Structural Homophily

Vincent Boucher

Structural Homophily

Vincent Boucher*

January 2012

Abstract

Homophily, or the fact that similar individuals tend to interact with each other, is a prominent feature of economic and social networks. Most existing theories of homophily are based on a descriptive approach and abstract away from equilibrium considerations. I show that the equilibrium structure of homophily has empirical power, as it can be used to recover underlying preference parameters.

I build a non-cooperative model of network formation, which produces a unique, empirically realistic equilibrium network. Individuals have homophilic preferences and face capacity constraints on the number of links. I develop a novel empirical method, based on the shape of the equilibrium network, which allows for the identification and estimation of the underlying homophilic preferences. I apply this new methodology to race-based choices regarding friendship decisions among American teenagers.

JEL Codes: D85,C72,C13

*Department of Economics, Université de Montreal and CIREQ: vincent.boucher@umontreal.ca
I would like to thank my advisors Onur Ozgur and Yann Bramoullé for sharing their time, and for their most valuable comments and discussions. I also want to thank Lars Ehlers and Marc Henry for their precious help, and for many discussions and comments. Thanks also to Paolo Pin for his helpful comments. I would also like to thank Ismael Mourifie, Louis-Philippe Béland, Yousef Msaid and David Karp, as well as the participants at various conferences, including those of the Canadian Economic Association (2011), Coalition Theory Network (2010), Societe Canadienne de Science Economique (2010), Econcon (2011), and Groupe de Recherche International (2011) for their questions, comments and suggestions. Finally, I gratefully acknowledge financial support from the CIREQ, the FQRSC and the SSHRC.
This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations.

1 Introduction

The fact that similar individuals tend to interact with each other is a prominent feature of social networks. The phenomenon, referred to as *homophily*, is increasingly being studied by economists.[1] Indeed, the structure of the social networks in which individuals interact has been shown to significantly influence many social outcomes such as segregation,[2] information transmission and learning,[3] and employment and wages.[4] Being able to understand, identify, and measure how the social characteristics of an individual influence network formation is therefore of central importance. However, most studies to date overlook the equilibrium implications of homophily, and disregard key factors such as the impact of time constraints.

In this paper, I develop an empirically realistic model of strategic network formation incorporating homophilic preferences and capacity constraints on the number of links. My analysis uncovers novel structural predictions generated by the equilibrium interplay between the individuals' homophilic preferences and capacity constraints. Building on the explicit structure of homophily obtained in equilibrium, I develop a new estimation technique that allows one to recover underlying preferences parameters. I show as an illustration that the formation of friendship networks among American teenagers is strongly influenced by racial considerations. I also show that this preference bias toward individuals of the same race varies considerably with respect to the racial group considered.

The emphasis on the equilibrium implications of homophilic preferences is new to the literature. The equilibrium network resulting from the theoretical model exhibits more structure than the known stylized facts regarding homophilic patterns in social networks.[5] The equilibrium network architecture allows for an original empirical methodology using a maximum likelihood approach. A key feature of the estimation strategy is that it recovers explicit preference parameters characterizing homophily in social networks. In other

[1] See for example Currarini et al. (2009), Bramoullé et al. (2012), and Rivas (2009).
[2] Echenique and Fryer (2007), Watts (2007), and Mele (2010).
[3] Golub and Jackson (2010a,2010b).
[4] van der Leij et al. (2009) and Patacchini and Zenou (2009).
[5] See Bramoullé et al. (2012), and Currarini et al (2009).

words, the estimation strategy allows for the identification of *preference interactions* from *constraint interactions*.[6]

The theoretical framework produces sharp predictions. There exists a generically unique, empirically realistic equilibrium network. A key assumption is that the homophilic preferences of individuals can be represented by a distance function on the set of characteristics of the individuals. This idea is implicitly or explicitly exploited by many papers looking at homophily in social networks.[7] This assumption allows me to introduce enough heterogeneity in the model to generate empirically realistic equilibrium networks. I also assume that individuals have link-separable utilities, and an explicit resource constraint, such as time. For example, while a teenager may prefer to be friends with other teenagers who have similar characteristics, he must take into account the fact that he has limited time to spend with the friends he chooses to have. Hence, the resource constraint explicitly introduces an upper bound on the number of bilateral relationships an individual can sustain.[8] The specific notion of homophily emerging in equilibrium results from the tension between the individuals' homophilic preferences and the individuals' resource constraint. These two premises imply a novel theoretical prediction on the shape of homophily in equilibrium. I call this specific network architecture *structural homophily*.

Structural homophily describes an explicit relationship between individuals' socioeconomic characteristics and the network architecture. An individual is characterized by a "social neighborhood" on the space of individual characteristics.[9] This neighborhood explicitly determines the set of acceptable bilateral relationships. In a network characterized by structural homophily, two individuals are linked if and only if they belong to the intersection of their neighborhoods. These neighborhoods are not directly observable, but are implied by equilibrium predictions of the theoretical model for a given a distance function. This novel theoretical prediction has empirical power.

[6]Manski (2000) distinguishes between three sources of social interactions: Preference interactions, Constraint interactions, and Expectations interactions.

[7]See for instance, Johnson and Gilles (2000), Marmaros and Sacerdote (2006), Iijima and Kamada (2010), Mele (2010) and Christakis et al. (2010).

[8]It relates to the sociological and psychological observation referred to as the Dunbar's number.

[9]It relates to the sociological notion of a "social niche"; see for instance McPherson et al. (2001)

I use structural homophily to develop an original estimation strategy. This strategy is based on the duality between the equilibrium network structure, and structural homophily. Any equilibrium network exhibits structural homophily, and any observed network that exhibits structural homophily is an equilibrium network. I develop a maximum likelihood approach, defined over a population of distinct social networks. The empirical method allows for the identification and estimation of prominent socioeconomic characteristics affecting the equilibrium network structure. This is relevant for policy making since it allows the policy maker to *target* relevant socioeconomic characteristics. As an illustration, I use data on the friendship networks of American teenagers provided by the Add Health database.[10] I focus the analysis on race-based choices and show that the same-race preference bias substantially varies with respect to racial group. Blacks have a stronger bias than Asians, while Whites have the smallest bias. The estimated coefficients are preference parameters, and hence do not depend on the distribution of the racial groups in the population, nor do they depend on the individuals' resource constraints.

This paper contributes to the theoretical and the empirical literature on network formation. Most theoretical models of network formation produce relatively structured equilibrium networks such as stars, circles or chains.[11] These models, although highly relevant from a theoretical perspective, are not well suited for empirical purposes. Indeed, the resulting set of equilibrium networks is both too large (many equilibrium networks) and too constraining (stars, chains, circles, etc.) to represent actual, observable, social networks. Most theoretical models assume that payoffs depend on detailed features of the network structure, but neglect the capacity constraints on the number of links an individual can make.[12] I show that the introduction of this constraint, combined with explicit ex-ante homophilic and link-separable utilities, implies the existence of a unique, empirically realistic equilibrium network.[13]

[10]Carolina Population Center, University of North Carolina at Chapel Hill; see http://www.cpc.unc.edu/projects/addhealth.

[11]Bala and Goyal (2000), Jackson (2008, chapter 6), Jackson and Rogers (1997), Jackson and Wolinsky (1996), and Johnson and Gilles (2000).

[12]Exceptions include Bloch and Dutta (2009) and Rubí-Barceló (2010).

[13]I concentrate on strategic models of network formation. There exists a large literature on random network formation, which is not directly concerned with the current setting. The interested reader can see

Two alternative explanations of homophily have been proposed. The first is through correlations in the meeting process:[14] individuals have no preference bias, but individuals with similar characteristics have a higher probability of meeting. The second is through preference biases:[15] individuals prefer to link with similar individuals. In this paper, I assume that individuals have homophilic preferences, but evolve in a deterministic world. I analyze the equilibrium implication of these preferences in a fully strategic, non-cooperative setting.

The empirical literature on network formation is still in an early stage. The few existing papers clearly identify homophily as a driving factor of the network formation process.[16] This paper contributes to the literature on strategic network formation by providing an estimation strategy based on the equilibrium structure of homophilic preferences. Equilibrium considerations are important, as they imply a departure from link-level estimation techniques. The model defines a precise dependence structure which allows for the definition of an explicit maximum likelihood estimator.[17]

The remainder of the paper is organized as follows. In section 2, I present the theoretical model and key definitions. In section 3, I find and characterize the (unique) equilibrium network. In section 4, I describe the empirical methodology and explore its properties using Monte Carlo simulations. In section 5, I present an application of race-based homophily in friendship networks using the Add Health database. I conclude in section 6.

2 The Theoretical Model

In this section, I present a non-cooperative model of network formation that characterizes the equilibrium effects of homophily. The model generically produces a unique equilibrium. I first provide a formal definition of Structural Homophily. Next, I outline the theoretical

for instance Jackson (2008, chapters 4 and 5) and the references therein.

[14] See for instance Bramoullé et al. (2012)

[15] See also Currarini et al. (2009), and Mele (2010)

[16] See for instance Christakis et al. (2010), Mele (2010), Currarini et al. (2010), and Franz et al. (2008)

[17] As opposed to the simulated maximum likelihood estimators, as in Christakis et al. (2010), and Mele (2010).

framework, and finally, I briefly present the main definitions and equilibrium concepts.

2.1 Structural Homophily

In order to introduce this new notion of homophily, we need some preliminary assumptions. There is a finite set of individuals N. Individuals may be linked together through a network. Let $g_i \subseteq N$ be the set of individuals linked to individual i for all $i \in N$. Each individual $i \in N$ is characterized by a type $\theta_i \in \Theta$, where Θ is the type space. An individual's type could represent, for instance, a series of socioeconomic characteristics. I consider a distance d on Θ. For notational simplicity, let $d_{ij} \equiv d(\theta_i, \theta_j)$ for any $i, j \in N$. Then, *structural homophily* is defined as follows.

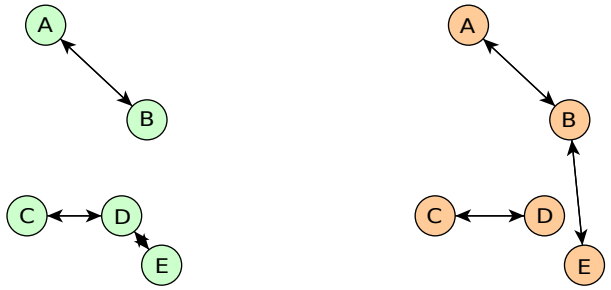
Definition 1 *A network g exhibits **structural homophily** with respect to a distance $d(.,.)$ if whenever two individuals, i and j, are not linked, either $d_{ij} \geq \max_{k \in g_i}\{d_{ik}\}$ or $d_{ij} \geq \max_{k \in g_j}\{d_{jk}\}$.*

This definition formalizes the fact that two individuals that are "close" should be linked. Intuitively, if two individuals are not linked, it is because, from the point of view of one of the individuals, the other is located relatively too far. Notice that this definition only makes sense when the creation of a link requires mutual consent. Figure 1 shows two examples of networks for $\Theta = \mathbb{R}^2$. The first network exhibits Structural Homophily, but the second does not. In Figure 1b, the closest individuals (i.e. D and E) are not linked, which is in contradiction with structural homophily since D is linked to C, and E is linked to B.

More insight can be obtained by drawing the equivalence (or indifference) curves corresponding to the *farthest link* for each individuals considered (i.e. for B and D in Figure 2a, and for D and E in Figure 2b). These equivalence curves define neighborhoods; every individual inside the neighborhood of i is at a distance smaller the distance between i and his farthest link. If both individuals belong to the intersection of the two neighborhoods generated by the equivalence curves (as in Figure 2b), then Structural Homophily

Figure 1: Structural Homophily

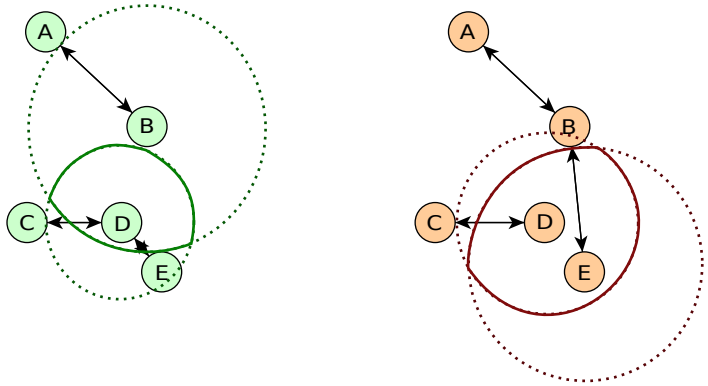
(a) Respected (b) Violated



is violated.[18]

Figure 2: Structural Homophily: Equivalence curves

(a) Respected (b) Violated



Structural homophily has an interpretation in terms of revealed preferences. Suppose that individuals have preferences over links with other individuals, and that such preferences are a function of the distance between the individuals. Suppose also that we observe the network (i.e. the individuals and their links), and the types of the individuals in the network (i.e. a series of individual characteristics). Then, under mutual consent, we should not observe networks such as the one depicted in Figure 2b. That is, structural homophily should hold.

[18]This closely relates to the *cutoff rule* of Iijima and Kamada (2010).

It it interesting to note that small-worlds networks respect structural homophily for a specific type space.[19] In a small world model, individuals are located on *islands*. In that setting, structural homophily implies that individuals are linked first with individuals of the same island. Hence, if there is a link between two islands, those islands have to be fully connected. I now present a social networking game, which produces Structural Homophily at equilibrium.

2.2 The Game

There are n individuals, each of whom is endowed with a fixed amount of resources $\bar{x}_i = \kappa_i \xi$, where $\xi \in \mathbb{R}_+$ and $\kappa_i \in \mathbb{N}$. We will see that, in equilibrium, κ_i is interpreted as the maximum number of links that an individual i can have. A strategy for an individual i is a vector $x_i = (x_i^1, ..., x_i^n) \in X_i$, where $X_i = \{x_i \in \mathbb{R}_+^n | x_i^j \leq \xi, \ and \ \sum_{j \in N} x_i^j \leq \kappa_i \xi\}$. Then, ξ plays the role of a link-level constraint. The introduction of the link-level constraint is motivated by the empirical fact that the number of links varies across individuals. Let $X = \times_{i \in N} X_i$. We say that there is a link between an individual i and an individual j iff $x_i^j > 0$ and $x_j^i > 0$. Let $g_i = \{j \in N | i \text{ and } j \text{ are linked}\}$, so $j \in g_i$ iff $i \in g_j$. That is, a link exists iff both individuals invest a strictly positive amount of resources in it. Notice that individual i can be linked to himself.

The utility of an individual is given by the function $u_i : X \to \mathbb{R}$. It is additive in the different links he has, and it is represented by :

$$u_i(x) = \sum_{j \in N \setminus \{i\}} v_i(x_i^j, x_j^i, d_{ij}) \cdot \mathbb{I}_{\{j \in g_i\}} + w_i(x_i^i) \cdot \mathbb{I}_{\{i \in g_i\}}$$

where $\mathbb{I}_{\{P\}}$ is an indicator function that takes value 1 if P is true, and 0 otherwise. The function $v_i(x, y, d)$ gives the value of any link for i. It is assumed to be twice continuously differentiable with $v_x(x, y, d) > 0$ if $y > 0$, $v_y(x, y, d) > 0$ if $x > 0$, and $v_d(x, y, d) < 0$ if $x, y > 0$. The function $w_i(x_i^i)$ represents the payoff received from the private investment of

[19] See for instance Jackson and Rogers (2005) and Galeotti et al. (2006).

i.[20] It is also twice continuously differentiable with $w'(x) > 0$. I also allow for the presence of fixed costs, i.e. $v_i(0,0,d) \leq 0$ and $w_i(0) \leq 0$. Notice that an individual benefits from a link only if both individuals invest in the link. The model induces a game Γ between the n individuals. Formally, we have $\Gamma = (N, \{X_i\}_{i \in N}, \{u_i\}_{i \in N})$.

The model has two important features. First, the initial endowment creates scarcity and induces a feasibility constraint. This effect is typical of any matching model. If some individual i invests resources in a link with individual j, he will have less available resources to create a link with another individual. That is, the feasibility constraint implies a tradeoff between the distance between two individuals, and the level of investment they put in the link. This is what Manski (2000) refers to as "constraint interactions". Second, the preferences are affected by the presence of direct externalities. The amount of resources invested by some individual in a given link directly affects the utility of the individuals he links to. That is, in Manski's terms, "preference interactions". Those two features will play an important role in equilibrium.

This completes the description of the game. I now present the main definitions.

2.3 Definitions

Before turning to the analysis of the model, I introduce some definitions. The collection of links between individuals generates a *network* $g = (N, E)$. A network is characterized by a set of individuals (here, N), and a set E of links, which are (unordered) pairs of individuals. The set of all possible networks is denoted by $\mathbb{G}$. Any network g can be represented by a $n \times n$ *adjacency matrix* A that takes values $a_{ij} = 1$ if $j \in g_i$, and 0 otherwise, for all $i, j \in N$. The *degree* $\delta_i(g)$ of an individual i is the number of links attached to i, i.e. $\delta_i(g) = |g_i|$.

I am interested in the following solution concepts:

Definition 2 *A Nash Equilibrium (NE) is a profile $x^* \in X$ such that $u_i(x_i^*, x_{-i}^*) \geq u_i(x_i, x_{-i}^*)$ for all $x_i \in X_i$, and for all $i \in N$.*

[20] The function w_i can also be interpreted as the private value of the resource x for i

The set of Nash equilibria is very large. Since an individual benefits only from a collaborative link when both individuals invest in the link, it will never be profitable to unilaterally start a new link. For this reason, I will focus on the following solution concept, introduced by Goyal and Vega-Redondo (2007).

Definition 3 *A Bilateral Equilibrium (BE) is a profile $x^* \in X$ such that :*
(1) x^* is a Nash Equilibrium
(2) There exists no $i, j \in N$, such that $u_i(x_i, x_j, x^*_{-i-j}) > u_i(x^*)$ and $u_j(x_i, x_j, x^*_{-i-j}) \geq u_j(x^*)$ for some $x_i \in X_i$ and $x_j \in X_j$.

This solution concept allows for bilateral deviations. This is a natural extension of individual rationality, since individuals can benefit from the creation of links. For certain economies, however, the BE concept will be too constraining. Accordingly, I also introduce the following weakened equilibrium concept.

Definition 4 *A Weak Bilateral Equilibrium (WBE) is a profile $x^* \in X$ such that :*
(1) x^* is a Nash Equilibrium
(2) There exists no $i, j \in N$, such that $u_i(x_i, x_j, x^*_{-i-j}) > u_i(x^*)$ and $u_j(x_i, x_j, x^*_{-i-j}) > u_j(x^*)$ for some $x_i \in X_i$ and $x_j \in X_j$.

In a WBE, a deviation must strictly increase the payoff of both individuals involved. Notice that $BE \subseteq WBE \subseteq NE$. I discuss the distinction between these concepts in section 3.1 (lemma 3.1 and proposition 3.5).

3 Equilibrium Characterization

I first show the existence of an equilibrium. Since the payoff functions are not continuous, we cannot directly use the standard fixed-point arguments. The existence of a NE is straightforward. Let $x_i^j = 0$ for all $j \neq i$. Then, for every individual, the maximization problem becomes: $\max_{x_i \in X_i} w(x_i^i) \cdot \mathbb{I}_{\{i \in g_i\}}$. The allocation $x^* \in X$ that maximizes this problem for all $i \in N$ is obviously a NE. In order to show the existence of a WBE (or a BE), I will need to introduce additional assumptions. The next result provides an intuition

on the additional restrictions imposed by the bilateral stability on the solution set. It states that if a deviation is jointly profitable, but not unilaterally profitable, the deviating individuals have to invest more in their collaborative link. All proofs can be found in appendix A.

Lemma 3.1 *If $x^* \in X$ is a NE, but not a WBE, given any deviating pair (i,j), with profitable deviations $x_i \in X_i$ and $x_j \in X_j$, we have $x_i^j > x_i^{j*}$ and $x_j^i > x_j^{i*}$.*

Since x^* is a NE, it is individually rational. Also, since the utility functions are additive in the different links, the action of individual j on individual i only affects i through the link between i and j. If x^* is not jointly rational for i and j, the incentive to deviate must come from the link i and j have together.

Throughout this section, I consider two alternative assumptions:

Assumption 1 (Finiteness) *For all $i,j \in N$, $x_i^j \in \{0,\xi\}$*

Assumption 2 (Convexity) *For all $i \in N$, $\frac{\partial^2 v_i}{\partial x^2}(x,y,d) \geq 0$, $\frac{\partial^2 w_i}{\partial x^2}(x) \geq 0$*

The finiteness assumption is extensively used in the literature.[21] Convexity is often assumed when the network formation process involves continuous strategies. For example, Bloch and Dutta (2009) define the strength of a link between individuals i and j as the sum of a (strictly) convex function of the individuals' investment, i.e. $s_{ij} = f(x_i^j) + f(x_j^i)$, with $f' > 0$ and $f'' > 0$. Rubi-Barceló (2010) uses a linear (hence convex) function to represent the payoff from scientific collaboration between two researchers.[22] I provide existence results and show that those two assumptions imply that the equilibrium network exhibits structural homophily.

The next results are based on an algorithm referred to as the *assignment algorithm*, and formally defined in Appendix B. The assignment algorithm uses as inputs: (1) the list of preferences $\{u_i(x)\}_{i\in N}$, (2) the individual characteristics $\{\theta_i\}_{i\in N}$, (3) the resource

[21] See for instance Jackson (2008) chapters 6 and 11.

[22] The value of a scientific collaboration as defined by Rubi-Barceló (2008, p.7) is interpreted as a distance in my model.

constraints $\{\kappa_i\}_{i\in N}$, and (4) the distance function d on Θ. It produces at least one allocation $x \in X$, and any allocation produced is such that $x_i^j \in \{0, \xi\}$ for all $i, j \in N$. When $x_i^j \in \{0, \xi\}$, the payoff that an individual receives from the links can be ranked using the distance function (a small distance implies a big payoff). Accordingly, the assignment algorithm proceeds first by linking the pairs of individuals with the smallest distances (provided that the link is profitable for both individuals, and leads to a higher payoff than the private investment). The following results show that any allocation constructed in that fashion is a WBE, and induces a network that exhibits structural homophily.

Let's start with the finite case. Under Finiteness, the involvement of an individual in some link does not affect the amount of resources he invests in his other (existing) links. The value of a link between two arbitrary individuals is then independent of the other (potential) links. Consequently, we have the following:

Theorem 3.2 (Finite Strategy Space) *Under Finiteness, an allocation is a WBE iff it is produced by the assignment algorithm.*

Under convexity, for a given link, it is also rational for both individuals to invest resources until the link-level constraint ξ is met, provided that it leads to a positive payoff. We then have the following:

Theorem 3.3 (Existence) *Under Convexity, any allocation produced by the assignment algorithm is a WBE.*

Proposition 3.4 gives sufficient conditions so that any individual *has* to invest up to the link-level constraint, in any WBE.

Proposition 3.4 (Uniqueness) *Suppose that the inequalities in Assumption 2 are strict, then any WBE can be produced by the assignment algorithm.*

Then, under Finiteness or Strict Convexity, any equilibrium can be constructed through the assignment algorithm. It is worth noting that under Finiteness, $x_i^j \in \{0, \xi\}$ by assumption, while under Strict Convexity it must hold only in equilibrium.

The above results show the existence of a WBE, but not of a BE. The intuition is the following. Suppose that Finiteness holds, and that the economy contains only 3 individuals: i, j, k. Suppose also that $d_{ij} = d_{ik} < d_{jk}$, and that $\bar{x}_i = \bar{x}_j = \bar{x}_k = \xi$. Finally, suppose that $v_i(\xi, \xi, d_{ij}) = v_j(\xi, \xi, d_{ij}) = v_k(\xi, \xi, d_{ik}) > 0$, while any other link has a negative value. Then, in this example, there is no BE, but there are two WBE (see Figure 3). The reason is that i is indifferent between a link with j or a link with k. So, if i is linked with j, but receives a proposition from k, he will be indifferent between keeping his link with j and replacing it with a link with k (while k would be strictly better off with such a deviation).

In many contexts, however, individuals have many characteristics, and the likelihood of such a circumstance is small. In the absence of such a circumstance, we can show the existence of a BE. Formally,

Proposition 3.5 *Suppose that $d_{ij} \neq d_{kl}$ for any $i \neq j$ and $k \neq l$, then any WBE produced by the assignment algorithm is a BE. Moreover, this equilibrium is unique.*

This implies that if for all $i \in N$, the types $\theta_i \in \Theta$ are drawn from a distribution with a dense support on Θ, then there exists a unique WBE, which is also a BE, $[a.s.]$

Figure 3: WBE and BE

(a) The First WBE (b) The Second WBE

Let's now turn to the characterization of the equilibrium network. Since the level of investment of an individual in a potential link does not depend on the number of links he has, the payoffs are only influenced by the distance. Suppose i and j are linked. Then, the creation of a new link between j and k has no spillover effects on i. This produces important consequences on the shape of the equilibrium network. The next proposition characterizes the allocations produced by the assignment algorithm.

Proposition 3.6 (Characterization) *Let g^* be the network generated by some allocation produced by the assignment algorithm, then*
(1) For all $i \in N$, $\delta_i(g^*) \leq \kappa_i$.
(2) The network g^* exhibits Structural Homophily.

The proof is immediate from the construction through the assignment algorithm. Since investments are maximal in every link, the number of links an individual can have is bounded by the resource constraint κ_i. Also, since the assignment algorithm creates links starting from the ones associated with the smallest distances, the induced network exhibits structural homophily. In essence, under Finiteness or (strict) Convexity, any equilibrium network can be constructed through the assignment algorithm, hence satisfying structural homophily.

Let's now turn to efficiency issues. There are many ways to define efficiency. The first one would be to consider the Pareto criterion. Given Finiteness or Convexity, any BE is Pareto efficient. In fact, we have an even stronger result, which is the fact that any BE is a Strong Nash equilibrium (Aumann, 1959).

Proposition 3.7 *Under Finiteness or Strict Convexity, any BE is a Strong Nash equilibrium.*

Since the utility functions are additive, bilateral stability implies stability in the sense of a Strong Nash equilibrium. However, since the utility functions are non-continuous (and utilities are not transferable), Pareto efficiency does not imply efficiency in the sense of the utilitarian criterion. Consider the following social welfare function:

$$W(x) \quad = \quad \sum_{i \in N} u_i(x)$$

In this case, efficiency is not guaranteed. In particular, one can find examples of economies where the unique BE is efficient (in the sense of the utilitarian and the Pareto criterion), as well as examples of economies where the unique BE is inefficient (in the sense of the utilitarian criterion). This inefficiency comes from two principle sources.

First, under the Finiteness assumption, any efficient allocation $z \in X$ is such that $z_i^j \in \{0, \xi\}$ for all $i, j \in N$ (by assumption). Since an individual values only his own payoff, while the social planner (SP) cares about all individuals, a collaborative link is more valuable for the SP than it is for an individual. (It enters the utility function of both the individuals involved.) The tradeoff between the individual and the collaborative links is then different for an individual than for the SP.

Second, under the (strict) Convexity assumption, another issue arises. Since the SP is willing to trade off the utilities of the individuals, an efficient allocation $z \in X$ need not be such that $z_i^j \in \{0, \xi\}$. For example, suppose that there are no fixed costs, then any network g^* such that $\delta_i(g^*) < \kappa_i$ for some $i \in N$ is inefficient. The reason is that if $\delta_i^* < \kappa_i$ for some $i \in N$, the creation of a link with some agent j (who is willing to invest a small amount ϵ) leads to $v_i(\xi, \epsilon, d_{ij})$ for i. If ϵ is small enough, the loss for j is compensated by the discrete jump in the utility of i. Hence, g^* is inefficient. However, it is possible that such a network g^* is induced by a BE.

This concludes the analysis of the theoretical model. In section 4, I develop an estimation technique derived from structural homophily, and present Monte Carlo simulations.

4 The Econometric Model

In this section, I present the econometric model. I use Structural Homophily in order to estimate the weights of the distance function.[23] I would like to emphasize that the method and results of this section are self-contained. If one was willing to *assume* structural homophily (instead of viewing it as the equilibrium outcome of the non-cooperative game presented in the last section), all the results of this section would apply.

In order to present the econometric model, I introduce the following definition:

Definition 5 *An observation q is*
1) a network $g = (N_q, E_q)$, and

[23] Čopič et al. (2009) also exploit homophily, although in a very different setting, in order to develop their estimation technique.

2) for each individual $i \in N_q$, a vector of R individual socioeconomic characteristics, i.e. $\{\theta^i\}_{i \in N}$, where θ^i is a $1 \times R$ vector.

For a given observation $q \in 1, ..., Q$, I note (g_q, θ_q), where θ_q is $n_q \times R$. Definition 5 implies that an econometrician does not observe the specific level of investment in a link (i.e the link-level constraint), nor does he observe the resource constraint κ_i.[24] Accordingly, given a set of observations $(g_q, \theta_q)_{q=1}^{Q}$, we do not possess enough information to construct the equilibrium network through the assignment algorithm, even assuming some structural form for the utility functions. Specifically, a standard econometric model would be the following. Given a parametric form for the payoff functions (i.e. $\{v_i(x, y, d), w_i(x)\}_{i \in N_q}$), and the distance function (i.e. $d(i, j)$), one would assume that the data is generated by:

$$g_q = \Lambda(\theta_q, \kappa_q, \xi_q, \varepsilon_q; \beta) \tag{1}$$

where Λ is the assignment algorithm, κ_q is the $n_q \times 1$ vector of individual resource constraints, ξ_q is the link-level resource constraint, ε_q is the error term, and β is the vector of parameters to be estimated. Provided that one observes $\theta_q, \kappa_q, \xi_q$, one could, in principle, estimate β. Since κ_q and ξ are typically unobserved in existing datasets, I use a different approach.[25] From section 3's results, I have established that any allocation produced by the assignment algorithm respects structural homophily.[26] My approach will then be to maximize *the likelihood that the observed network exhibits structural homophily.* Accordingly, the distance function will play a central role. I assume the following structural form for the distance function:

$$\ln(d_{ij}) = \sum_{l=1}^{L} \beta_l \rho_l(\theta_i, \theta_j) + \varepsilon_{ij} \tag{2}$$

where $\varepsilon \sim_{iid} N(0, 1)$, and $\rho_l(.,.)$ is a dimension-wise distance function.[27] The vector

[24] Notice that while κ_i is an upper bound to $\delta_i(g)$, they are not necessarily equals. See proposition 3.6.

[25] There are also severe computational and identification issues using the specification in (1).

[26] Also, by observing a network that exhibits structural homophily, one can always find some $v_i(x, y, d)$, κ_i and ξ such that it is produced by the assignment algorithm.

[27] For instance, if $\Theta \in \mathbb{R}^2$, one could choose $\rho_l(\theta_i, \theta_j) = |\theta_i^l - \theta_j^l|$. The proposed structural form is by

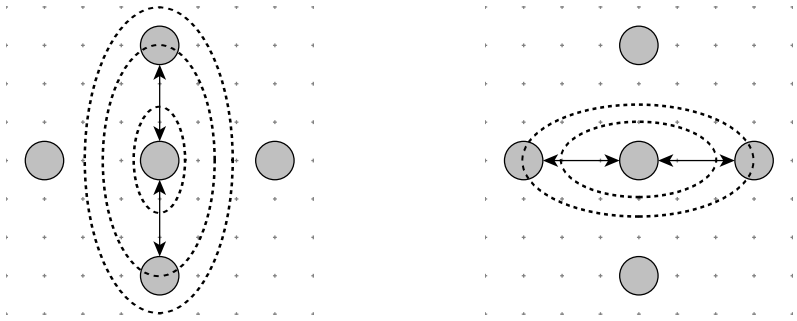
$(\beta_1, ..., \beta_L) \in \Xi \subset \mathbb{R}^L$ are the weights of the distance function. Equation (2) highlights two important features of the model.

First, instead of trying to specifically identify the parameters of the utility function, I limit myself to the estimation of the relative importance of the social characteristics in the network formation process. That is, I only seek to estimate the parameters of the distance function, and not the parameters of the utility functions (for instance, I do not estimate the value of the resource for the individuals). This is illustrated in Figure 4. In Figure 4a, the individuals place more value on the characteristic on the *horizontal* axis. Then, the "closest" individuals for the central node are the ones on the top and bottom. Symmetrically, in Figure 4b, the individuals place more value on the characteristic on the *vertical* axis. Then, the "closest" individuals for the central node are the ones on the left and right. My aim is to estimate the relative weights placed on each characteristics.[28]

Figure 4: Changing the Weight of the Distance Function

(a) Relative Importance on the Horizontal Characteristic

(b) Relative Importance on the Vertical Characteristic



Second, I assume that the distance function is observed with noise. That is, there exists a set of variables, observed by the individuals within the model, but unobserved by

no means the only possibility. Any positive and symmetric function could be used. I prefer to use the specification in 2 to simplify the exposition.

[28]Centered ellipses like those depicted in Figure 4 are implied by the additive form we assumed in (2). The generalization to more general class of distance functions such as in Henry and Mourifie (2011) is straightforward.

an econometrician, that affects the distance function.[29] This assumption is not standard and deserves a discussion.

A typical method to introduce unobserved heterogeneity into this type of models would be to assume that the value of a link depends on some unobserved set of characteristics, i.e. $v_i(x, y, d) + \varepsilon_{ij}$. However, this cannot be identified from a model where the distance is observed with noise, since we can always define a symmetric function $\tilde{d} : \Theta^2 \to \mathbb{R}$ such that $v_i(\xi, \xi, \tilde{d_{ij}}) = v_i(\xi, \xi, d_{ij}) + \varepsilon_{ij}$ for all $i \neq j$.[30]

Now, given (2), we can compute the probability (conditional on an observation) that a network exhibits structural homophily. Let $\Psi = 1 - \Phi$, where Φ is the c.d.f. of the standard normal distribution, and let $\gamma = \left(\beta_1/\sqrt{2}, ..., \beta_L/\sqrt{2}\right)$. The probability that a network g (given a set of characteristics θ) exhibit Structural Homophily is (algebraic manipulations can be found in appendix C) :

$$\begin{aligned} \mathbb{P}(sh|g, \theta, \gamma) &= \Pi_{ij \notin g} \left\{ \Pi_{k \in g_i} \Psi \left[(s_{ik} - s_{ij})\gamma' \right] + \Pi_{k \in g_j} \Psi \left[(s_{jk} - s_{ij})\gamma' \right] \right. \\ &\quad \left. - \Pi_{k \in g_i} \Psi \left[(s_{ik} - s_{ij})\gamma' \right] \Pi_{k \in g_j} \Psi \left[(s_{jk} - s_{ij})\gamma' \right] \right\} \end{aligned} \tag{3}$$

where s_{ij} is the $1 \times L$ vector of dimension-wise distance, i.e. $s_{ij}^l = \rho_l(\theta_i, \theta_j)$.[31]

Then, given that there are Q observations, I propose the following maximum likelihood estimator:

$$\ell(\beta|\theta) = \frac{1}{Q} \sum_{q=1}^{Q} \ln[\mathbb{P}(sh|g_q, \theta_q, \gamma)] \tag{4}$$

Provided that there exists a unique $\gamma^0 \in \Xi$ which maximizes (4), the maximum likelihood estimator is well-behaved, and γ can be consistently estimated.[32]

[29] For instance, ε_{ij} can be interpreted as a measurement error.

[30] If v_i is quasi-linear in the distance, i.e. $v(x, y, d) = f(x, y) - d$, the two models are equivalent.

[31] Equation 3 assumes that there is no isolated individual (i.e. no individual i is such that $g_i \in \{\emptyset, \{i\}\}$). This is done without loss of generality since for any pair of individuals in which one of the individual is isolated, the condition imposed by structural homophily is trivially respected.

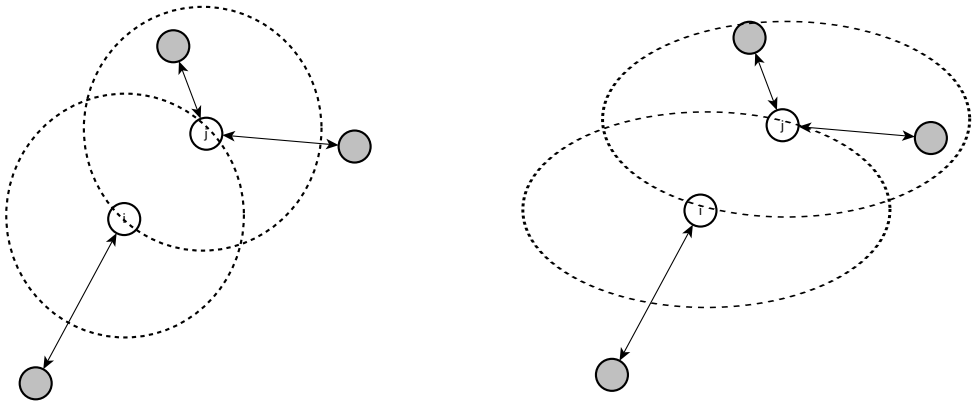
[32] Although the function in (3) looks peculiar, the MLE setting is standard and the estimation of (4) requires only usual the usual set of assumptions. See for instance Cameron and Trivedi (2005, p. 142-143) for the asymptotic properties of the maximum of likelihood estimator.

The identification's strategy is based on a link-deference approach. A link exists if no individual refused it. There are two reasons for an individual to refuse a link: (1) because he has no resources left (constraint interactions), or (2) because the other individual is too distant (preference interactions). I want to identify the preference effect, given that the resource constraint is unobserved. The estimation strategy can be viewed as to minimize the probability that structural homophily is violated.

Lets consider two alternative parameters β and β'. Suppose that we observe two individuals, i and j, not linked together, as in Figure 5. According to β and β', i is linked to an individual, farther from him than j. This means that i would have been willing to create a link with j, but that j refused. This implies that j cannot be linked to individuals farther from him than i. If he does, structural homophily is violated. Thus, if j is linked to farther individuals than i under β, but not under β', then β' is chosen over β to represent individuals' preferences.

Figure 5: Admissible Parameters, $\Theta = \mathbb{R}^2$

(a) Distance Weights according to β (b) Distance Weights according to β'



This shows why isolated individuals (i.e. individuals that have no link) provide no information: whatever the parameters' values, they never contradict structural homophily. In other words, for isolated individuals, we cannot identify whether they are isolated be-

cause they have limited resources, or because they have strong homophilic preferences. From a revealed preference approach, we gain information about an individual' preferences by observing his choices. If an individual is not connected, he does not "consume" any resource. We therefore cannot say anything about his preferences.

I now explore the properties of this method through Monte Carlo simulations.

4.1 Monte Carlo Simulations

I now present some Monte Carlo simulations. One of the advantage of section 3 is that it provides a simple algorithm allowing for the construction of the equilibrium network. Using the assignment algorithm, I will explore the finite sample properties of the estimator defined in the previous section. For simplicity and because of computational limitations, I assume that $\Theta = \mathbb{R}^2$ (this could represent, for example, the geographic position of the individuals), and $\rho_l(\theta_i, \theta_j) = |\theta_i^l - \theta_j^l|$. For all $i \in N$, I assume that $\theta_i \sim_{iid} N(\mathbf{0}, \sigma^2\mathbf{I})$. Thus, σ^2 controls for the dispersion of the individuals on the plane. As assumed, I let $\varepsilon_{ij} \sim N(0,1)$. I run 1000 replications of an economy composed of 150 independent populations (networks), each of which has 20 individuals, and I vary κ_i and σ^2 (I assume that κ_i is drawn from a uniform distribution).

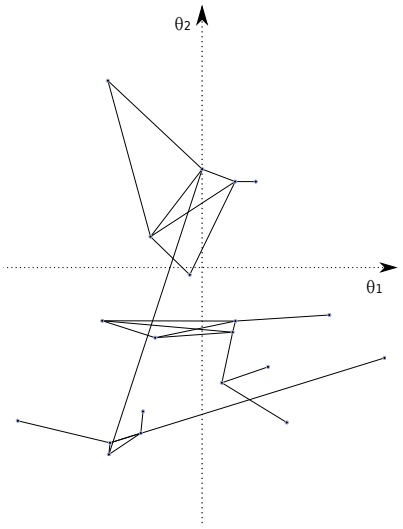
The simulated networks are generated using the assignment algorithm, assuming that $v_i(\xi, \xi, d_{ij}) > 0$ for all $i, j \in N$ and that $w_i(\xi) < 0$ for all $i \in N$. I assume that the weights are $\beta = (2, 6)$, so the distance is $d(\theta_i, \theta_i) = 2|\theta_i^1 - \theta_j^1| + 6|\theta_i^2 - \theta_j^2|$. Figure 6 displays a typical equilibrium network for this economy. Figure 6a shows the simulated network on the plane while Figure 6b rearranges the individuals in order to see clearly the network structure. Notice that the individuals value the vertical characteristic more than the horizontal one.

The small size of each observation (i.e. 20 individuals in every network) has an impact on the precision of the estimator. Take the following limiting case. Suppose that, as in the simulation framework, every link is profitable. Then, if the resource constraint is large enough, the equilibrium network is the complete network, and Structural Homophily is not binding. As a result, the model in (4) is not identified. I now explore the precision

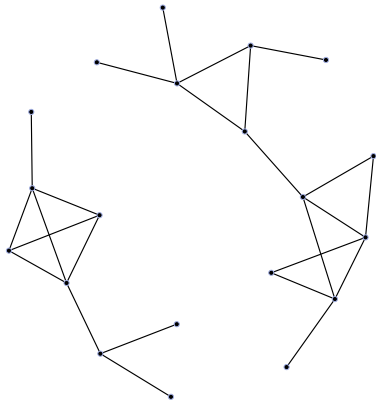
[33] Using the Kamada-Kawai algorithm is a standard way of drawing networks on the plane.

Figure 6: Typical network, with $\beta = [2\ 6]$, and $\kappa_i \sim U[1, 4]$

(a) In the type space



(b) K.K. representation[33]



of the estimator when individuals have a relatively large resource constraint, compared to the size of the population. I find that the estimator performs better when the maximal number of links is small compared to the size of the population, and that the precision of the estimator can be improved by increasing the dispersion of the population on the type space.

Table 1 and Figure 7 to 10 (Appendix D) show the simulation results. Since the parameters are only scale-identified, I report only the relative estimates. Simulations show that as the number of links increases (relative to the size of the population), the precision of the estimator is increased, but the estimates can be slightly biased upward. However, this problem vanishes as the distribution of the population over the type space increases.

I now turn to the implementation of the estimation technique. In the next section, I use the Add Health database to address the role of race in the formation of friendship networks.

Table 1: Monte Carlo Simulations

	Standard Deviation (σ)			
κ_i	10	12	14	16
$\{1,2\}$	3.031	3.024	3.02	3.01
	(0.026)	(0.028)	(0.02)	(0.02)
$\{3,4\}$	3.077	3.045	3.03	3.02
	(0.027)	(0.028)	(0.03)	(0.02)
$\{5,6\}$	3.089	3.050	3.03	3.03
	(0.029)	(0.029)	(0.03)	(0.03)
$\{7,8\}$	3.104	3.069	3.05	3.03
	(0.032)	(0.030)	(0.03)	(0.03)
$\{9,10\}$	3.107	3.081	3.05	3.04
	(0.033)	(0.030)	(0.03)	(0.03)
$\{11,12\}$	3.112	3.082	3.05	3.04
	(0.034)	(0.033)	(0.03)	(0.03)
$\{13,14\}$	3.117	3.082	3.05	3.04
	(0.044)	(0.039)	(0.04)	(0.04)
$\{15,16\}$	3.122	3.090	3.06	3.05
	(0.047)	(0.071)	(0.06)	(0.06)

5 Empirical Application: High-School Friendship Networks

I wish to estimate the weights of the distance function that leads to the formation of the friendship networks of American teenagers. I am particularly interested in the role of race, as previous studies have suggested there is a significant race-based preference bias in the choice of friendship relations among teenagers. Currarini et al. (2010) use a search model in order to estimate the preference bias for Asians, Blacks, Hispanics and Whites. They show that Asians have the largest preference bias, followed by Whites, Hispanics and Blacks. Using a different approach, Mele (2011) estimates the role that homophilic preferences toward race plays in the formation of friendship networks. He shows that all racial groups have strong homophilic preferences, although he does not capture any strong differences between groups. Interestingly, I find strong evidence that the racial preference bias varies across racial groups, although I find that Blacks have the strongest bias, followed by Asians and Whites.

As in the two papers mentioned, I use the Add Health database as it is particularly well suited for my model. Recall that the model presented in sections 2 and 3 assumes that the individuals of the same population meet with probability one. A convincing empirical implementation then requires that the observed populations are small enough. To that effect, the Add Health database provides information on students' high-schools, which are quite small entities.[34] Specifically sample includes the race, and the friendship networks of 5,466 teenagers, coming from 98 high schools in the U.S. The variable of interest is race. I assume that a student's type is his or her race. Thus the type space has 4 dimensions: White, Black, Asian, Native. Formally, $\Theta = \{0,1\}^4$, so a student who considers himself as Black-Asian would be of type $\theta = (0,1,1,0)$. I assume the following distance function:

$$\ln d(x_i, x_j) = \sum_{r=1}^{4} \beta_r \mathbb{I}_{\{x_i^r \neq x_j^r\}} + \varepsilon_{ij} \tag{5}$$

where $\mathbb{I}_{\{P\}}$ is an indicator function that takes value 1 if P is true, and 0 otherwise. For instance, the distance between a teenager i who is White, and a teenager j who is Black, is $d(x_i, x_j) = \beta_{white} + \beta_{black}$. The β's measure the relative strength of the preference bias toward individuals of the same racial group, e.g. being Black, v.s. being non-Black.

The Add Health questionnaire asks each teenager to identify their best friends (up to 10, and a maximum 5 males and 5 females). I assume that two individuals are friends only if they attend the same school. This assumption is standard in the literature using Add Health data. This allows each school (the set of teenagers and the network) to be treated as an observation. Thus, the database contains 98 observations (i.e. 98 schools). Table 2 summarizes the data:

I estimate the model (4), using the distance function in (5). The estimated weights $(\hat{\beta}_1, ..., \hat{\beta}_4)$ and the corresponding standard errors are shown in Table 3. Since the weights are only scale-identified, I report the relative effects. The estimation shows that the weight associated with the Blacks' dimension is the greatest (2.270 times greater than the Whites',

[34]For that reason, and for computational reasons, I limit myself to schools for which I observe less than 300 students, which is about 68% of the schools in the database. I also remove the isolated individuals, as they provide no relevant information (see p.18, last paragraph).

Table 2: Descriptive Statistics

Variable	Mean	Standard Deviation
White	0.733	0.442
Black	0.150	0.357
Asian/Pacific	0.031	0.174
Native	0.062	0.242
Degree	2.064	1.284

and 1.796 times greater than the Asians'). The Asians' dimension is the second in magnitude (1.264 times greater than the Whites'). I find no statistically significant relative weight for the Natives' dimension. The interpretation is that the preference bias toward same-race students is greater for Black than for Asians and Whites. Notice that this is independent of the relative proportion of each racial group in the population, and the (unobserved) individuals' time constraints.

These results show that, even if the distance function is the same for every individual, we can still represent situations where homophilic preferences differ with respect to an individual's type. In this application, for instance, everyone is weighting the dimension Black/non-Black using the same function. However, structural homophily is binding only for Blacks in that dimension of the type space. The estimated parameters can then be interpreted as relative biases.

Table 3: Relative Estimated Weights (White normalized to 1)[†]

	Black	Asian	Native
Estimate[††]	2.270**	1.264**	-0.199
SE	(0.244)	(0.157)	(0,150)
Robust SE[†††]	[0.304]	[0.294]	[0.171]

[†] S.E computed using the delta method.
[††] ** for 1% significance level.
[†††] Robust SE using the (sandwich) variance-covariance matrix for pseudo-m.l.e.

Turning back to the distance functions, one can reconstruct the distance between the different racial groups from the estimates in Table 3. Recall that, for instance, the distance between a Black and a White is $d(black, white) = \beta_{black} + \beta_{white}$. Then, according to Table

3, the distance between Blacks and Asians is the greatest ($d = 3.534$), followed by the distance between Blacks and Whites ($d = 3.270$) and the one between Whites and Asians ($d = 2.264$). This shows that, in order to correctly specify the impact of homophilic preferences on the creation of links, one has to take in to account the impact of the preference biases of both individuals involved. Structural homophily allows to identify those preference biases.

I now discuss the limitations of my approach and suggest some potential generalizations.

6 Going Further

I have shown that structural homophily can be obtained by a non-cooperative game of network formation. Under Finiteness or (strict) Convexity, any Bilateral Equilibrium of the game features structural homophily. I also have shown that structural homophily has empirical implications. I develop an estimation technique that can be used to estimate some parameters of the model, namely the weights of the distance function. I can then identify which social characteristics significantly influence the network formation process. Being able to estimate the magnitude of these relevant characteristics is an important step in the process of designing efficient policies, as it allows the policy makers to target relevant characteristics. To illustrate this method, I estimated the weights of the distance function in the context of friendship networks for teenagers. I found significant differences in the homophilic preference bias between racial groups.

The model developed in this paper is a first step toward a better understanding of network formation processes under time constraints. However, there are still many unanswered questions. For instance, the results in section 3 are based on the Finiteness or Convexity assumption. Those are arguably strong assumptions as they imply that individuals invest as much as they can in their existing links. This may not be true in general. However, the study of the model under a concavity assumption faces difficult existence issues. One could address this issue by considering weaker solution concepts such as Pairwise Stability (Jackson and Wolinsky, 1996) which potentially exhibit less structured equilibrium networks.

Another potential extension would be to introduce probabilities of meeting between individuals. Without meeting probabilities, the set of potential friends is the same for every individuals, i.e. the whole population. In general, in large population, some individuals may not know themselves, which would obviously prevent them from creating a link. A simple way to introduce meeting probabilities would be to assume that the set of potential friends is limited to individuals that have "met". Hence, individuals can only invest resources in links with individuals in a subset of the population. In that case, the (ex-post) strategy space would not be the same for every individual, but structural homophily would still hold in equilibrium. More elaborate models could however assume that meeting friends is a costly process. The individuals would then be allowed to endogenously choose the amount of resource they spend searching for friends.[35] As the estimation technique does not require the observation of the time constraints, structural homophily is likely to hold in equilibrium. However, in both extensions, the estimated parameters may not be interpreted in terms of preferences. If homophily affects the preferences *and* the random meeting process, it is unclear how those two effects can be identified.

[35]A nice example of a search model with homophilic preferences is Currarini et al. (2009).

References

Aumann, R. J. (1959) "Acceptable Points in General Cooperative n-person Games" In Contribution to the theory of game IV, Annals of Mathematical Study 40, 287-324

Bloch F. and Dutta B. (2009) "Communication Networks with Endogenous Link Strength", Games and Economic Behavior, 66(1), 39-56

Bramoullé Y., Currarini, S., Jackson, M.O., Pin P. and Rogers B. (2012) "Homophily and Long-Run Integration in Social Networks", Journal of Economic Theory,*Forthcoming*

Cameron A.C. and Trivedi P.K. (2005) "Microeconometrics, Methods and Applications", Cambridge University Press

Christakis N., Fowler J., Imbens G.W. and Kalyanaraman K. (2010) "An Empirical Model for Strategic Network Formation", *Working Paper*

Čopič J., Jackson M.O. and Kirman A. (2009) "Identifying Community Structures from Network Data via Maximum Likelihood Methods", B.E. Press Journal of Theoretical Economics Vol. 9 : Iss. 1 (Contributions), Article 30.

Currarini S., Jackson M. O., Pin P. (2009) "An Economic Model of Friendship: Homophily, Minorities, and Segregation", Econometrica, 77, 1003-1045

Currarini S., Jackson M. O., Pin P. (2010) "Identifying the roles of race-based choice and chance in high school friendship network formation," Proceedings of the National Academy of Sciences of the United States of America, 107(11): 4857-4861

Echenique F. and Fryer R.G. (2007) "A Mesure of Segregation Based on Social Interactions" The Quarterly Journal of Economics, 122(2), 441-485

Franz S., Marsili M. and Pin P. (2008) "Observed Choices and Underlying Opportunities", *Working Paper*

Galeotti A., Goyal S. and Kamphorst J. (2006) "Network Formation with Heterogenous Players", Games and Economic Behavior, 54(2), 353-373

Golub B. and Jackson M.O. (2010a) "Naive Learning in Social Networks: Convergence, Influence and the Wisdom of Crowds", the American Economic Journal: Microeconomics 2(1): 112-149

Golub B. and Jackson M.O. (2010b) "Using selection bias to explain the observed structure of Internet diffusions", Proceedings of the National Academy of Sciences, 107(24): 10833-10836

Goyal S. and Vega-Redondo F. (2007) "Structural Holes in Social Networks", Journal of Economic Theory, 137, 460-492

Henry M. and Mourifié I. (2011) "Euclidean Revealed Preferences: Testing the Spatial Voting Model", Journal of Applied Econometrics, *Forthcoming*

Iijima R. and Kamada Y. (2010) "Social Distance and Network Structures", *Working Paper*

Irving R. W. (1985) "An efficient algorithm for the "stable roommates" problem", Journal of Algorithms, 6(4), 577-595

Jackson M. O. (2008) Social and Economic Networks, Princeton University Press

Jackson M.O. and Rogers B.W. (2005) "The Economics of Small Worlds", Journal of the European Economic Association, 3(2-3), 617-627

Jackson M.O. and Rogers B.W. (2007) "Meeting Strangers and Friends of Friends: How Random are Socially Generated Networks?", American Economic Review, 97(3), 890-915

Jackson M. O. and Wolinsky (1996) "A Strategic Model of Social and Economic Networks", Journal of Economic Theory, 71, 44-74

Johnson C. and Gilles R. P. (2000) "Spatial Social Networks", Review of Economic Design, 5, 273-299

van der Leij M., Rolfe M. and Toomet O. (2009) "On the Relationship Between Unexplained Wage Gap and Social Network Connections for Ethnical Groups", *Working Paper*

Manski, C.F. (2000) "Economic Analysis of Social Interactions", The Journal of Economic Perspectives, 14(3), 115-136

Marmaros D. and Sacerdote B. (2006) "How Do Friendships Form?", The Quarterly Journal of Economics, 121(1), 79-119

Mele A. (2010) "A Structural Model of Segregation in Social Networks", *Working Paper*

McPherson M., Smith-Lovin L., and Cook M J. (2001) "Birds of a Feather: Homophily in Social Networks", Annual Review of Sociology, 27:415-444

Patacchini E. and Zenou Y. (2009) "Ethnic Networks and Employment Outcomes", *Working Paper*

Rivas J. (2009) "Friendship Selection", International Journal of Game Theory, 38, 521-538

Rubí-Barceló A. (2010) "Core/periphery scientific collaboration networks among very similar researchers", Theory and Decision, *Forthcoming*.

Watts A. (2007) "Formation of Segregated and Integrated Groups", International Journal of Game Theory, 35:505-519

Appendix A

Proof of lemma 3.1

Let x^* be some NE, and suppose that (i, j) is a deviating pair in the sense of a WBE. Let $(\tilde{x}_i, \tilde{x}_j)$ be some joint deviation for (i, j). We need to show that $\tilde{x}_i^j > x_i^{j*}$ and $\tilde{x}_j^i > x_j^{i*}$.

Since $(\tilde{x}_i, \tilde{x}_j)$ is a profitable deviation (in the sense of a WBE), we have

$$\begin{aligned} u_i(\tilde{x}_i, \tilde{x}_j, x^*_{-i-j}) &> u_i(x^*) \\ u_j(\tilde{x}_i, \tilde{x}_j, x^*_{-i-j}) &> u_j(x^*) \end{aligned} \tag{6}$$

Since x^* is a NE, we have

$$\begin{aligned} u_i(x_i, x^*_{-i}) &\leq u_i(x^*) \\ u_j(x_j, x^*_{-j}) &\leq u_j(x^*) \end{aligned} \tag{7}$$

for all x_i, and x_j. In particular, condition (7) holds for $x_i = \tilde{x}_i$ and $x_j = \tilde{x}_j$. Putting conditions (6) and (7) together, we have : $u_i(\tilde{x}_i, \tilde{x}_j, x^*_{-i-j}) > u_i(\tilde{x}_i, x^*_{-i})$ and $u_j(\tilde{x}_i, \tilde{x}_j, x^*_{-i-j}) > u_j(\tilde{x}_j, x^*_{-j})$. Since the utility function is linear in the links, this is equivalent to $v_i(\tilde{x}_i^j, \tilde{x}_j^i, d_{ij}) > v_i(\tilde{x}_i^j, x_j^{i*}, d_{ij})$ and $v_j(\tilde{x}_i^j, \tilde{x}_j^i, d_{ij}) > v_j(\tilde{x}_j^i, x_i^{j*}, d_{ij})$. The production functions are strictly increasing in the second argument, so we must have $\tilde{x}_i^j > x_i^{j*}$ and $\tilde{x}_j^i > x_j^{i*}$. (If $x_i^{j*} = x_j^{i*} = 0$, we have $v_i(\tilde{x}_i^j, \tilde{x}_j^i, d_{ij}) > 0$ and $v_j(\tilde{x}_j^i, \tilde{x}_i^j, d_{ij}) > 0$, and the result is straightforward.) $\square$

Proof of theorem 3.2

First, we show that $\tilde{x}$ produced by the assignment algorithm (see appendix B) is a NE. By construction, we have $v_i(\xi, \xi, d_{ij}) \geq 0$, and $w_i(\xi) \geq 0$, hence removing a link is never profitable. Now, the only link that an individual can unilaterally create is the individual link. Suppose that it is profitable to do so for $i \in N$. Then either $[\delta_i < \kappa_i$ and $w_i(\xi) > 0]$, or $[\delta_i = \kappa_i$ and $w_i(\xi) > \min_{j \in g_i} v_i(\xi, \xi, d_{ij})]$. By construction, both are impossible.

Now, suppose that $\tilde{x}$ is a NE, but not a WBE. That is, there exists $i, j \in N$ such that $j \notin g_i$ (from lemma 3.1, since $x_i^j \in \{0, \xi\}$) who want to deviate, i.e. create a link between them. There are 2 cases:

1. $\delta_i = \kappa_i$. Then, i needs to remove a link in order to create a new link. (Since $\tilde{x}$ is a NE, he won't remove more than one link.) Then, this implies that there exists $k \in g_i$ such that $v_i(\xi, \xi, d_{ij}) > v_i(\xi, \xi, d_{ik}) \geq 0$. This implies that $d_{ij} < d_{ik}$.

   We now turn to j. If $\delta_j = \kappa_j$, the same argument applies for j, then $v_j(\xi, \xi, d_{ij}) > v_j(\xi, \xi, d_{jl})$ for some $l \in g_j$ (and $v_i(\xi, \xi, d_{ij}) > v_i(\xi, \xi, d_{ik})$). Since we have $d_{ij} < d_{ik}$ and $d_{ij} < d_{jl}$, this contradicts the fact that $\tilde{x}$ was created by the assignment algorithm.

   If $\delta_j < \kappa_j$, j has at least ξ to invest. Together with the fact that $d_{ij} < d_{ik}$, this contradicts the fact that $\tilde{x}$ is produced by the assignment algorithm.

2. $\delta_i < k_i$ and $\delta_j < k_j$. This is impossible since, from the assignment algorithm, it implies that $v_i(\xi, \xi, d_{ij}) < 0$ or $v_j(\xi, \xi, d_{ij}) < 0$.

□

Proof of theorem 3.3

We need to show that the allocation $\tilde{x} \in X$, which is produced by the assignment algorithm (see appendix B), is a WBE of Γ.

We first show that $\tilde{x}$ is a NE. Suppose that it is not; that is, there exists $i \in N$ such that $\tilde{x}_i$ is not individually rational. Since for any $i, j \in N$, we have $x_i^j \in \{0, \xi\}$. This means that i wants to create an additional link. (Unilaterally reducing the investment in a link necessarily lowers i's payoff.) The only link that i can create on his own is the individual link. There are two cases:

1. $\tilde{x}_i^i = 0$ and $\delta_i < \kappa_i$. Then, by construction from the assignment algorithm, this implies that $w_i(\xi) < 0$. So i has no individual profitable deviation, since $w_i(\tilde{x}_i^j) < w_i(\xi)$.

2. $\tilde{x}_i^i = 0$ and $\delta_i = \kappa_i$. Then, if i has a profitable deviation, there exists $J \subseteq g_i$ such that $w_i(\sum_{j \in J} \epsilon_j) > \sum_{j \in J} \{v_i(\xi, \xi, d_{ij}) - v_i(\xi - \epsilon_j, \xi, d_{ij})\}$. That is, i is reducing his

investments in links in J in order to invest in his individual link. Let $d^* = \max_{j\in J} d_{ij}$, we have

$$\begin{aligned} w_i(\sum_{j\in J}\epsilon_j) &> \sum_{j\in J}\{v_i(\xi,\xi,d_{ij}) - v_i(\xi-\epsilon_j,\xi,d_{ij})\} \\ &\geq \sum_{j\in J}\{v_i(\xi,\xi,d^*) - v_i(\xi-\epsilon_j,\xi,d^*)\} \quad (8) \\ &\geq v_i(\xi,\xi,d^*) - v_i(\xi-\sum_{j\in J}\epsilon_j,\xi,d^*) \quad (9) \end{aligned}$$

where (8) follows from $v_{xd}(x,\xi,d) \leq 0$, and (9) follows from $v_{xx}(x,\xi,d) \geq 0$. Now, since $v_{xx}(x,\xi,d) \geq 0$, if (8) is true for $\sum_{j\in J}\epsilon_j < \xi$, it is also true for $\sum_{j\in J}\epsilon_j = \xi$, hence $w_i(\xi) > v_i(\xi,\xi,d^*)$. This contradicts the fact that $\tilde{x}$ was created by the assignment algorithm.

We still need to show that $\tilde{x}$ is a WBE. Suppose that it's not, i.e. there exists (i,j) and (x_i,x_j) such that $u_i(x_i,x_j,\tilde{x}_{-i-j}) > u_i(\tilde{x})$ and $u_j(x_j,x_i,\tilde{x}_{-i-j}) > u_j(\tilde{x})$. From the construction of $\tilde{x}$, it must be the case that i,j are such that $\tilde{x}_i^j = \tilde{x}_j^i = 0$. Again, we have 2 cases:

1. $\delta_i < \kappa_i$ and $\delta_j < \kappa_j$. This is impossible since, from the assignment algorithm, it implies that $v_i(\xi,\xi,d_{ij}) < 0$.

2. $\delta_i = \kappa_i$. Then, if i has a profitable deviation, there exists $K \subseteq g_i$ such that $v_i(\sum_{k\in K}\epsilon_k, x_j^i, d_{ij}) > \sum_{k\in K}\{v_i(\xi,\xi,d_{ik}) - v_i(\xi-\epsilon_k,\xi,d_{ik})\}$. Let $d_i^* = \max_{k\in K} d_{ik}$, we have

$$\begin{aligned} v_i(\sum_{k\in K}\epsilon_k, x_j^i, d_{ij}) &> \sum_{k\in K}\{v_i(\xi,\xi,d_{ik}) - v_i(\xi-\epsilon_k,\xi,d_{ik})\} \\ &\geq \sum_{k\in K}\{v_i(\xi,\xi,d_i^*) - v_i(\xi-\epsilon_k,\xi,d_i^*)\} \quad (10) \\ &\geq v_i(\xi,\xi,d_i^*) - v_i(\xi-\sum_{k\in K}\epsilon_j,\xi,d_i^*) \quad (11) \end{aligned}$$

where (10) follows from $v_{xd}(x,\xi,d) \leq 0$, and (11) follows from $v_{xx}(x,\xi,d) \geq 0$. Now, since $v_{xx}(x,\xi,d) \geq 0$, if (11) is true for $\sum_{k\in K}\epsilon_k < \xi$, it is also true for $\sum_{k\in K}\epsilon_k = \xi$,

hence $v_i(\xi, x_j^i, d_{ij}) > v_i(\xi, \xi, d_i^*)$.

We now turn to j. If $\delta_j = \kappa_j$, the same argument applies for j; then $v_j(\xi, \xi, d_{ij}) > v_j(\xi, \xi, d_j^*)$ (and $v_i(\xi, \xi, d_{ij}) > v_i(\xi, \xi, d_i^*)$). Since we have $d_{ij} < d_i^*$ and $d_{ij} < d_j^*$, this contradicts the fact that $\tilde{x}$ was created by the assignment algorithm.

If $\delta_j < \kappa_j$, j has at least ξ to invest (and it is profitable to invest up to ξ since $v_x(x, y, d) > 0$), then together with the fact that $d_{ij} < d_i^*$, this contradicts the fact that $\tilde{x}$ is produced by the assignment algorithm.

□

Proof of proposition 3.4

From theorem 3.3, it is sufficient to show that for any $i, j \in N$, $x_i^j \in \{0, \xi\}$, at any NE.

Consider some $i, j \in N$, and suppose that $x_i^j \in (0, \xi)$. I show that this implies that there exists $k \in N$ such that $x_i^k \in (0, \xi)$. Suppose otherwise. Then, i still has resources available. Since $v_x(x, y, d) > 0$, i could increase x_i^j and be better off. Hence, x is not a NE, so it is not a WBE. Hence, there exists $k \in N \setminus \{i\}$ such that $x_i^k \in (0, \xi)$. There are 2 cases:

1. $[k = i]$. Since x is a NE, we must have the following.

   - If $x_i^i + x_i^j \geq \xi$, then

$$w_i(x_i^i) + v_i(x_i^j, x_j^i, d_{ij}) \geq w_i(\xi) + v_i(x_i^j + x_i^i - \xi, x_j^i, d_{ij})$$
$$w_i(x_i^i) + v_i(x_i^j, x_j^i, d_{ij}) \geq w_i(x_i^j + x_i^i - \xi) + v_i(\xi, x_j^i, d_{ij})$$

     Rewriting, we have

$$w_i(\xi) - w_i(x_i^i) \leq v_i(x_i^j, x_j^i, d_{ij}) - v_i(x_i^j + x_i^i - \xi, x_j^i, d_{ij})$$
$$w_i(x_i^i) - w_i(x_i^j + x_i^i - \xi) \geq v_i(\xi, x_j^i, d_{ij}) - v_i(x_i^j, x_j^i, d_{ij})$$

Since $v_{xx}(x,y,d) > 0$, we have $v_i(\xi, x_j^i, d_{ij}) - v_i(x_i^j, x_j^i, d_{ij}) > v_i(x_i^j, x_j^i, d_{ij}) - v_i(x_i^j + x_i^i - \xi, x_j^i, d_{ij})$, and since $w''(x) > 0$, we have $w_i(\xi) - w_i(x_i^i) > w_i(x_i^i) - w_i(x_i^j + x_i^i - \xi)$. This is in contradiction with the above conditions, hence x is not a NE.

- If $x_i^i + x_i^j < \xi$, then

$$w_i(x_i^i) + v_i(x_i^j, x_j^i, d_{ij}) \geq w_i(x_i^i + x_i^j) + v_i(0, x_j^i, d_{ij})$$
$$w_i(x_i^i) + v_i(x_i^j, x_j^i, d_{ij}) \geq w_i(0) + v_i(x_i^i + x_i^j, x_j^i, d_{ij})$$

Rewriting, we have

$$w_i(x_i^i + x_i^j) - w_i(x_i^i) \leq v_i(x_i^j, x_j^i, d_{ij}) - v_i(0, x_j^i, d_{ij})$$
$$w_i(x_i^i) - w_i(0) \geq v_i(x_i^i + x_i^j, x_j^i, d_{ij}) - v_i(x_i^j, x_j^i, d_{ij})$$

Since $v_{xx}(x,y,d) > 0$, we have $v_i(x_i^j + x_i^i, x_j^i, d_{ij}) - v_i(x_i^j, x_j^i, d_{ij}) > v_i(x_i^j, x_j^i, d_{ij}) - v_i(0, x_j^i, d_{ij})$, and since $w''(x) > 0$, we have $w_i(x_i^j + x_i^i) - w_i(x_i^i) > w_i(x_i^i) - w_i(0)$. Again, this is in contradiction with the above conditions, hence x is not a NE.

$i \neq k$ and $i \neq j$.

Since x is a NE, we must have the following:

- If $x_i^k + x_i^j \geq \xi$, then

$$v_i(x_i^k, x_k^i, d_{ik}) + v_i(x_i^j, x_j^i, d_{ij}) \geq v_i(\xi, x_k^i, d_{ik}) + v_i(x_i^j + x_i^k - \xi, x_j^i, d_{ij})$$
$$v_i(x_i^k, x_k^i, d_{ik}) + v_i(x_i^j, x_j^i, d_{ij}) \geq v_i(x_i^j + x_i^k - \xi, x_k^i, d_{ik}) + v_i(\xi, x_j^i, d_{ij})$$

Rewriting, we have

$$v_i(\xi, x_k^i, d_{ik}) - v_i(x_i^k, x_k^i, d_{ik}) \leq v_i(x_i^j, x_j^i, d_{ij}) - v_i(x_i^j + x_i^k - \xi, x_j^i, d_{ij})$$
$$v_i(x_i^k, x_k^i, d_{ik}) - v_i(x_i^j + x_i^k - \xi, x_k^i, d_{ik}) \geq v_i(\xi, x_j^i, d_{ij}) - v_i(x_i^j, x_j^i, d_{ij})$$

Since $v_{xx}(x,y,d) > 0$, we have $v_i(\xi, x_j^i, d_{ij}) - v_i(x_i^j, x_j^i, d_{ij}) > v_i(x_i^j, x_j^i, d_{ij}) - v_i(x_i^j + x_i^k - \xi, x_j^i, d_{ij})$, and $v_i(\xi, x_k^i, d_{ik}) - v_i(x_i^k, x_k^i, d_{ik}) > v_i(x_i^k, x_k^i, d_{ik}) - v_i(x_i^j + x_i^k - \xi, x_k^i, d_{ik})$. This is in contradiction with the above conditions, hence x is not a NE.

- If $x_i^i + x_i^j < \xi$, then

$$v_i(x_i^k, x_k^i, d_{ik}) + v_i(x_i^j, x_j^i, d_{ij}) \geq v_i(x_i^j + x_i^k, x_k^i, d_{ik}) + v_i(0, x_j^i, d_{ij})$$
$$v_i(x_i^k, x_k^i, d_{ik}) + v_i(x_i^j, x_j^i, d_{ij}) \geq v_i(0, x_k^i, d_{ik}) + v_i(x_i^j + x_i^k, x_j^i, d_{ij})$$

Rewriting, we have

$$v_i(x_i^j + x_i^k, x_k^i, d_{ik}) - v_i(x_i^k, x_k^i, d_{ik}) \leq v_i(x_i^j, x_j^i, d_{ij}) - v_i(0, x_j^i, d_{ij})$$
$$v_i(x_i^k, x_k^i, d_{ik}) - v_i(0, x_k^i, d_{ik}) \geq v_i(x_i^j + x_i^k, x_j^i, d_{ij}) - v_i(x_i^j, x_j^i, d_{ij})$$

Since $v_{xx}(x,y,d) > 0$, we have $v_i(x_i^j + x_i^k, x_j^i, d_{ij}) - v_i(x_i^j, x_j^i, d_{ij}) > v_i(x_i^j, x_j^i, d_{ij}) - v_i(0, x_j^i, d_{ij})$, and $v_i(x_i^j + x_i^k, x_k^i, d_{ik}) - v_i(x_i^k, x_k^i, d_{ik}) > v_i(x_i^k, x_k^i, d_{ik}) - v_i(0, x_k^i, d_{ik})$. This is in contradiction with the above conditions, hence x is not a NE.

□

Proof of proposition 3.5

The proof is obvious from the proof of theorem 3.2 and theorem 3.3. One only has to remark that for any $i, j, k \in N$, $v_i(\xi, \xi, d_{ij}) \geq v_i(\xi, \xi, d_{ik})$ implies that $v_i(\xi, \xi, d_{ij}) > v_i(\xi, \xi, d_{ik})$ if we assume that $d_{ij} \neq d_{kl}$. □

Proof of proposition 3.7

The fact that any Strong NE needs to be produced by the assignment algorithm follows from propositions 3.2 and 3.4. Suppose that $x^* \in X$ is a BE, but not a Strong NE. There exists $S \subset N$ and $x_S \in \times_{i \in S} X_i$ such that $u_i(x_S, x^*_{-S}) > u_i(x^*)$ for all $i \in S$. We will

show that under Strict Convexity or Finiteness, this implies that there exists a bilateral deviation.

Under Finiteness, $x_i \in \{0, \xi\}^n$ for all $i \in S$. Using the same argument as the one used in lemma 3.1, there exist at least one project created under a deviation by coalition S. That is, $\exists i, j \in S$, such that $x_i^{j*} = x_j^{i*} = 0$ and $x_i^j = x_j^i = \xi$. Since the utility functions are additive, this implies that i, j have a profitable bilateral deviation. Since resources invested in the link (i, j) must have come either from unused resources or from the deleation of another link since $x_i^j \in \{0, \xi\}$ for all $i, j \in N$.

Under Convexity, if it is profitable to withdraw resources from one link and invest in two new links, it is even better to invest in only one of those links. (This is exactly the argument used in proposition 3.3). Specifically, suppose that there exists $i, j, k \in S$ such that $x_i^j, x_i^k > 0$, and $x_i^{j*} = x_i^{k*} = 0$. Then, either $x_i^j = \xi$ and $x_i^k = 0$ or $x_i^j = 0$ and $x_i^k = \xi$ is better for i. Then, i is willing to make a bilateral deviation with j (wlog). Since the utilities are linear, it is also profitable for k (since it is under a joint deviation in S). Hence, there exists a bilateral deviation between i and j. $\square$

Appendix B

The Assignment Algorithm

I generate a network g (represented by the adjacency matrix A) in which every individual invests as much as possible in every active link (i.e. $x_i^j \in \{0, \xi_i\}$ for all $i, j \in N$).

Let $\eta_i^j = v_i(\xi, \xi, d_{ij})$ for all $i, j \in N$ such that $i \neq j$, and $\eta_i^i = w_i(\xi)$, for all $i \in N$. This function represents the value of a link between two individuals. Now, define the (not necessarily unique) ordered list L^0 as follows: $L^0 = (d_{ij})_{i,j \in N: i<j}$, such that $L_1^0 \leq L_2^0 \leq ... \leq L_m^0$. The list L^0 is an ordered list of distance values, for all pairs of individuals. The number of elements in L^0 is the number of possible pairings between individuals in N, i.e. $n(n-1)/2$. Let L_l^0 be the element of position l in the list L^0. I note $(L_l^0)^{-1} = (i, j)$ if $L_l^0 = d_{ij}$.

The algorithm computes g and takes $L^t = L^0$ as inputs. It operates in two steps.

1 *Take the first element of the list L^t, i.e. L_1^t. Let $L_1^t = d_{ij}$.*
If $a_{ii} = 0$ or $a_{jj} = 0$,

1. *If $\eta_i^i \geq \eta_i^j$ and $\eta_i^i \geq 0$, then $a_{ii} = 1$*
2. *If $\eta_j^j \geq \eta_j^i$ and $\eta_j^j \geq 0$, then $a_{jj} = 1$*

Otherwise,

1. *If $\eta_i^j \geq 0$ and $\eta_j^i \geq 0$, then set $a_{ij} = a_{ji} = 1$.*
2. *If $\eta_i^j < 0$, then generate $L^{*i} = L^t \setminus \{d_{ik}\}_{k \in N: d_{ik} \in L^t}$. (That is, remove all distances associates with i, since all the following distances will be greater than d_{ij}.)*
3. *If $\eta_j^i < 0$, then generate $L^{*i} = L^t \setminus \{d_{jk}\}_{k \in N: d_{jk} \in L^t}$, i.e. do the same for j as we did for i.*

Generate $L^{t+1} = \{(d \in L^{i*} \cap L^{j*}) \setminus d_{ij}\}$.

2 *Repeat (1) for* $t = 1, ...$ *until* $|L^t| = 0$ *or until* $\exists i \in N$ *such that* $\delta_i = \kappa_i$.

For all $i \in N$ *such that* $\delta_i = \kappa_i$*, generate* $L^{*i} = L^t \setminus \{d_{ik}\}_{k \in N: d_{ik} \in L^t}$*. (That is, remove all distances associated with* i*, since he has no resources left.) Then, generate* $L^{t+1} = \cap_{i \in N} L^{i*}$ *and repeat (1).*

After the algorithm stops, I generate the allocation $\tilde{x}$ as follows. For all $i, j \in N$, if $a_{ij} = 1$, $\tilde{x}_i^j = \xi$, otherwise $\tilde{x}_i^j = 0$. Notice that by definition $\tilde{x} \in X$.

Appendix C

The Likelihood Function

I assume that no individual is isolated. The definition of structural homophily is: For all $ij \notin g$, $d_{ij} \geq d_{ik}$ for all $k \in g_i$ or $d_{ij} \geq d_{jk}$ for all $k \in g_j$. Then, since the ε_{ij} are independents, and $ln(d) \geq ln(d')$ iff $d \geq d'$, the probability that g exhibits structural homophily is

$$\Pi_{ij \notin g} \left\{ \Pi_{k \in g_i} \mathbb{P}(d_{ij} \geq d_{ik}) + \Pi_{k \in g_j} \mathbb{P}(d_{ij} \geq d_{jk}) - \Pi_{k \in g_i} \mathbb{P}(d_{ij} \geq d_{ik}) \Pi_{k \in g_j} \mathbb{P}(d_{ij} \geq d_{jk}) \right\}$$

This gives:

$$\mathbb{P}(d_{ij} \geq d_{ik}) = \mathbb{P}(\sum_{r=1}^{R} \beta_r \rho_r(\theta_i, \theta_j) + \varepsilon_{ij} \geq \sum_{r=1}^{R} \beta_r \rho_r(\theta_i, \theta_k) + \varepsilon_{ik})$$

At this point, the normalization of ε is necessary for the identification of β. Simplifying the last expression, we have:

$$\begin{aligned} \mathbb{P}(d_{ij} \geq d_{ik}) &= \mathbb{P}(Z \geq \sum_{r=1}^{R} \beta_r [\rho_r(\theta_i, \theta_k) - \rho_r(\theta_i, \theta_j)]) \\ &= 1 - \Phi(\sum_{r=1}^{R} \beta_r [\rho_r(\theta_i, \theta_k) - \rho_r(\theta_i, \theta_j)]) \end{aligned}$$

Appendix D

Figure 7: Standard deviation: 10

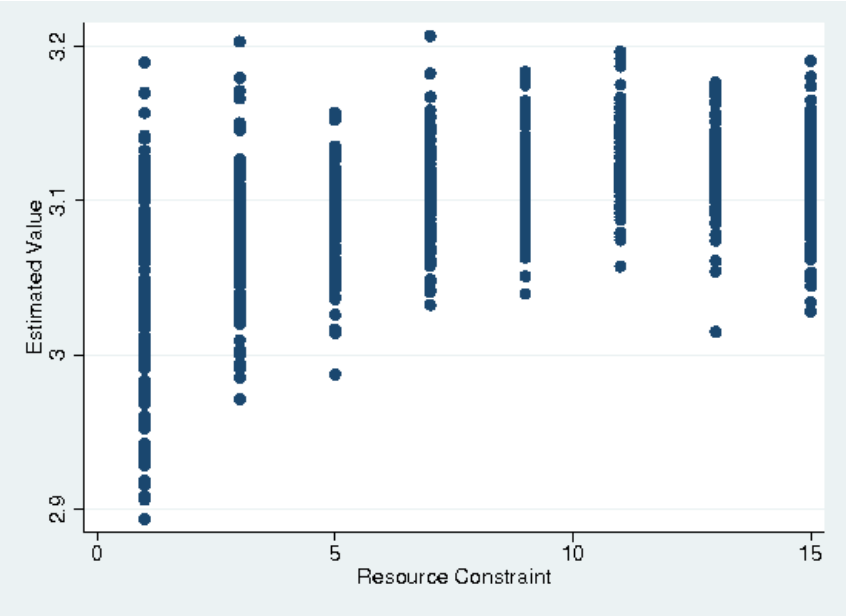


Figure 8: Standard deviation: 12

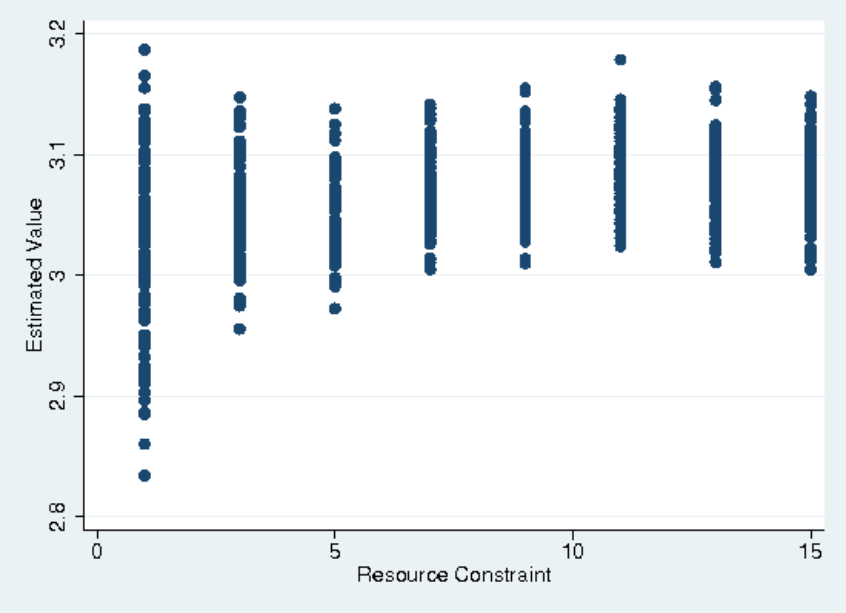


Figure 9: Standard deviation: 14

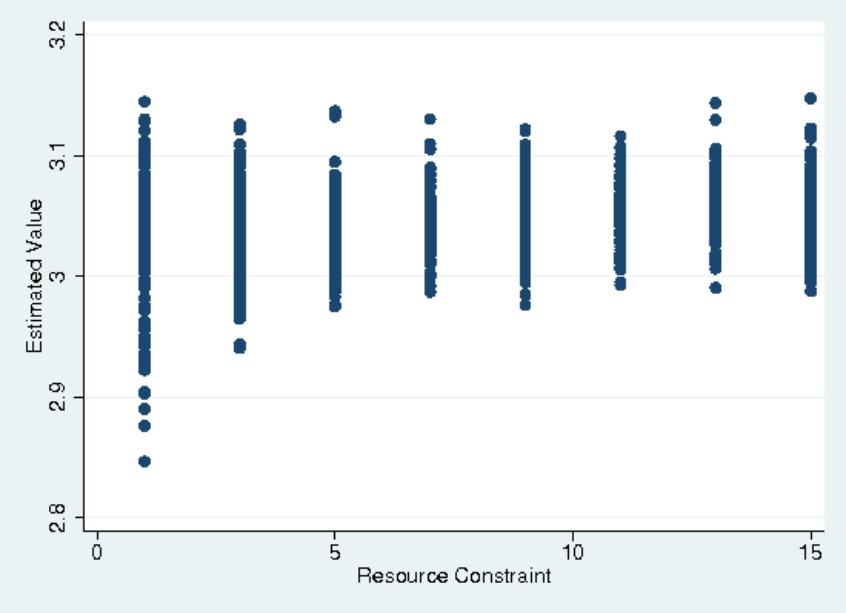


Figure 10: Standard deviation: 16