the data, she could leave the parameters unrestricted, obtaining the same results of this paper, with some restrictions still arising to ensure that the model is well-behaved.

An important consideration to be done is about interpreting the model that has been vastly used to measure peer effects as a best response. Given the complexity of strategies in this case, after introducing the strategic choice of peers, this interpretation loses appealing. The way we push our interpretation here is to take the linear-in-means model as a behavioral rule, and the true best response is associated to the choice of links, anticipating the changes that would derive from any addition or deletion of a link.

Now let us look at player's incentives to form links. Here we show that any link toward better students than themselves is profitable, while it is costly sending links towards worse students.

Lemma **2.** *For all agents i, $U_i(y, g + ik) > U_i(y, g + ij)$ for all $y_k > y_j$. Moreover, $y_i^*(g + ik) > y_i^*(g + ij)$, for each $\alpha \in [0, 1)$.*

This raises two considerations, depending if we are in a context that is described by a directed network or an undirected network. First let us consider the implications of the choice of one of these two. Directed network would make sense in a context where the indirect benefits are relevant. In schools this would translate into a framework where students are influenced by others through indirect interactions, that is by observation or competition, or even simple externalities created by one's behavior. While this is a mechanic that is likely to exist, we claim that its relevance is marginal with respect to the influence that students have on each other through mutual interactions and direct exchange. Nevertheless, under directed network because we would observe a specific network structure arise under stability, which is called *nested-split graph*. A nested split graph is a configuration where, given an ordering of players, the "best node" receives $N - 1$ links, the second best $N - 2$ and so on and so forth. No links from a "better" node is sent to a "worse" node. In other words, it is a hierarchical structure such that the neighborhood of an agent with low centrality is a subset of the neighborhood of another agent with higher centrality (neighborhoods are nested). In the literature there exist several definitions of *nested split graphs*. One possible definition states that if ij have a directed link in the configuration G, and there is an agent k that has an in-degree d_k^{in} greater than agent j, that implies that i is linked with k, too. Formally we have the following

Definition **1.** *A graph is called a nested split graph if*

$$ij \in G \quad \text{and} \quad d_k^{in} \geq d_j^{in} \Rightarrow ik \in G$$

Here we provide the formal statement about the result on nested split graph in case of directed network.

Corollary **3.** *If links are directed, all stable networks are nested-split graphs.*

The proof follows directly from Lemma 2, and is therefore omitted. Thus all links sent towards "better" students increase strictly both utility and the equilibrium outcome. It is trivial then to check that the nested split graph is the only stable network in case of directed configuration. For the reasons mentioned above, we do not believe that this is the most interesting case an therefore we focus on the case of undirected network, where links must be reciprocated to be successfully formed.

Given our choice to work with an undirected network then, the first observation is that under this setting the stable network can only be the empty network. Under the simple rule for which no link downward is accepted, while all links are sent upwards, we see immediately that it does not exist a link that is profitable for two players.

This case is however totally uninteresting for the purpose of this paper, so we move to a more reasonable scenario where there are benefit attached to social relations that are not related to outcome. On one hand this could be motivated by the fact that we observe individuals having social intercourses all the time, and therefore there must be some intrinsic value to a social relation. Moreover in a school context the number of friends is usually a proxy of popularity among students. This motivates us enough to move to the next section where we will be looking to a context where agents can choose their own links.

Provided that, we can now move to the endogenous network. We will add an initial stage of the game, where players will form the best set of links, taking into account the effects on behavior and other's choices, too.

3.3 Stable networks with peer effects

In order to justify the presence of links we have to introduce a benefit deriving from links. Such benefit may be interpreted as a value of linking to someone, independently of their characteristics. This is a simplification, but we will extend the setting to more general preferences in the section 4. Thus let us consider the following preferences.

$$U_i = \sum_{j \in N(i)} \delta - \frac{1}{2} y_i^2 + \alpha \frac{1}{d_i} \sum_{j \in N(i)} y_i y_j + (1 - \alpha) y_i \theta_i \tag{7}$$

The set of strategies for a player i is given by $S_i = Y \times G_i$ where $G_i = \{0, 1\}^{N-1}$ and $Y \in \mathbb{R}^+$. The set of all players strategies is given by $S_1 \times ... \times S_N$ and a strategy profile $s = (x, g) \in S$ defines the vector of outcomes $y = (y_1, ..., y_N)$ and players linking strategies $g = (g_1, ..., g_N)$.

Definition **2.** ***TIMING***: *The timing of the game is:*

- *Agents form the network,*

- *Outcomes are determined as the solution of the system of best responses.*

In order to determine the equilibrium concept here, we have to define a deviation for a player. First of all we want links to be mutually accepted. Severing a link on the other hand must require a single player deviation. Therefore we want the equilibrium to be pairwise stable. Here it follows the definition of pairwise stability initially introduced by Jackson and Wolinsky (1996), that we report formally here to let the paper be self-contained.

- $\forall ij \in g, \; U_i(y, g) \geq U_i(y, g - ij) \quad \text{and} \quad U_j(y, g) \geq U_j(y, g - ij)$
- $\forall ij \notin g, \; \text{if } U_i(y, g) < U_i(y, g + ij) \quad \text{then} \quad U_j(y, g) > U_j(y, g + ij)$

A strategy profile $s^* = (y*, g*)$ is a Nash equilibrium if and only if for all $s_i \in S_i$ and all $i \in N$

$$U_i(s^*) \geq U_i(s_i, s^*_{-i})$$

Combining pairwise stability (PS) and Nash equilibrium, we are in the equilibrium concept of Pairwise Nash stability (PNS). Therefore a deviation here consists into severing any number of links, which is a unilateral choice, and forming any number of links under mutual acceptance.

As we know, equilibria may first of all fail to exist, and if there exist an equilibrium, it is not unique in general. While we cannot show analytically existence and uniqueness we will provide a characterization of all stable networks. Before going into the result we find useful to provide now an example that may clarify the main mechanisms of the model.

Example 1. Lawrence and Kent[4] are two freshmen seeking to join a fraternity. Initially they meet Greg, which comes from a prestigious house family, and has always excelled at school. θ_G stands here for Greg's type, and we assume it is $\theta_G = 4$. Since $\theta_L =$ and $\theta_K =$, respectively, the utility deriving from link would be they meet John.

There are $n = 4$ agents, Lawrence, Kent, Greg and John. Their types are respectively $\theta_L = 2.5$, $\theta_K = 2.5$, $\theta_G = 4$ and $\theta_J = 1$. The parameter $\alpha = 0.5$. Greg considers forming a link with Lawrence and Kent, but realizes that is not profitable. Call the network where Greg does not form links with neither Lawrence nor Kent g, while the network with links g'. In fact the reservation utility for Greg is $U_G(g) = 8$, while if

[4]This example is freely inspired by the movie National Lampoon's Animal House (1978).

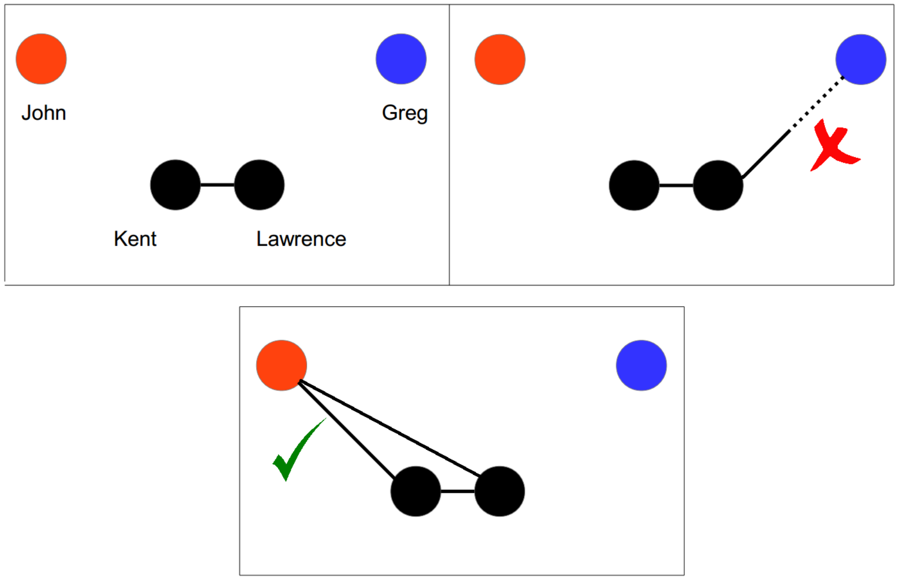

Figure 1: The figures summarizes the example. We have a scenario where for both Lawrence and Kent it would be profitable to form a link with Greg, but he refuses. However for them it is still profitable to link with John, which leads to the stable network represented in the bottom panel of the figure. All details about parameters are described in the text.

linked it would be $U_G(g') = 5.28125$, and his grade would be $y^*_G(g') = 3.25$. Thus, the marginal utility for Greg of adding this link is $mu_G(g', g) = 5.28125 - 8 = -2.71875$. Now, consider $\delta = 1.25$. In such a case adding two links would not be enough to compensate for the loss in utility deriving from the reduction in Greg's grades.

After meeting Greg,[5] Lawrence and Kent meet John. Now they consider whether to form the link. Call g the network where Kent and Lawrence are linked together, and g' the network where they add the link with John. Grades would go down for both of them since $y^*_L = \alpha\bar{\theta}_L + (1-\alpha)\theta_L = \alpha\frac{1}{2}(\theta_K + \theta_J) + (1-\alpha)\theta_L = 2.125$, and similarly $y^*_K = 2.125$. Utility for either Lawrence or Kent are $U_{L,K}(g') = 2.5859375$, while $U_{L,K}(g) = 3.125$. Including the benefit $\delta = 1.25$ we get respectively $U_{L,K}(g') = 5.0859375$ and $U_{L,K}(g) = 4.375$. Thus we see how with such benefit the stable network is the one where Lawrence and Kent trade-off their grades for friendship with John, since the more desirable link with Greg could not be sustained, because for Greg it was not profitable to add the links.

[5]The sequentiality here is only for flavor reasons, while it has no relation with the equilibrium conditions.

Thus all links sent towards "better" students increase strictly both utility and the equilibrium outcome. This result helps us also in the characterization of equilibria. In order to do so, we have to assume an algorithm that let us both guide towards a solution and will help in the proof. We therefore provide the following definition.

Definition **3.** ***NETWORK FORMATION PROCESS***: *Players meet over time $T = 1, 2, ..., t, ...$ At each period a couple ij is selected to decide whether to form a link or sever it, if already existing. Players selected this way can simultaneously sever any existing link with all $k \in N(i)$ and all $h \in N(j)$.*

Considering the network formation process above, the network reaches a stable state when no couple of agents want to deviate anymore by adding a link, and no agent wants to sever one or more links. The intuition of what follows is that all the links that are accepted by two agent are formed. Such an announcement strategy will be also PS if no other mutually beneficial link remains, and no agent wants to unilaterally sever any link formed this way. Given the previous remark we know that all links to better quality students are desired by the agents. At the same time, while links to worse quality students are detrimental for utility, they may be counterbalanced by the positive utility deriving from the benefit of having an additional link. A key part in the definition of the network formation process is that agents can revise the set of links along the process. This is relevant because otherwise the algorithm may pick some situations where it is no longer optimal to keep a link, but there is no way to get rid of it.

The algorithm was first introduced by Watts (2001), and it represents a dynamic context. Although we do not discuss dynamics in this paper, we could interpret this network formation stage as an initial period of trial and error. Initially students do not know each other, so they will try new friendships. Over time, they will keep only the connections that matter, and will keep those for a long time. Before proceeding with the result we provide the formal definition of the configuration that turns out to characterize the stable networks here. This definition can be found in Dutta and Jackson (2013).

Definition **4.** ***LOCALLY COMPLETE NETWORK***: *Call $i \leftrightarrow j := \{h \in N | i \leq h \leq j\}$ the set of agents between i, j in the well ordered set. Moreover $g^{i \leftrightarrow j}$, which is the complete network on $i \leftrightarrow j$ is called to be a clique in g if $g^{i \leftrightarrow j} \subset g$.*

Thus a network is called locally complete when for each $i < j | ij \in g$ implies $i \leftrightarrow j$ is a clique in g.

Informally, local completeness means that if we order agents on the line according to their type, take any couple of agents and then all agents that are situated between them will form a complete network. This definition allows us to characterize the set of stable networks.

PROPOSITION **4.** *Every stable state is a locally complete network.*

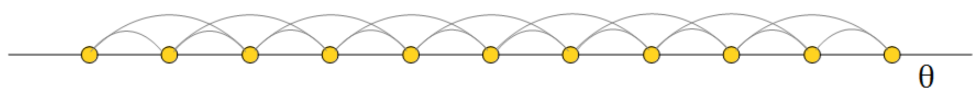

Figure 2: A simple and specific example of a locally complete network.

We cannot rule out the hypothesis that the network will not converge to a stable state. We know from Jackson and Watts (2002), that either we reach a stable state, or we are in a cycle.[6] One of the key remarks from the characterization of stable networks is that *homophily* emerges. Only links with agents similar in characteristics are formed. This is a consequence of preferences on one hand, but also of the fact that agents anticipates their equilibrium behavior after that the network is successfully formed. It is important to stress this feature because it is well documented in empirical works that homophily is a key feature in real world networks. Another remark, is that both peer effects and homophily push towards similarity in behavior, although peer effects would be stronger in absence of homophily, since interactions happen among more heterogeneous agents.

The characterization provided in Proposition 4 is quite strong. The set of possible networks in absence of structure is incredibly large, and the characterization reduces it dramatically. This simple version of the model has therefore a nice structure that allows to derive neat analytical results. However, to account for the complexity of real world scenarios, we extend the model in order to accommodate for further dimensions of heterogeneity. We discuss this extension in the next section.

[6]See Jackson and Watts (2002), Lemma 1, pag. 273.

4 Rationalization of network structures

So far, we have proposed a simplified version of the model, where basically agents make the choice of links only taking into consideration the characteristics relevant for their effort production. Interactions are assumed to be beneficial per se, since the benefit is exogenous, and completely unrelated to any characteristic. As we have seen, this gives rise to peculiar network structures, characterized by strong homophily on grades. While this is somehow observed in reality, it is also documented that structures as regular as locally complete graphs are hardly encountered. However, the model produces relevant features observed in real social networks. The main empirical regularities that our model can reproduce can be listed as follows:[7]

1. Low diameter, as the maximum distance between any two agents. If segregation do not arise the network is quite dense and the diameter is low.

2. High clustering, as the number of close triads over the sum of all possible close triads. Local completeness implies high levels of clustering.

3. Homophily.

Although we presented a strong characterization that reduces dramatically the number of stable networks, the regularity demanded by local completeness is often rejected by the data. Through this section we aim at providing the tools to extend the set-up proposed so far, in order to gain flexibility and potentially explain every possible network structure.

To do so we focus on characteristics unrelated to grades. An econometrician observes a number of variables, and through a simple analysis it is possible to determine which characteristics correlate with effort, and which do not.

Call the set of observables of an individual i that do not contribute to effort production $\tau_i = \{\tau_i^1, ..., \tau_i^M\}$. One possible way of incorporating these extra components is by manipulating the benefit component. That is, we consider the benefit of the interaction being a function of characteristics of both an agent i and an agent j, that we denote by $\delta_i(\tau_i, \tau_j)$. This generates for each agent a vector of benefits δ_i whose entries $\delta_{i,j}$ are determined by dyad-specific interaction between characteristics. This representation enables a high level of flexibility through which we can make assumption

[7] See Jackson and Rogers (2007) for an extensive discussion of real world networks main characteristics.

on underlying factors that matter for link decisions, other than grade-related components. One natural way to proceed would be to assume that benefits are decreasing in proximity of characteristics, that is agents have preferences for homophily. This approach would be similar to Boucher (2015). On one hand this adds more structure, but the parameters could be directly obtained from the data.

EXAMPLE 2. It is broadly documented that homophily is an important aspect of social interactions. Therefore here we can assume that benefit increases as there are similarities over characteristics. Formally we may have $\delta_i^m(\tau_i^m, \tau_i^m - i) = \beta_m |\tau_i^m - \tau_{-i}^m|^{-1}$, which determines the benefit attached over a single characteristic. The total benefit for an agent i can be linearly collected into $\delta_i = \sum_{m \in M} \delta_i^m$.

This slight modification allows us to rationalize possibly every network structure observed in reality. Moreover, all the coefficients could be tested with enough information available about agents' characteristics and the network structure. Let us assume that there is a given distribution θ, and along with it there is another dimension τ, representing a taste not correlated with grade-defining characteristics. We assume $\theta \in [2, 4]$, which is in a reasonable range of theoretical grades for the students considered in this example. Moreover, $\tau \in [0, 1]$ is assumed to enter the model through the benefit of the interaction. Thus we have $\delta_i^m(\tau_i^m, \tau_i^m - i) = \beta_m |\tau_i^m - \tau_{-i}^m|^{-1}$.

One remark that follows this result, is to stress the importance of understanding the set of relevant characteristics over which students will form links. Although some of those may be unobserved to the econometrician, if one has information on freshmen, observing a set of observables and their admission test results and then recording the links formed by the students, we would be able to detect such characteristics.

This is crucial for a broad understanding of the motives of link formation, that in general are quite hard to detect and few works aim to do so with a robust empirical analysis. On the other hand, this set of relevant characteristics can be argued to be completely context dependent. In such a case, although we lose the external validity of the eventual findings, it would be in the interest of a school planner willing to design an optimal policy because the success of that policy relies entirely on the willingness of students to interact after an alteration of the composition of the classroom.

5 Numerical exercise

The scope of this section is two-fold. On one hand we show how the model can explain real world problems of relevance. On the other hand we believe it is an interesting way to show the mechanics of the model in practice, and the richness of its implications.

This exercise is strongly related to an empirical investigation based on field experiment, found in the paper by Carrell et al. (2013). There are several reasons why we refer to that paper. First of the authors there were initially interested on maximizing the educational outcome of students in a school, the *USAFA Academy*. They wanted to exploit solely peer effects to achieve that goal. Thus, given that they had full disposable power composing classrooms, the authors tried to determine the optimal way of doing so. Initially they were driven by some benchmark testing, were classrooms were randomized. Randomization allowed them to identify peer effects through the classical linear-in-means model proposed initially by Manski (1993).

In particular they provide empirical evidence of heterogeneous peer effects among different quartiles of the educational outcomes. Results could be summarized by saying that significant and strong positive peer effects affected *low achievement students* (LAS), while nothing was found on *medium achievement students* (MAS) or *high achievement students* (HAS). Moreover, the outcome of LAS proved to be sensitive to the proportion of HAS being present in the classroom. In other words, an additional percentage of HAS in the classroom have consistent positive effects on the outcome of LAS, while presence of other students lead to insignificant results.

The authors pushed forward the inquiry, and basically they asked themselves, in light of the previous findings, how could be set an optimal policy regarding the composition of classrooms, with the aim to maximize students' outcomes. A simple linear programming problem was set therefore to determine the composition of classrooms that would increase LAS's output, without harming other students. The program yielded a theoretically pareto-improving composition of classrooms such that: LAS are mixed with HAS, and MAS are let to form homogeneous classrooms. In theory, this maximize LAS output, leaving every one else equal. Remember that there is no evidence of negative peer effects at the top tiers of the distribution of outcomes.

Therefore the experiment was set, and for two years the authors let half of the classrooms be the control group, where students are conventionally selected at random within freshmen. The other half, given the distribution of outcomes, was divided in

LAS, MAS and HAS. LAS and HAS formed the so called *bimodal* classrooms, while MAS were put in the *homogeneous* ones.

Results were totally unexpected by the authors.[8] On average, treated LAS were outperformed by control group LAS (-0.061). Treated MAS did consistently better than the control group (+0.082), while HAS exhibited no significant change. Because of this unexpected outcome, the authors performed a post-experimental survey asking about friendships and how group of study took place. The survey provided evidence of segregation in the bimodal classrooms. The article finally concludes like follows:

"*Despite our emphasis on endogenous sorting, we are unable to reject a related story in which the presence of middle ability students is a crucial part of generating positive peer effects for the lower ability students.*"[9]

This evidence motivates us to proceed as follows: *(i)* we generate a population of "students" picking at random values within an interval of ex-ante educational outcomes (GPA), consistent with the summary statistics of the afore mentioned experiment. *(ii)* The model is fully simulated. As an outcome we obtain stable networks and all their statistics, and ex-post educational outcome. The difference between ex-post and ex-ante leaves us with a measure of peer effects, given the networks that took place strategically. *(iii)* Perform a similar analysis of the paper by Carrell et al. (2013) and show that results from the simulated model consistently reproduce the experimental results.

5.1 Simulations

The algorithmic approach for solving the stable network has the advantage of providing the tools to simulate the model.[10] So the approach to this exercise is to simulate some data, and then have the network take place strategically. We generate a sample of $n = 1020$ individuals, with $\theta_i \in [2, 4]$ for all $i \in N$. In terms of output we have therefore stable networks and vectors of ex-post schooling performances. Other parameters are set as follows: i) $\alpha = 0.5$ and $\delta = 0.036$. In a first step we build the

[8] See Carrell et al. (2013), pag. 871.

[9] See Carrell et al. (2013), pag.881.

[10] All the codes used to produce this section are available at `https://github.com/ugobolletta/Peer-effects-in-school`. The results shown here can be exactly reproduced following the instructions provided.

sample generating uniformly at random values for agents' θs. This will be our baseline to measure peer effects in the control group. Then we build a sample of treated groups following Carrell et al. (2013). As in that paper, students are sorted into categories calculating terciles of the distribution, which allows us to label observations according to their category. In this way we can reproduce an experimental scenario where we can study the effect of the composition of classrooms on individuals' outcomes. The observations in the treated group are $n_t = 1020$. This number allows us to distinguish between Bimodal and Homogeneous treatments while still having plenty of statistical power to produce consistent estimates. The empirical strategy is simple because we exploit the exogenous treatments and also random assignment. Therefore the effect here could be identified with a simple pre-post as the independent variable, regressed over the dummy treatment.

$$y_i - \theta_i = \beta_0 + \beta_1 treatment_i + \varepsilon_i \tag{8}$$

where standard errors are clustered at the classroom level. Identification follows by random treatment allocation, and the estimates are reported in table 1.

The routine

1. Evaluate $y(g)$,
2. Pick an i and j at random,
3. Add the link between i, j, update $y(g + \{ij\})$ and evaluate payoff,
4. **if** $U_i(\cdot, g + \{ij\}) > U_i(\cdot, g)$ **and** $U_j(\cdot, g + \{ij\}) > U_j(\cdot, g)$ **end**.
5. **elseif** $i < j$, **for each** $k \in N_i(g)$ with $k < i$ eliminate links sequentially, from the further left, update $y(g + \{ij\} - \{ki\})$, evaluate payoff. **If** $U_i(\cdot, g + \{ij\} - \{ki\}) > U_i(\cdot, g)$ **and** $U_j(\cdot, g + \{ij\} - \{ki\}) > U_j(\cdot, g)$ **end**.
6. Same as above if $j < i$.
7. Iterate
8. **Return** $G, \mathbf{y}$

VARIABLES	(1) Difference (LAS)	(2) Difference (MAS)	(3) Difference (HAS)
Bimodal	-0.0553***		0.00777*
	(0.0113)		(0.00458)
Homogeneous		0.0391***	
		(0.00785)	
Observations	662	672	676
R-squared	0.103	0.073	0.007

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1: Here we show the results of the regression described by equation 8. The first row reports the treatment effect for individuals in the Bimodal treatment, which by construction are LAS and HAS only. LAS performed worse that their counterpart in the control group, while HAS has benefited from the treatment but the magnitude of the effect is not sizable and is poorly estimated. The MAS that where pooled together in the Homogeneous treatment performed better than in the control group, consistently with empirical results. The differences in sample sizes are due to slight imbalance between control group and treatment groups.

The mechanism behind the results is the network configuration that emerged by the strategic decisions of individuals. The results can therefore be explained through a neat segregation of groups in the Bimodal treatment group. It seems clear that the presence of MAS is important because in absence of external incentive agents in HAS seem not to be willing to interact with LAS agents. Thus, the students in the middle category favor those in the lowest category by improving their outcomes. We report, along with the results of the regression in Table 1, a visualization of the network structure for the control group, the Bimodal treatment and the Homogeneous treatment in Figure 3.

6 Discussion

Designing interventions to improve outcomes of targeted categories of students is something worth investigating. With the goal in mind of providing education of quality for the largest number of individuals, an educative system could provide incentives in order to foster peer effects, and get a step closer to that goal, while keeping the

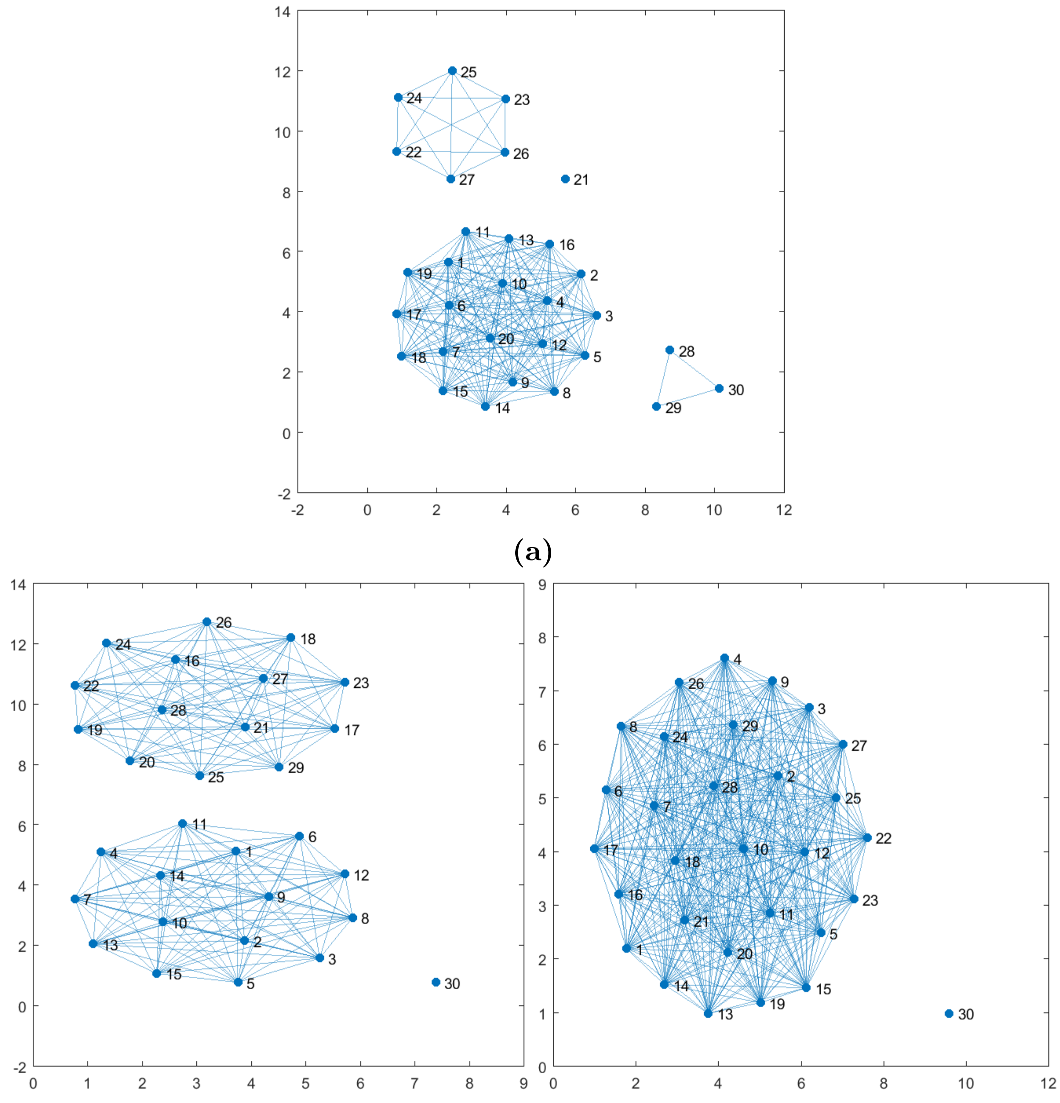

Figure 3: Figure (a) reports the stable network for one of the classrooms belonging to the control group. Agents are labeled according to their ex-ante ability, from lower to higher. We see how top students tend to avoid interactions with the others, which instead formed a dense component of interactions. The picture is very different in Figures (b) and (c), which reports respectively a stable network in the Bimodal group (b), and in the Homogeneous group (c).

costs contained.

Through this paper we contribute in this direction providing a theory that draws some guidelines on how to design policies while taking into account the unobserved responses of individuals to the exogenous intervention of the planner. We learned that segregation should be avoided, and that having an environment rich of social interactions could benefit the students that perform poorly. Another implication of the model proposed here, is that it seems that the better performing individuals are those that do not accept to be held back by interacting with their colleagues. It is possible that a well-designed incentive scheme, like short-time sessions where students of different ability interact with each other, could provide high returns, with low losses. Evidence of such mechanism is provided by Li et al. (2014). Having the good students interacting in short repeated sessions of study with those that lagged behind, improved consistently the outcome of the latter category. We believe that these kind of instruments should be privileged by schools, under the availability of a great knowledge of the environment.

It is important that further empirical research is conducted on this topic. Consider a scenario where an econometrician can observe both network and agent's characteristics. Then we may be able to use the model proposed here to estimate peer effects, calibrating the key parameters in order to match the observed data. Moreover, if we would observe students in multiple time points, we may use the information on ex-ante characteristics, such those of freshmen, then calibrate the network, which would yield an extensive understanding of the magnitude of peer effects and the motives behind link formation. There are two possible approaches to extend the use of the model that we propose: *i)* to use an existing dataset and match the parameters so that we minimize the predictions of the model to the observed outcomes. Once we learn the context we could then use further the theory to simulate policy experiments in order to detect the preferred strategy. *ii)* To perform an educated data collection on the field and implement the interventions with real subjects. The approaches range from a semi-structural analysis to experimental. We believe that promoting the inter-relation between theory and empirics is the only way to design a successful "optimal policy". On the contrary, in the absence of sufficient information, it may be preferable to not interfere with the "natural variation".

ACKNOWLEDGMENTS: We would like to thank in particular Giulio Zanella, Paolo Pin, Yann Bramoullè and Berno Buechel for their insightful comments and suggestions. Moreover a special thanks goes to the attendants of Internal seminar series in University of Bologna and Aix-Marseille, the Behavioral & Experimental Lunch Seminar in Stanford University and the CTN Workshop (2017) for their constructive questions and comments that helped me improving this work. I am thankful to AMSE for the economic support. All errors and mistakes remain my own.

References

Badev, A. (2013). Discrete games in endogenous networks: Theory and policy.

Ballester, C., A. Calvó-Armengol, and Y. Zenou (2006). Who's who in networks. wanted: the key player. *Econometrica 74*(5), 1403–1417.

Barnes, J., K. M. Beaver, J. T. Young, and M. TenEyck (2014). A behavior genetic analysis of the tendency for youth to associate according to gpa. *Social Networks 38*, 41–49.

Bolletta, U. and P. Pin (2018). Polarization when people choose their peers. *SSRN Manuscript DOI: http://dx.doi.org/10.2139/ssrn.3245800*.

Boucher, V. (2015). Structural homophily. *International Economic Review 56*(1), 235–264.

Boucher, V. (2016). Conformism and self-selection in social networks. *Journal of Public Economics 136*, 30–44.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009). Identification of peer effects through social networks. *Journal of econometrics 150*(1), 41–55.

Bramoullé, Y. and R. Kranton (2007). Public goods in networks. *Journal of Economic Theory 135*(1), 478–494.

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. Patacchini, and Y. Zenou (2009). Peer effects and social networks in education. *The Review of Economic Studies 76*(4), 1239–1267.

Carrell, S. E., B. I. Sacerdote, and J. E. West (2013). From natural variation to optimal policy? the importance of endogenous peer group formation. *Econometrica 81*(3), 855–882.

Chetty, R., N. Hendren, and L. F. Katz (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review 106*(4), 855–902.

De Giorgi, G. and M. Pellizzari (2014). Understanding social interactions: Evidence from the classroom. *The Economic Journal 124*(579), 917–953.

Dutta, B. and M. O. Jackson (2013). *Networks and groups: Models of strategic formation*. Springer Science & Business Media.

Galeotti, A. and S. Goyal (2010). The law of the few. *The American Economic Review*, 1468–1492.

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.

Jackson, M. O. and B. W. Rogers (2007). Meeting strangers and friends of friends: How random are social networks? *The American economic review*, 890–915.

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

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

Kinateder, M. and L. P. Merlino (2014). Public goods in endogenous networks. *Unpublished Manuscript*.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. *Econometrica 75*(1), 83–119.

Li, T., L. Han, L. Zhang, and S. Rozelle (2014). Encouraging classroom peer interactions: Evidence from chinese migrant schools. *Journal of Public Economics*.

Ludwig, J., G. J. Duncan, L. A. Gennetian, L. F. Katz, R. C. Kessler, J. R. Kling, and L. Sanbonmatsu (2012). Neighborhood effects on the long-term well-being of low-income adults. *Science 337*(6101), 1505–1510.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies 60*(3), 531–542.

Marmaros, D. and B. Sacerdote (2006). How do friendships form? *The Quarterly Journal of Economics 121*(1), 79–119.

Saia, A. (2018). Random interactions in the chamber: Legislators' behavior and political distance. *Journal of Public Economics 164*, 225–240.

Watts, A. (2001). A dynamic model of network formation. *Games and Economic Behavior 34*(2), 331–341.

Appendix A

Proof of Lemma 1

Proof. The problem is represented by a system of N equations in N unknown variables. By linearity of equations and restrictions on the parameter α the solution of the system is always unique. If $\alpha = 1$ there is a continuum of equilibria, but we are not interested in such a case, and the restriction on the parameter is assumed throughout the rest of the paper. To show that $y_i^* \in [\underline{\theta}, \tilde{\theta}]$ we could use a simple contradictory argument. For instance just assume that there is an agent ℓ such that $y_\ell > \tilde{\theta}$ and that it is the maximum value among all y_i^*. Given that behavior in equilibrium is a convex combination of $\bar{\theta}_\ell$ and θ_ℓ, as stated in equation 5, that $\theta_\ell \in [\underline{\theta}, \tilde{\theta}]$, $y_\ell > \tilde{\theta}$ if and only if $\bar{\theta}_\ell > y_\ell$, which contradicts the initial assumption. Same argument can be applied to show that $y_i^* \geq \underline{\theta}$. □

Proof of Remark 1

Proof. As a first step, recall the best reply functions. Given that in equilibrium all agents must play the best response, we can rewrite 5 as follows

$$y_i = \alpha \frac{1}{d(i)} \sum_{j \in N(i)} (\alpha\{\bar{y}_j\} + (1-\alpha)\theta_j) + (1-\alpha)\theta_i$$

Now assume that all have equally sized reference groups. This would hold for the the complete network, or any regular k-degree network. In the special case of the complete network, $k = N - 1$. Therefore all average values are defined for a size of population of $N(i) = k$, for all i. Collecting terms we can write it as

$$y_i = \alpha^2 \bar{y}_{N(i)} + \alpha(1-\alpha)\{\bar{\theta}_j\}_{j \in N(i)} + (1-\alpha)\theta_i$$

where $\bar{y}_{N(i)} = \frac{1}{d(i)} \sum_{j \in N(i)} \{\bar{y}_k\}_{k \in N(j)}$. Note that it is exactly the linear in means model of Manski (1993). In fact, taking the expected value of both sides we have that

$$E(y) = \alpha^2 E(y) + \alpha(1-\alpha)E(\theta) + (1-\alpha)\theta_i + \tilde{\varepsilon}_g$$

from which we get

$$E(y) = \beta_1 \theta_i + \beta_2 E(\theta) + \tilde{\varepsilon}_g$$

where $\beta_1 = \frac{1-\alpha}{1-\alpha^2}$ and $\beta_2 = \frac{\alpha}{1+\alpha}$. □

Proof of Lemma 2

Proof. First we notice that we can rearrange equation 4 like follows:

$$U_i(y_i^*, y_{-i}, \theta_i, g) = -\frac{1}{2} y_i^{*2} + \alpha y_i^* \frac{1}{d_i} \sum_{j \in N_i(g)} y_j + (1-\alpha) y_i^* \theta_i \tag{9}$$

From 9, combining with 5 we get the following

$$U_i(\cdot) = -\frac{1}{2}\left(\alpha \bar{y}_i + (1-\alpha)\theta_i\right) + \alpha\left(\alpha \bar{y}_i + (1-\alpha)\theta_i\right)\bar{y}_i + (1-\alpha)\left(\alpha \bar{y}_i + (1-\alpha)\theta_i\right)\theta_i \tag{10}$$

Then simply solving and rearranging terms we get

$$\begin{aligned} U_i(\cdot) &= \frac{1}{2}\alpha^2 \bar{y}_i^2 + \alpha(1-\alpha)\bar{y}_i\theta_i + \frac{1}{2}(1-\alpha)^2\theta_i^2 \\ &= \left(\frac{1}{\sqrt{2}}\alpha \bar{y}_i + \frac{1}{\sqrt{2}}(1-\alpha)\theta_i\right)^2 \end{aligned} \tag{11}$$

From 11 it is easy to check that $\frac{\partial U_i(\cdot)}{\partial \bar{y}_i} > 0$. Now call $\bar{y}_i'$ the term such that $k \in N_i$ and $j \notin N_i$ and $\bar{y}_i$ the term such that $j \in N_i$ and $k \notin N_i$. The simple observation that $\bar{y}_i' > \bar{y}_i$ completes the proof. □

Proof of Proposition 4

Proof. We prove the result by contradiction. We find beneficial to state formally the definition of the marginal payoff coming from a link, which will reduce consistently the notational burden of the proof.

Definition 5. ***LINK MARGINAL PAYOFF***: *Let $g \in G$. For all $i, j \in N$ such that $ij \in g$*

$$mu_i(g, ij) = u_i(g) - u_i(g - ij)$$

is the marginal payoff to i from the link ij in g.

Every ij with $j > i$ is such that $mu_i(g, ij) > 0$. So announcing links to any $j > i$ is strictly profitable, irregardless of the value of δ, as it was proved in Lemma 1.

Assume we have reached a stable state. We know that no couple can benefit from the addition of a link that has not yet been formed, and no agent want to sever any existing link. We claim that such stable network is locally complete. Assume it is not. This means that we have at least one agent k such that $i < k < j$, with $g_{ij} = 1$ and such that $g_{ik} = 0$ or $g_{kj} = 0$. Consider the first case. We know that $g_{ij} = 1$, so $mu_i(g, ij) > 0$ and $mu_j(g, ij) > 0$. We know also from Lemma 2 that $mu_i(g, ik) > 0$ for any $N_i(g)$. However, since $mu_j(g, ij) > 0$ it follows that $mu_k(g, ik) > 0$. Thus g_{ik} is profitable for both i and k. A first contradiction. Now let us consider the second case. Similarly we know from Lemma 2 that $mu_k(g, kj) > 0$. However we know that $mu_j(g, ij) > 0$, and thus either $mu_j(g, kj) > 0$ irregardless of the existing link with i, or $mu_j(g, kj) > 0$ if a the link g_{ij} would be deleted. In the latter case we know that $mu_j(g, kj) > mu_j(g, ij)$, so j would be better off deleting the link with i and form it with k. Another contradiction. Therefore we conclude that the possible configurations are either $g_{ik} = g_{kj} = g_{ij} = 1$, or $g_{kj} = 1$ and $g_{ij} = 0$, both of which are locally complete. □