2-1-2013

Discrete Games in Endogenous Networks: Theory and Policy

Anton Badev
University of Pennsylvania, anton.badev@gmail.com

Badev, Anton. 2013. "Discrete Games in Endogenous Networks: Theory and Policy." PSC Working Paper Series, PSC 13-05.

Discrete Games in Endogenous Networks: Theory and Policy

Abstract
This paper develops a framework for analyzing individuals' choices in the presence of endogenous social networks and implements it with data on teen smoking decisions and friendship networks. By allowing actions and friendships to be jointly chosen, the framework extends the literature on social interactions, which either models choices, taking the social network as given, or which models friendship selection without incorporating additional choices. In the context of a large population network game, this paper also introduces the notion of k-player Nash stability. This solution concept subsumes the Nash equilibrium and, as k decreases, gradually relaxes the assumptions of rationality and coordination underlying the Nash play. I show how the strategic interactions of the static one-shot play are embedded in an evolutionary model of network formation, which I estimate with social network data from United States high schools. The empirical analysis demonstrates the importance of modeling the joint decisions of friendships and smoking in evaluating existing and proposed new policies targeting teen smoking prevalence. These include policies related to school racial desegregation, separating middle and high school grades, and anti-smoking campaigns. Neglecting the endogeneity of the friendship network leads to a downward bias of 10% to 15% on the predicted effect of these policies on adolescent smoking rates.

Keywords
Adolescnet smoking, Discrete games, Multiplicity, Social networks

Disciplines
Demography, Population, and Ecology | Social and Behavioral Sciences | Sociology

Comments
Badev, Anton. 2013. "Discrete Games in Endogenous Networks: Theory and Policy." PSC Working Paper Series, PSC 13-05.

Discrete Games in Endogenous Networks: Theory and Policy*

Anton Badev†

First version: 6/1/2012
This version: 2/1/2013

Abstract. This paper develops a framework for analyzing individuals' choices in the presence of endogenous social networks and implements it with data on teen smoking decisions and friendship networks. By allowing actions and friendships to be jointly chosen, the framework extends the literature on social interactions, which either models choices, taking the social network as given, or which models friendship selection without incorporating additional choices. In the context of a large population network game, this paper also introduces the notion of k-player Nash stability. This solution concept subsumes the Nash equilibrium and, as k decreases, gradually relaxes the assumptions of rationality and coordination underlying the Nash play. I show how the strategic interactions of the static one-shot play are embedded in an evolutionary model of network formation, which I estimate with social network data from United States high schools. The empirical analysis demonstrates the importance of modeling the joint decisions of friendships and smoking in evaluating existing and proposed new policies targeting teen smoking prevalence. These include policies related to school racial desegregation, separating middle and high school grades, and anti-smoking campaigns. Neglecting the endogeneity of the friendship network leads to a downward bias of 10% to 15% on the predicted effect of these policies on adolescent smoking rates.

JEL Codes: D85, C73, L19
Keywords: Social Networks, Adolescent Smoking, Multiplicity, Discrete Games.

---

*I thank Kenneth Wolpin, Petra Todd, and George Mailath for continuous guidance throughout this project. I have benefited from discussions with Antonio Merlo, Steven Durlauf, Holger Sieg, Hanming Fang, Angelo Mele, James Heckman, Flavio Cunha, Katja Seim, Ali Jadbabaie, and Michael Kearns. I also thank the seminar participants at the 2012 North American Summer Meeting of the Econometric Society, the 2012 Summer School on Socioeconomic Inequality, the 2012 Extreme Science and Engineering Discovery Environment (XSEDE) conference, the 10th Annual Quantitative Marketing and Economics Conference, and the Penn Empirical Micro and Theory Clubs. I gratefully acknowledge financial support from the TRIO (PARC/Boettner/NICHD) Pilot Project Competition. This work used the XSEDE, which is supported by National Science Foundation grant number OCI-1053575. All errors are mine.
†Department of Economics, University of Pennsylvania. E-mail: anton.badev@gmail.com.

1 Introduction

This paper develops a framework for analyzing individuals' choices in the presence of endogenous social networks and implements it with data on teen smoking decisions and friendship networks. By allowing actions and friendships to be jointly chosen, the framework extends the literature on social interactions, which either models choices, taking the social network as given, or which models friendship selection without incorporating additional choices. In the context of a large population network game, this paper also introduces the notion of k-player Nash stability. This solution concept subsumes the Nash equilibrium and, as k decreases, gradually relaxes the assumptions of rationality and coordination underlying the Nash play. I show how the strategic interactions of the static one-shot play are embedded in an evolutionary model of network formation, which I estimate with social network data from United States high schools. The empirical analysis demonstrates the importance of modeling the joint decisions of friendships and smoking in evaluating existing and proposed new policies targeting teen smoking prevalence. These include policies related to school racial desegregation, separating middle and high school grades, and anti-smoking campaigns. Neglecting the endogeneity of the friendship network leads to a downward bias of 10% to 15% on the predicted effect of these policies on adolescent smoking rates.

Arguably, human beings are both products and creators of their social environments. In particular, humans engage in a complex system of social relations, also known as social interactions, which determine their social environment, influence their socio-economic choices, and affect the market outcomes. While a large body of social science research in the past few decades has contributed to the understanding of how social interactions shape individuals' behavior, social networks present powerful mechanisms to achieve policy-related goals that have remained largely unexplored.

The issues of identification and estimation of social interaction models have been analyzed in a variety of settings, including models of local interactions (i.e., only the effects of immediate friendships are accounted for) and/or global interactions (only the effects of average intra-group behavior are accounted for).[1] As result, the main identification challenges on the empirical work with social interaction models are well understood. In particular, the correlation between an individual's choices and those of her peers may work through three channels (Manski, 1993). An individual's choices may be directly influenced by the choices of her peers (endogenous effects) or by characteristics of her peers (contextual effects). In addition, the correlation between individuals' choices in a group may be due to common unobservable group-level factors (correlated

[1] See Brock and Durlauf (2001, 2007); Krauth (2005b); the more recent work of Bramoullé, Djebbari and Fortin (2009); Calvó-Armengol, Patacchini and Zenou (2009); De Giorgi, Pellizzari and Redaelli (2010); and the survey in Blume, Brock, Durlauf and Ioannides (2010).

effects). In this paper, I argue that to fully understand the nature of social interactions and address the issues of identifying self-selection into peer networks, we must model the co-evolution of social networks and individual behavior.

The model has the following basic features. Individuals' payoffs from smoking and friendships depend on the decisions of friends as well as on the decisions of the overall population. For a population of size n, I define the notion of k-player Nash stability as follows: In a k-player Nash stable network state, any subset of k players is in a Nash equilibrium of the game between them when only the friendships between the k players are decided together with their action statuses. Consequently, in a k-player Nash stable network state, no player has an incentive to alter simultaneously $k-1$ of his friendships and his action. For $k < n$, this equilibrium notion imposes less rationality than Nash equilibrium play (i.e., when $k = n$) and is, therefore, a more tenable assumption in large population games. However, k-player Nash stability suffers well known deficiencies, on which Kandori, Mailath and Rob (1993) remark:

> While the Nash equilibrium concept has been used extensively in many diverse contexts, game theory has been unsuccessful in explaining *how* players know that a Nash equilibrium will be played. Moreover, the traditional theory is silent on how players know *which* Nash equilibrium is to be played if a game has multiple equally plausible Nash equilibria.

To address the above concerns I develop an "evolutionary" model of network formation, motivated by a long line of theoretical work in evolutionary game theory.[2] In the model the network configuration arises as the outcome of agents' optimal behavior over time rather than as the instantaneous outcome of a one-shot game with many players. More specifically, in every period an individual meets $k-1$ potential friends and decides whether or not to befriend each of them and whether to revise her action status. I demonstrate that this decentralized adaptive behavior results in a k-player Nash stable state being reached in finite time. In the presence of transient preference shocks, however, the k-player meeting process induces a stationary distribution on the set of all possible network states. I show that this distribution is invariant to the choice of k if the preference shocks follow the Gumbel distribution[3]. This invariance warrants the use of the stationary distribution as a behaviorally grounded statistical device to probabilistically rank equilibria in the family of k-player Nash stable states.

It is important to emphasize that the purpose of this paper is to empirically study the strategic interactions that prevail in social networks from a new perspective. More specifically,

[2]See the seminal papers of Foster and Young (1990), Kandori et al. (1993), and Blume (1993). Some of the ideas I exploit are encountered in Cournot (1838, Chapter VII).

[3]Also known as logit, extreme value type I, and log-Weibull distribution. A brief historical account of the development of the relevant statistical theory can be found in Gumbel (1958, Chapter I).

I show how the strategic incentives of the static one-shot play are embedded in a behavioral model where the network evolves over time. In this model, a k-player Nash equilibrium receives the highest probability among all the states in its immediate neighborhood. Furthermore, the model naturally delivers a coherent statistical structure suitable for empirical implementation of the proposed framework. In an application, I use this feature to carry out a likelihood-based estimation with data on adolescent smoking and friendships.

Direct estimation of the model through maximum likelihood is computationally unfeasible.[4] Instead, I approximate the likelihood using Markov chain Monte Carlo methods introduced by Geyer and Thompson (1992). The model is estimated with data on smoking behavior, friendship networks, and home environment (parental education background and parental smoking behavior) from the National Longitudinal Study of Adolescent Health. This is a longitudinal study of a nationally representative sample of adolescents in the United States, who were in grades 7–12 during the 1994–95 school year.

The estimation results indicate a substantial role for the social environment in shaping adolescents' smoking decisions. In particular, their conditional probability of smoking increases with an additional smoking friend by 7.2 percentage points (ppt) for 7th and 8th graders, and 6.1 ppt for those in grades 9 through 12. In addition, students in schools where 50% of the population smokes (independently of their friendship statuses) are 15.2 ppt more likely to smoke than students in schools with no smokers at all. The estimated marginal effects of various dimensions of the home environment are substantial as well. The presence of a smoker in the household increases (on average) the likelihood of an adolescent smoking by 18.5 ppt. If the pupil's mother has completed high school or college, this likelihood falls (on average) by 5.7 or 7.9 ppt respectively.

I use the estimated model to perform a number of counterfactual experiments. First, I analyze how changes in the racial composition of schools affect adolescent smoking. Simulations from the model suggest that redistributing students from racially segregated schools into racially balanced schools decreases the overall smoking prevalence. Second, I use the model to document similar effects on the overall smoking when students are segregated by grades within schools (e.g., middle school and high school students are separated), which prevents friendship links between younger and older students. Third, I examine the possibility of cascade effects from anti-smoking campaigns, and find that even if substantial percentage of students are not directly exposed to the campaign, it can still have a large impact on the overall prevalence of smoking. Fourth, I examine the elasticity of adolescent smoking to varying levels of changes in tobacco prices. Finally, I compare the predictions made by the model when friendship networks are

[4] Evaluation of the likelihood for a given set of parameters requires the calculation of an intractable normalizing constant–a summation with 2^{n^2} terms. For networks of size 10, this summation involves 2^{100} terms.

allowed to adjust to the proposed interventions against the model's predictions when friendship networks are not allowed to adjust. The simulations suggest that neglecting the endogeneity of the friendship network leads to a downward bias of 10% to 15% on the predicted effect of these policies on adolescent smoking rates.

The literature on social interactions can be grouped into two broad classes of studies. One class of studies develops social interaction models that take the network structure as given. These range from models that directly relate an individual's choices to mean characteristics of his peer groups[5] to models with elaborate equilibrium micro-foundations, such as those in Brock and Durlauf (2001, 2007); Krauth (2005b); Calvó-Armengol, Patacchini and Zenou (2009); Giorgi and Pellizzari (2011); Bisin, Moro and Topa (2011). Another large literature, mostly theoretical, focuses on the strategic interactions behind network formation abstracting from any other individual choices. Jackson and Wolinsky (1996); Bala and Goyal (2000); Dutta, Ghosal and Ray (2005); and Mele (2010) are prominent papers in this literature.[6] Depending on the specific context, conditioning the model on the peer network or focusing only on network formation may be appropriate, but to study adolescent smoking and high school friendship networks it is necessary to model these two decisions jointly.

A handful of theoretical papers consider both network formation along with other choices potentially affected by the network (See Jackson and Watts, 2002; Goyal and Vega-Redondo, 2005; Staudigl, 2011, 2012; Köenig et al., 2012). The theoretical frameworks available, however, are meant to provide focused insights into isolated features of networks and deliver sharp predictions, but are not easily adapted for the purposes of estimation.[7] In particular, the prevailing perspective has been to use stochastic dynamics, where uncertainty vanishes as time goes by, as an equilibrium selection device, and to study conditions under which the selected (stochastically stable) states are unique, risk dominant, or efficient. As opposed to obtaining a sharper prediction, my approach is to use the uncertainty of a behavioral model of decision making to probabilistically rank the possible network states. In addition, I examine the behavioral dynamics of the network in relation to a broader class of equilibria of the one-shot play.

[5]See Powell et al. (2005), Ali and Dwyer (2009), and the references therein. A reduced form approach to social interactions suffers severe identification issues (Manski, 1993). Moreover, the empirical analysis from such studies often is not generalizable to changes in the policy environment. See the next footnote.

[6]I abstract from a large statistical literature which approaches the network from a purely stochastic view, e.g., in their settings, the network links are formed and dissolved at random according to some probability distribution (See Reka and Barabási, 2002). A similar perspective is shared by Goldsmith-Pinkham and Imbens (2011), and Hsieh and Lee (2011), who propose econometric models of networks and actions. Overall these models lack the equilibrium microfoundation of the strategic models of network formation, limiting their scope for policy analysis.

[7]Köenig, Tessone and Zenou (2012) presents an interesting exception. They analyze a dynamic model of network formation with endogenous actions which is capable of generating the so-called nested split graph. In addition they obtain analytic characterizations of network statistics. Their model, however, is not directly applicable to the analysis of social interactions in that it does not accommodate heterogeneity of players and takes the link destruction process exogenously.

Typically models of social interactions consider a situation where agents' payoffs (and hence optimal decisions) depend on the actions of other agents. Such environments have been formalized as discrete games—an extension of discrete choice models allowing for strategic interdependence of agents' payoffs.[8] Multiplicity is a common problem in discrete games of complete information, and extant approaches to deal with it include: modeling the outcomes that are unique even in the presence of multiplicity (Bresnahan and Reiss, 1990; Berry, 1992), imposing an order of moves (Berry, 1992), assuming an ad-hoc equilibrium selection rule (Hartmann, 2010), obtaining the identified set estimates (Tamer, 2003; Ciliberto and Tamer, 2009), empirically estimating an equilibrium selection rule (Bajari, Hong and Ryan, 2010), and conditioning on the observed equilibrium (i.e. being agnostic about the equilibrium selection mechanism; Bisin et al. (2011)).[9] In contrast, this paper proposes a behavioral model of network formation which can address the indeterminacy of the static framework, and yet embeds the strategic incentives from the one-shot play. In particular, the model delivers a probabilistic ranking of the equilibria in the family of k-player Nash stable states.

Finally, this paper is related to a large body of empirical work on social interactions and risky teen behaviors. Of this literature, the closest to this work is Nakajima (2007), which estimates a model of smoking decisions in an exogenously-given peer groups. In his setting, individuals sequentially update their smoking decisions, taking into account the overall smoking prevalence in the peer group they belong to. Another related work, albeit from a different point of view, is Mele (2010). He developes a model of network formation in which individuals update their best friends over time, and proposes an implementation strategy via Bayesian methods. Lastly, there is a literature which uses laboratory experiments and quasi-experiments, such as exogenous assignments to peer groups, to document the presence of social effects on a wide range of behaviors.[10] However, there are well known limitations on the interpretation of these estimates as local treatment effects (Heckman and Vytlacil, 2005). Moreover, these frameworks are not generalizable to peer effects outside the specific context of the experiment.[11]

[8]Discrete games have been used in a variety of contexts including firms' market entry decisions (Bresnahan and Reiss, 1990; Berry, 1992), firms' product quality choice (Mazzeo, 2002), and individuals' decisions in the presence of peer effects (Brock and Durlauf, 2001; Bisin, Moro and Topa, 2011) among others.

[9]In a recent work, Narayanan (2011) takes a Bayesian perspective to the problem of multiplicity. There, the prior reflects the analyst's uncertainty about equilibrium selection. His approach goes beyond estimating the identified set in that the posterior may localize high density regions within the interval estimates and thus provide more information about the parameters.

[10]See Sacerdote (2001) and the review in Durlauf and Ioannides (2010).

[11]See Durlauf and Ioannides (2010) for discussion.

Background on tobacco smoking

Tobacco is the single greatest preventable cause of death in the world today.[12] In the United States alone, cigarette smoking causes approximately 443,000 deaths each year (accounting for one in every five deaths) and imposes an economic burden of more than $193 billion a year in health care costs and loss of productivity. Approximately 1 million young people under 18 years of age start smoking each year; about 80% of adults who are smokers started smoking before they were 18 (Kessler et al., 1996; Liang et al., 2001). Despite an overall decline in smoking prevalence from 2005 to 2010, when the percentage of current smokers decreased from 20.9% to 19.3%, the reduction in teen smoking has been less pronounced. In fact, the proportions of 8th and 10th graders who smoke increased slightly in 2010. As with many human behaviors, social interactions (peer influence) have often been pointed to as a major driving force behind adolescent smoking choices.

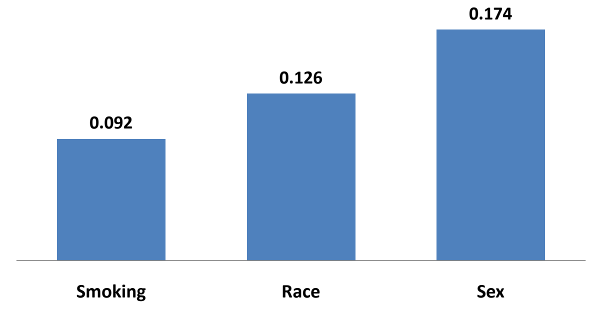
Figure 1 illustrates the relationship between friendships and smoking in the data sample used for the estimation. The left panel displays the smoking mixing matrix, which groups friendship nominations with respect to the smoking status of the nominator and the nominee. For example, the top left number in the table, 304, is the number of friendships in which a smoker nominates a smoker. On the other side, 418 is the number of friendship nominations in which a smoker nominates a non-smoker. If friendships were drawn at random then smokers and non-smokers would be nominated in proportion reflecting the size of the two groups in the sample. Thus smokers are biased in their nominations, because 42% of their friends smoke, while only 21% of the sample does so. To understand the extent to which the magnitude of this correlation is unusual, the right panel of Figure 1 compares friendship patterns across different socio-demographic groups. In particular, the diagram plots the Freeman segregation indexes (FSI) for sex, race, and smoking. A higher value implies a low likelihood that two individuals from different groups are friends. A value of 1 implies no friendships between the different groups. Clearly, the segregation behavior of individuals with respect to smoking is comparable to that of race and gender.

The rest of the paper is organized as follows. Section 2 develops the theoretical framework, which Section 3 extends to an estimable empirical model. Section 4 presents the data, the estimation method, and the parametrization of the model. In Section 5, I discuss the empirical results together with the model fit. Section 6 concludes.

[12] The World Health Organization, *Report on the Global Tobacco Epidemic* (2008). The statistics for the U.S. are compiled from reports by the Surgeon General (2010), National Center for Health Statistics (2011), and Monitoring the Future (2011).

Figure 1: Smoking and Friendships

		Nominee	
		Smoker	Nonsmoker
Nominator	Smoker	**42% (304)**	58% (418)
	Nonsmoker	16% (499)	**84% (2562)**



Source: The National Longitudinal Study of Adolescent Health (Add Health) - Wave I, 1994-95 school year (Estimation sample: 14 schools, $1,125$ students, 21% smokers).

2 A Model of Friends Selection and Individual Actions

This section presents a theoretical model of social interactions, in which individuals make both decisions regarding their friendship relations and decisions regarding their actions (smoking being the example of the empirical application). I develop a model, introduce a class of equilibria, and then discuss the existence and characterization of these equilibria. Finally, I consider a sequential decision process where every period an individual adapts her strategy to the current state of the network, and show that this process is absorbed in an equilibrium. Although the model in this section cannot be estimated (it features no uncertainty and produces a sharp prediction with respect to the observed outcome), it serves to provide a game-theoretic interpretation of the high-probability states in the empirical model in the next section.

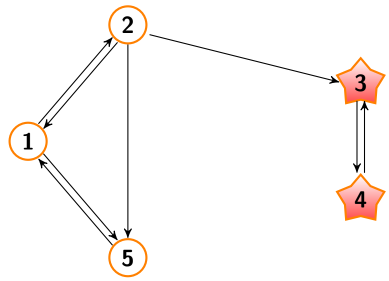
2.1 Setup

Consider a finite-size set of peers $I = \{1, 2, ..., n\}$, where every individual is identified by a vector of exogenous attributes X_i. In addition, individuals decide on a binary action $a_i \in \{0, 1\}$. The collection of all student cohorts in a given high school at a given time period is a good example of such a pool. Another example is the population of a geographically isolated area.[13] Peers' attributes X_i may include age, race, gender, belonging to a grade-class, etc., and the action a_i may refer to any binary choice such as the decision to smoke, to commit a crime, or to purchase a visible good (iphone, clothing, and jewelry), for example.

In addition to actions and attributes, individuals feature relationships that form a social

[13]Any closed collection of individuals who draw friends from within themselves will fit the assumptions of the model. Alternatively, this assumption (if the data do not span the complete network, for example) can be relaxed if one conditions on the existing friendships with outsiders of the pool. Peers who are not in the data are not part of the model, in the sense that the links to them can be thought of as part of the fixed attributes X_i. Then the model will explain the formation of new friendships conditioning on the ones with outsiders.

Figure 2: Example of Network State



Note: Example with $n = 5$ individuals and 8 nominations. In 3 cases the nominations are reciprocal - $g_{12} = g_{21} = 1$, $g_{15} = g_{51} = 1$, and $g_{34} = g_{43} = 1$. The circle-shaped clear node indicates an action status of 0, i.e., $a_1 = a_2 = a_5 = 0$, while the star-shaped shaded node indicates an action status of 1, i.e., $a_3 = a_4 = 1$.

network; the individuals are tied by interdependencies, which I will think of as friendships. The relationship between peers i and j is summarized in the binary variables g_{ij} and g_{ji}. In particular, if i nominates j as a friend, then $g_{ij} = 1$ (i forms a one–directional link to j). Clearly, this notation does not preclude the existence of asymmetric relationships, i.e., individual i may nominate j as a friend but not vice versa.[14]

Figure 2 illustrates the setup with $n = 5$ individuals. In the graph, each individual in $I = \{1, 2, 3, 4, 5\}$ is depicted as vertex and a friendship nomination is shown as a directed edge. In the example, the relationship between individuals 1 and 2 is symmetric, i.e., $g_{12} = g_{21} = 1$. However, this is not the case for individuals 2 and 3 – individual 2 has nominated 3 but the converse is not true ($g_{23} = 1$ and $g_{32} = 0$). Finally, in Figure 2 the star-shaped nodes denote individuals with an action status equal to 1 ($a_3 = a_4 = 1$).[15]

At any point in time the outcome is completely determined by the network state S, which includes the state of all *edges* (friendship statuses) g_{ij} and *vertices* (action statuses) a_i. Formally:[16]

$$S = \left(\{g_{ij}\}_{i,j=1}^{n}, \{a_i\}_{i=1}^{n} \right).$$

Let $\mathbf{S}$ be the set of all possible (directed and unweighted) n-node networks. Note that, even for small networks, $\mathbf{S}$ is a large set – $|\mathbf{S}| = 2^{n^2}$ – and grows exponentially with n.[17] Finally, it will

[14]One can argue that human frienships are symmetric by nature and should be modeled as such. However, I am interested in friendship relations as a mean to transmit influence, i.e., I am thinking about friends as role models. Moreover, only about 35% of the friendships are reciprocal in my estimation sample.

[15]The model in this paper does not rely on any specific assumptions about the type of actions, other than being a binary choice that is potentially related to the friendship network.

[16]To keep the algebraic expressions comprehensible, I keep g_{ii} in the state space and assume $g_{ii} = 0$ for all i. Consequently, I reserve $n^2 + n$ variables to index a state space of size n^2.

[17]For example, for $n = 10$, $|\mathbf{S}| = 2^{100}$.

be convenient to denote the state of the network including all edges and vertices but one edge with S_{-ij}. Similarly S_{-i} fixes the network state excluding vertex (the action of agent) i.

Choices and preferences

In the model, individuals select their friends and choose their actions. Formally, the decision of individual i is given by $S_{(i)} = (a_i, \{g_{ij}\}_{j \neq i}) \in \mathbf{S}_{(i)}$. Here $\mathbf{S}_{(i)}$ is i's choice set i.e. the set of all possible friendships and action statuses. An individual chooses $S_{(i)} \in \mathbf{S}_{(i)}$ to maximize her utility $u_i : \mathbf{S} \times \mathbf{X} \longrightarrow \mathbb{R}$, which maps the state $S \in \mathbf{S}$ and the attributes of the population $X = (X_1, ..., X_n) \in \mathbf{X}$ to $\mathbb{R}$. Individual i's utility u_i depends on her action status, her friendship links, the action statuses of her immediate friends and those of the population in the following way:

$$\begin{aligned} u_i(S,X) &= a_i v(X_i) + a_i \sum_j g_{ij} g_{ji} a_j \phi_{ij} + a_i \sum_j a_j h(X_i, X_j) + a_i \sum_j g_{ij} c(X_i, X_j) \\ &\quad + \sum_j g_{ij} w(X_i, X_j) + \sum_j g_{ij} g_{ji} m(X_i, X_j) \\ &\quad + \sum_j \sum_k \left(g_{ij} g_{jk} g_{ki} + g_{ij} g_{ik} g_{jk} \right) q(X_i, X_j, X_k) \end{aligned} \tag{1}$$

Here $v(.)$, $c(.,.)$, $h(.,.)$, $w(.,.)$, $m(.,.)$, and $q(.,.,.)$ are functions of individuals' demographic characteristics. The explicit dependence of ϕ_{ij} on the demographic characteristics of i and j is suppressed here, although this is without loss of generality since the attributes are fixed and the parameter is indexed by i and j.

The incentives encoded in (1) are easier to understand, if we consider the incremental payoffs from changing just one dimension of an individual's decision, i.e., the incremental payoff of changing the action status or a single friendship. Below I analyze each in turn.

Action choice Consider the incremental payoff for individual i from changing her action status $\Delta_{a_i} u_i(S,X) = u_i(a_i = 1, S_{-i}, X) - u_i(a_i = 0, S_{-i}, X)$, holding her friendships and the choices of the rest of the population fixed. I allow $\Delta_{a_i} u_i(S,X)$ to depend on i's observables, her friendship links, and the choices of the overall population. From (1) it can be seen that:

$$\Delta_{a_i} u_i(S,X) = v(X_i) + \underbrace{\sum_{j \neq i} a_j g_{ij} g_{ji} \phi_{ij}}_{\text{local peer effect}} + \underbrace{\sum_{j \neq i} a_j h(X_i, X_j)}_{\text{aggregate peer effect}} + \underbrace{\sum_{j \neq i} g_{ij} c(X_i, X_j)}_{\text{contextual effects}} \tag{2}$$

All terms in (2) have a straightforward interpretation. First, the utility of choice $a_{i,t} = 1$ may depend on an individual's attributes - $v(X_i)$, e.g., it is well known that blacks smoke less and students in higher grades smoke more. The next two (summation) terms capture peer effects. The local peer effect terms - $a_j g_{ij} g_{ji} \phi_{ij}$ - allow friends' behavior to have a direct influence on an individual's utility. That is, one is influenced strongly by the behavior of her friends as opposed to casual individuals. On the other hand, the function $h(.,.)$ in the second summation captures the aggregate peer effect - the impact of the behavior of the overall population. Intuitively, a person may be potentially influenced from observing the behavior of the surrounding population, irrespectively of whether these are friends or not. Note how in $h(.,.)$ the influence between individuals is contingent on their attributes X_i and X_j. For example, males may be more likely to be affected by the observed behavior of other males as opposed to the observed behavior of females. Finally, the contextual effects term captures the influence of an individual's friends characteristics (rather than choices) on i's decision. In particular, the attributes of the social environment one self-selects into may impact an individual's choices separately from friends' behavior (Manski (1993)).

Observe that in (2) the influence between peers is symmetric in that only mutual friendships affect an individual's payoff $\Delta_{a_i} u_i(S, X)$ (the local peer effect terms are $g_{ij} g_{ji} a_j \phi_{ij}$). In general, however, one can argue that peer influence may be asymmetric and terms of the form $g_{ij} a_j \psi_{ij}$ should be considered. Unfortunately, with only a cross-section of networks, it is difficult if not impossible, given the presence of links between two smokers (i.e. $g_{ji} a_i a_j = 1$), to separately identify the causal effect of the sender on the receiver versus that of the receiver on the sender. More specifically, if one wants to model the effect of one-directional friendships on smoking, identification (and tractability) concerns will require the peer effect term in the incremental utility of i to smoke $\Delta_{a_i} u_i(S, X)$ to look like $a_j(g_{ji}\psi_{ji} + g_{ij}\psi_{ij})$. That is, not only does a friendship receiver who is a smoker (i.e., who is nominated as a friend) affect nominator's utility from smoking but a friendship sender who is a smoker (i.e., who nominates somebody as a friend) also influences the nominee's utility from smoking. The model can be readily extended to accommodate such considerations; however, given the limited amount of network data available for estimation, I choose not to do so. Given the restriction of modeling the local peer effects in a symmetric way, arguably mutual friendships, i.e., friendships where the relationship is perceived as reciprocal, have the potential for a larger influence on an individual's behavior.

Friends selection Suppose that at some point individual i is considering the possibility of befriending j.[18] On the margin, i's incremental payoff is defined as $\Delta_{g_{ij}} u_i(S, X) = u_i(g_{ij} =$

[18] Again the decisions do not need to be sequential. Rather I am presenting the *incremental* payoffs of a friendship so the associations between functional forms and the object of interest are apparent.

$1, S_{-ij}, X) - u_i(g_{ij} = 0, S_{-ij}, X)$ and can be obtained from (1):

$$\Delta_{g_{ij}} u_i(S,X) = w(X_i,X_j) + \underbrace{g_{ji} m\left(X_i,X_j\right)}_{\text{reciprocity}} + \underbrace{a_i a_j g_{ji} \phi_{ij}}_{\text{choice segregation}} + \underbrace{a_i c(X_i,X_j)}_{\text{contextual effect}} + \sum_k \underbrace{(g_{jk}+g_{kj})(g_{ki}+g_{ik}) q(X_i,X_j,X_k)}_{\text{common friends}} \tag{3}$$

The first term $w(X_i, X_j)$ captures the baseline incentives of i to befriend j, which may or may not depend on their degree of similarity, i.e., same sex, gender, race, etc. The term $g_{ji} m\left(X_i, X_j\right)$ defines the extra utility of a reciprocal friendship versus a one-sided nomination. The third term in the utility function deserves special attention. It reflects the degree of similarity in the choices of i and j, which is allowed to create addition stimuli for i to nominate j as a friend. Thus action similarity (i.e., $a_i a_j = 1$) has the potential to influence i's decision to befriend j. As before, the contextual effect term is motivated by the observation that certain behaviors are better accepted in some but not other social circles. Finally, the last term is meant to capture the empirical regularity that individuals who share common friends are usually friends themselves (so-called clustering).

It is important to emphasize that the labels of the terms in (2) and (3) are figurative and are only meant to draw intuition from well-recognized, in the literature on social interactions, identification concerns. In this model, for example, the question of whether a friend's choices or a friend's attributes impact an individual's behavior is no longer an issue. Indeed, because friendships are choices, it does not make sense to condition on who one's friends are to explain, say, teen smoking behavior. In the context of the model, individuals select their friends so that their friends' attributes are just an attribute of their choice (as suggested by the term $a_i g_{ij} c(X_i, X_j)$ in the utility).

2.2 k-player Nash stable network states and Nash partitions

I consider a class of (non-cooperative) equilibria that gradually relax the assumptions of rationality and coordination underlying the Nash equilibrium concept, yet may have predictive power. For $1 < k \leq n$, k-player Nash stability extends the notion of pair-wise stability in Jackson and Wolinsky (1996) to non-cooperative settings in the spirit of Bala and Goyal (2000).[19] Broadly speaking, stable states are states that are robust to a certain class of deviations. In sizable populations, however, the set of possible deviations may be too large for individuals to optimize on. In our particular settings, for population of size n an individual has to compare

[19] Bala and Goyal (2000) is an early theoretical work discussing what in this model is n-player Nash stability.

her current play against 2^n possible alternatives.[20] The novelty of my approach is to consider equilibria for which the best response consideration set is smaller. That is, whenever individuals decide whether a "profitable" deviation is available, they do not consider all possible friendship combinations (which even for a population of size 100 is 2^{99}). Instead, individuals consider altering only part of their friendships at a time.

Intuitively, a network state is k-player Nash stable, for $1 < k \leq n$, if any set of k agents play a Nash equilibrium of the induced game between the k of them, conditioning on the rest of the network. More specifically, in the induced game each participant chooses actions and friendship links (only to the participants) simultaneously. To formalize this notion, let Γ be the n-player normal form game $\{\mathbf{S}_{(i)}, u_i\}_{i \in I}$ where as before $\mathbf{S}_{(i)}$ is the choice set (strategy space) of agent i.

Definition 1 *Let $S \in \mathbf{S}$ and $I_k = \{i_1, i_2, ..., i_k\} \subseteq I$, for $1 < k \leq n$. Define the* **restriction of Γ to I_k at** *S as the k-player normal form game $\Gamma|_{I_k}$, in the following way: (i) The set of players is $\{i_1, i_2, ..., i_k\}$; (ii) The strategy space is $\mathbf{S}_{(i|I_k)} = \{0, 1\}^k$, for $i \in I_k$, with typical element $S_{(i|I_k)} = (a_i, \{g_{ij}\}_{j \in I_k})$; (iii) The payoffs are given by the restriction $u_i|_{\mathbf{S}_{(I_k)}}$ of u_i in (1) on $\mathbf{S}_{(I_k)} = \times_{i \in I_k} \mathbf{S}_{(i|I_k)}$.*

Definition 2 *For $1 < k \leq n$, the network state $S \in \mathbf{S}$ is said to be* **k-player Nash stable** *provided that for any set of k players $I_k \subseteq I$, the restriction of S on I_k*

$$S|_{I_k} = (\{a_i\}_{i \in I_k}, \{g_{ij}\}_{i,j \in I_k}) \subseteq S \tag{4}$$

is a Nash equilibrium of $\Gamma|_{I_k}$ at S.

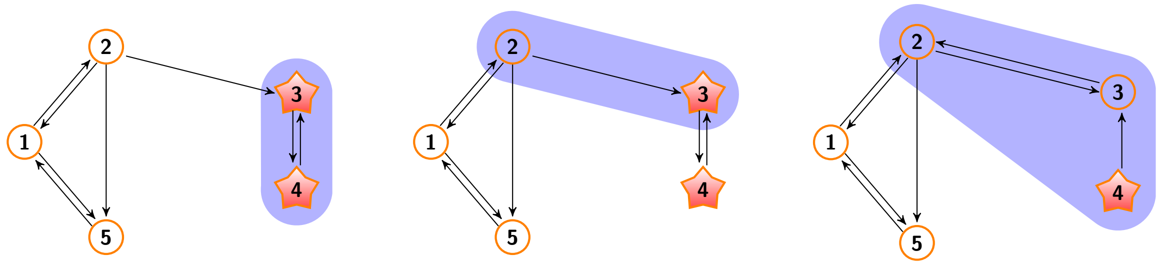
Note how in the above definition k-player Nash stability weakens the notion of a Nash equilibrium. In the former, every player chooses a best response with respect to at most $k - 1$ other players, while in the latter a player chooses a best response with respect to the whole population. To better understand the properties of this concept, consider the following definition.

Definition 3 *Consider a partition $\iota_m = \{I_1, I_2, ..., I_m\}$ of the set of players I i.e. $I = \cup I_j$ and $I_j \cap I_k = \emptyset$ for $j \neq k$. For the normal form game Γ, define a state $S \in \mathbf{S}$ to be* **partition Nash stable with respect to partition** *ι_m provided $S|_{I_k}$ is a Nash equilibrium of $\Gamma|_{I_k}$ for all $I_k \in \iota_m$.*

Figure 3 illustrates the notion of k-player Nash stability in a network with $n = 5$ individuals. The left two diagrams show a network state which is 2-player Nash stable. In such a state for

[20] Recall that the size of a player's strategy space $\mathbf{S}_{(i)}$ is $|\mathbf{S}_{(i)}| = 2^n$.

Figure 3: Example 2- and 3- player Nash stability



(a.) $\Gamma|_{\{3,4\}}$ (b.) $\Gamma|_{\{2,3\}}$ (c.) $\Gamma|_{\{2,3,4\}}$

every pair, the individuals within the pair are forming relationships with each other optimally (i.e., each individual plays best response to the strategy of the other). For example Diagram ($a.$) considers the pair consisting of individuals 3 and 4. In the game between them $\Gamma|_{\{3,4\}}$, whenever individual 3 chooses (a_3, g_{34}) and individual 4 chooses (a_4, g_{43}), the shaded outcome constitutes a Nash equilibrium. Similarly the shaded region in Diagram ($b.$) depicts a Nash equilibrium of $\Gamma|_{\{2,3\}}$. Now suppose that although individual 3 best response in $\Gamma|_{\{2,3\}}$ and $\Gamma|_{\{3,4\}}$ is consistent with the depicted state, his best response in $\Gamma|_{\{2,3,4\}}$ would be to sever his friendship with 4, nominate 2 as a friend, and choose $a_3 = 0$. Diagram ($c.$) depicts such a deviation which is not permissible in the context of 2-player Nash stability.

Denote the set of all k-player Nash stable states with S^*_{k-NS}. Intuitively, as k increases, k-player Nash stability places stronger requirements on what outcomes players consider as optimal (in the sense of being a best repose to the play of the rest). Indeed, the larger k is the larger is the consideration set over which a player determines her best response. Consequently, one would expect S^*_{k-NS} to shrink as k increases. Moreover, a moment of reflection reveals that k-player Nash stability is directly related to partition Nash stability. In particular, a network state is k-player Nash if and only if it is partition Nash stable with respect to any partition with maximum component of size less than k. The lemma below formalizes this intuition. (Its proof is in the appendix on p. 45.)

Lemma 1 *Let Γ be the n-player normal form game $\{\mathbf{S}_{(i)}, u_i\}$. Then the following holds:*

(i) A state $S \in \mathbf{S}$ is k-player Nash stable if and only if S is partition Nash stable with respect to all partitions $\iota_m = \{I_1, I_2, ..., I_m\}$ such that $|I_r| \leq k, \forall 1 \leq r \leq m$.

(ii) The family of k-player Nash stable states is ordered as

$$S^*_{2-NS} \supseteq ... \supseteq S^*_{k-NS} \supseteq ... \supseteq S^*_{n-NS}.$$

2.3 Potential function, existence and characterization of equilibrium

This section discusses the conditions under which the preferences in the model can be represented by a potential function.[21] This will allow me to draw on the theory of potential games and obtain a closed-form expression for the likelihood of the model. The latter facilitates substantially the empirical implementation of the proposed framework and makes transparent the argument for identification of the model's parameters.

The potential $\mathcal{P}$ is a single (not indexed by players) function on the set of network realizations $\mathbf{S}$ (and individual attributes $\mathbf{X}$), which summarizes the marginal incentives of the players given in (2) and (3). More specifically, for all $S_{(i)}, S'_{(i)} \in \mathbf{S}_{(i)}$ the potential has to satisfy:

$$u_i(S_{(i)}, S_{-(i)}, \mathbf{X}) - u_i(S'_{(i)}, S_{-(i)}, \mathbf{X}) = \mathcal{P}(S_{(i)}, S_{-(i)}, \mathbf{X}) - \mathcal{P}_i(S'_{(i)}, S_{-(i)}, \mathbf{X}).$$

In the context of the model, the following assumption is equivalent to the above condition:

Assumption 1 *For functions h, m, and q in (1)*

$$\begin{aligned} h\left(X_i, X_j\right) &= h\left(X_j, X_i\right) \\ m\left(X_i, X_j\right) &= m\left(X_j, X_i\right) \\ q\left(X_i, X_j, X_k\right) &= q\left(X_{i'}, X_{j'}, X_{k'}\right) \end{aligned}$$

hold for any $i, j, k, i', j', k', \in I$ such that (i', j', k') is a permutation of (i, j, k). Furthermore

$$\phi_{ij} = \phi_{ji}.$$

Assumption 1 places a symmetry type of restrictions on the direction of influence between individuals. In essence, these are identification restrictions. For example, that $\phi_{i,j} = \phi_{j,i}$ implies that the magnitude of potential influence on behavior from i to j equals the one from j to i. Whenever i and j are mutual friends and smoke, one cannot separately tell which way the causality went, i.e., it was i who caused j to smoke or vice versa. Importantly, however, the symmetry assumption does not imply that the overall peer effect is the same between individuals who are friends. Indeed, two friends do not need to have the same network of friends and hence the same peer pressure on their decisions.

With assumption 1 the following proposition holds (its proof is in the appendix on p. 45):

[21] Congestion games were the first class of games exhibiting this property (Beckmann et al. (1956); Rosenthal (1973)). Monderer and Shapley (1996) recognize that congestion games are instances of games with potential, propose several notions of potential functions for games in strategic form, and obtain the characterization of potential games.

Proposition 1 *With assumption 1, the preferences given in (2) and (3) are summarized with the* potential function $\mathcal{P} : \mathbf{S} \times \mathbf{X}_n \longrightarrow \mathbb{R}$ *:*

$$\begin{aligned}
\mathcal{P}(S, X) \quad = \quad & \sum_i a_i v(X_i) + \sum_i \sum_j a_i g_{ij} c(X_i, X_j) \qquad (5) \\
& + \frac{1}{2} \sum_i \sum_j a_i a_j g_{ij} g_{ji} \phi_{ij} + \frac{1}{2} \sum_i \sum_j a_i a_j h(X_i, X_j) \\
& + \sum_i \sum_j g_{ij} w\left(X_i, X_j\right) + \frac{1}{2} \sum_i \sum_j g_{ij} g_{ji} m\left(X_i, X_j\right) \\
& + \sum_i \sum_j \sum_k \left(\frac{g_{ij} g_{jk} g_{ki}}{3} + g_{ij} g_{ik} g_{jk} \right) q(X_i, X_j, X_k)
\end{aligned}$$

In particular the one-shot network formation game is a potential game*.*

The conclusion of proposition 1 precludes some pathological behavior whenever I introduce temporal dimensions of the play. It also facilitates the empirical implementation of the framework and helps us understand the generality of the decision protocol. In particular, it is useful in demonstrating that modes of perturbed best response dynamics are observationally equivalent (see theorem 2 on p. 21). Before we proceed to a more substantive analysis, the proposition below establishes existence. Its proof follows trivially from the existence of a potential function and the fact that the potential attains maximum on a finite domain.[22]

Proposition 2 *With assumption 1, for any $1 < k \leq n$, there exists at least one k-payer Nash stable state. Moreover the set of all k-player Nash stable states is characterized by:*

$$S^*_{k-NS} = \left\{ S \in \mathbf{S} : S_{(i|I_r)} \in \arg\max_{\tilde{S}_{(i|I_r)}} \mathcal{P}(\tilde{S}_{(i|I_r)}, S_{-(i|I_r)}, X) \quad \forall I_r \subseteq I, |I_r| \leq k \right\} \qquad (6)$$

2.4 A behavioral model of network formation

Broadly speaking, k-player Nash stable network states, including the Nash equilibrium of the one-shot game, are states that are robust to a certain class of deviations. Importantly, however, stability concepts dispense with any behavioral model of network formation. While there is some argument for the state to persist once it is reached, it is not clear what behavioral mechanism will drive the system to such a state. Moreover, it is not clear if these states are stable under a perturbation not in the neighborhood of an equilibrium.

To address this issue, consider a time dimension of the play in the spirit of Blume (1993), where individuals make their choices sequentially and revise these choices frequently. In each

[22]The proof, omitted here for brevity, closely follows Rosenthal (1973), and Monderer and Shapley (1996).

period, which need not be of equal length, one agent adapts her strategy with respect to the current state of the network.[23] Note that, in rationalizing the individual's responses, such models put all the weight on the contemporaneous environment rather than on the human ability to forecast the actions of the other. This is an adequate visualization of the reality under at least two scenarios. First, if the action of a single individual cannot substantially affect the future development of the network. In such cases individuals do not need to behave strategically with respect to the future response of the network. Second, if individuals form stationary expectations about the future that are, in fact, consistent with the current network state. In this case, although individuals need to act strategically with respect to the future, their best response to the current state is indeed a best response to the expected network state tomorrow as well. Thus, the proposed approach is likely to be plausible in studying decisions such as smoking and drinking, while it seems inappropriate in situations where an individual's forecast about the future is likely to be important (for example, to study fertility or career choices).

Meeting process of dimension k and adaptive network dynamics

In every period a randomly chosen individual (say i) meets a random set I_k of $k \leq n-1$ individuals from the population I. All individuals in I_k are potential friends and upon meeting them i decides jointly whether to befriend any (or all) of them and whether or not to change her action status a_i. In particular, i chooses what is currently her best option with respect to her friends under consideration, taking into account what their choices are (i.e., their action statuses, for example), what the choices of the population overall are, and who $i's$ current friends are who not under consideration. It is clear that the sequence of meetings and players' optimal decisions induce a sequence of network states (S_t), which is indexed by time subscript t.

The rules of the play in period t are determined by the outcome of a (stochastic) meeting process μ_t, which outputs who makes choices and who are the individuals considered as potential friends. Formally:

$$\Pr\left(\mu_t = (i, I_k)\,|S_{t-1}, X\right) = \mu_{i,I_k}\left(S_{t-1}, X\right) \tag{7}$$

In the simplest case, any meeting is equally probable so that:

$$\mu_{i,I_k}\left(S_{t-1}, X\right) = \frac{1}{n}\frac{1}{\binom{n-1}{k}}$$

Note, however, that (7) is general enough to accommodate meeting technologies with bias for

[23] In my settings, the sequential mode of decision making is similar to a process where every period a set I_k of agents play a Nash equilibrium of the induced game $\Gamma|_{I_k}$ at S_t. These two models have similar asymptotic behavior if a certain equilibrium selection rule for $\Gamma|_{I_k}$ is assumed, which is not in the spirit of my pursuit. In particular, I am interested in how decentralized adaptive behavior justifies the emergence of an equilibrium.

people with similar actions and/or characteristics, as, for example,

$$\mu_{i,I_k}\left(S_{t-1},X\right)\propto\exp\left\{-\max_{j\in I_k}\{d(a_i,a_j)\}-\max_{j\in I_k}\{d(X_i,X_j)\}\right\}$$

with d being a generic distance function.

As mentioned earlier, the goal of considering a model of sequential decision making is to provide behavioral justification of the notion of Nash stability. In this respect it is desirable for Nash stable states to be *attractors* of the proposed dynamic, that is, with the time for the system to approach such a state. Moreover, one would expect an equilibrium to be an *absorbing* state: once reached, the system will remain there forever. A moment of reflection reveals that the set of k-player Nash stable states obeys the latter. The former property is somewhat more technical and requires us to make the following, fairly mild, assumption:

Assumption 2 *For any configuration of the network and the attributes of the population, any meeting is possible:*

$$\Pr\left(\mu_t=(i,I_k)|S,X\right)>0$$

for all $i\in I$, $I_k\subsetneq I$, $I_k\neq\emptyset$, $S\in\mathbf{S}$ *and* $X\in\mathbf{X}$.

The above discussion is formalized in proposition 3. (Its proof is in the appendix on p. 46.)

Proposition 3 *Let* $1<k\leq n$ *and consider a meeting process of dimension* $k-1$ *(i.e. where an individual meets* $k-1$ *friends every period). With assumptions 1 and 2 the following hold*

1. *Any* $S\in S^*_{k-NS}$ *is absorbing i.e. if* $S_t\in S^*_{k-NS}$ *then* $S_{t'}=S_t$ *for all* $t'>t$.
2. *Independently of the initial condition (distribution)*

$$\Pr\left(\lim_{t\to\infty}S_t\in S^*_{k-NS}\right)=1$$

The first part of proposition 3 is implied immediately from the definition of S^*_{k-NS}. In this setting, the second part follows elegantly from the observation that the potential function $\{\mathcal{P}_t\}$ is a sub-martingale.

While the dynamic perspective introduced above addresses the concern raised by Kandori, Mailath and Rob (1993), namely, it explains how equilibrium is reached from the behavior of the agents, it exhibits some undesirable properties. First and foremost, it does not provide a convenient statistical framework that can be estimated given network data. In the model, the uncertainty is driven solely by the meeting process, which is likely to be unobservable. Thus, which equilibrium will be played is entirely determined by the realization of μ. Moreover, it is not *a priori* clear which mode of the meeting process is observed in the data, i.e., what is k, and different k will result in observing different classes of equilibria. These concerns are addressed in the next section.

3 An Estimable Framework

To reflect more realistically the fact that there are unobserved factors that affect players' considerations, this section introduces random preference shocks to the utility from friendships and actions. Factors relevant to the context of application could be, for example, appearance characteristics, contextual effects, mood, emotional predisposition, etc. The incorporation of preference shocks into the sequential model of network formation eliminates absorbing states and allows us to obtain a non-degenerate likelihood. The second part of this section explores particular distributional assumptions that guarantee a closed-form expression for the likelihood and demonstrates the invariance of the model to the dimension of the meeting process k.

3.1 Behavioral model of network formation revisited

The meeting process in the behavioral model determines who gets to play and what the choices are. In particular, a meeting process of dimension k defines a discrete consideration set of cardinality 2^k. That is, the individual chooses among $k-1$ potential friends and her action status a_i. The assumption below introduces to this discrete choice problem a random preference shock very much in the spirit of a random utility model.[24]

Assumption 3 *Suppose that the utilities in (1) contain a random preference shock. More specifically, let*

$$\overline{u}_i(S,X) = u_i(S,X) + \epsilon_{i,S}$$

with $\epsilon_{i,S} \sim$ *i.i.d. across time, individuals and network states. Moreover, suppose that* ϵ *has c.d.f. and unbounded support on* $\mathbb{R}$.

The matching process $\{\mu_t\}_{t=1}^{\infty}$ and the sequence of optimal choices, in terms of friends selection and individual actions, induce a Markov chain of network configurations on $\mathbf{S}$.[25] The above assumption guarantees that this chain obeys some desirable properties, which are formalized below. First, I lay out a formal statement of a theorem and then discuss its implications. (The proof is in the appendix on p. 46.)

[24]See Thurstone (1927); Marschak (1960); McFadden (1974).
[25]See the stochastic-choice dynamics in Blume (1993).

Theorem 1 *Let $1 < k \leq n$ and suppose assumptions 1, 2, and 3 hold. The Markov chain generated by the network formation game has the following properties:*

1. *There exists a unique stationary distribution $\pi_k \in \Delta(\mathbf{S})$. That is, if the chain has unconditional distribution π_k at period t, it has the same distribution for any $t' > t$.*

2. *Independently of the initial condition*

$$\lim_{t \to \infty} \Pr(S_t = S) = \pi_k(S)$$

   that is, the S_t converges in distribution to $S \sim \pi_k$.

3. *For any function $f : \mathbf{S} \longrightarrow \mathbb{R}$ the ergodic theorem holds*

$$\frac{1}{T} \sum_{t=0}^{T} f(S_t) \xrightarrow{a.s} \bar{f}$$

   where $\overline{f} = \int f(S)\, d\pi_k$.

The conclusion of theorem 1 is relevant for the empirical implementation of the model in a number of directions. First, in the absence of information about the initial network state, part 2 of the theorem warrants the use of the stationary distribution as the likelihood for the model. That is the probability that a specific network state $S \in \mathbf{S}$ is observed can be thought of as being $\pi_k(S)$. Moreover, the uniqueness of π_k precludes any ambiguity in the model's predictions. Finally, to understand the third part of the theorem, note that in the special case when f is the indicator function of state S, it just states that the proportion of time spent at each state converges almost surely to the equilibrium probability of that state. More generally, such a result allows for moments of the stationary distribution to be approximated from a given sequence of realizations of the meeting process and individuals' choices.[26]

The following assumptions are needed to obtain a closed-form expression for the likelihood.[27]

Assumption 4 *Suppose that the preference shock ϵ is distributed Gumbel$(\underline{\mu}, \underline{\beta})$.*

Assumption 5 *Suppose that for the meeting probability μ: (i) $\Pr(\mu_t = (i, I_k))$ does not depend on the relationship status between i and any $j \in I_k$. (ii) $\Pr(\mu_t = (i, I_k))$ does not depend on a_i. These together imply:*

$$\mu_{i,I_k}(S, X) = \mu_{i,I_k}(S', X)$$

[26] This part becomes useful at the estimation stage when, for a given set of parameters, I need to integrate over the stationary distribution.

[27] See Appendix B for details on the Gumbel distribution.

for all $S = (S_{(i|I_k)}, S_{-(i|I_k)})$ *and* $S' = (S'_{(i|I_k)}, S_{-(i|I_k)})$.

The above assumptions suffice to obtain a closed-form expression for the stationary distribution of π_k of S_t. Moreover, this stationary distribution will not depend on k. Although different k lead to different transient dynamics, the statistical properties of the disturbance term guarantee that π_k is invariant to k. To provide intuition behind this result consider two states $S, S' \in \mathbf{S}$. It can be shown that the probability of moving from S to S' is proportional to the probability of returning from S' to S by a factor that does not depend on k. The formal argument can be found in the appendix (p. 47).

Theorem 2 *Under assumptions 1-5, the stationary distribution* π_k*, from theorem 1, is given up to a constant by:*

$$\pi_k(S, X) \propto \exp\left(\frac{\mathcal{P}(S, X)}{\underline{\beta}}\right) \tag{8}$$

In particular, $\pi_k(S, X)$ *does not depend on* k*.*

The closed-form expression for the stationary distribution has advantages for the empirical implementation of the proposed framework, where π_k can be treated as the likelihood. First, one can explore a transparent argument for the identification of the model's parameters. It is clear that, given the variation in the data of individual choices $\{a_i\}_{i=1}^n$, friendships $\{g_{ij}\}_{i,j=1}^n$ and attributes $\{X_i\}_{i=1}^n$, functional forms for v, w, m, h, q will be identified as long as the different parameters induce different likelihoods of the data. Second, a closed-form expression for π_k facilitates the use of likelihood-based estimation methods. Last, but not least, theorem 2 demonstrates that the behavioral model with preference shocks indeed provides a unifying framework for the meeting processes with different dimensions, i.e., no matter what the dimension of the meeting process is the probability of observing S is the same.

3.2 Discussion

The family of stochastic best response dynamics (indexed by the dimension k) considered above has a particular meaning in our settings. Their (common) stationary distribution allows us to probabilistically *rank* the different classes of Nash stable network states. Below I discuss this point further, extend the family of k-dimension meeting processes to a random-k-k-player meeting process and comment on the link with the notions of a quantal response equilibrium (McKelvey and Palfrey (1995)) and a correlated equilibrium (Aumann (1974)).

Probabilistic ranking and long-run equilibria

The stationary distribution obtained in theorem 2 gives an intuitive (probabilistic) ranking of the family of k-player Nash stable equilibria. Under π, all k-Nash stable states receive a positive probability (across all possible values of k). Moreover, and realistically, a network state will receive a positive probability, although it may not be an equilibrium in any sense. It will be desirable, however, that in the vicinity of an equilibrium, the equilibrium to receive the highest probability. A somewhat related point is on the relation of the mode of π (i.e. the state with the highest probability) with the limiting notion of a long-run equilibrium from evolutionary game theory (see Kandori et al. (1993)). My goal in this section is to provide a new perspective to the theoretical results on equilibrium selection from evolutionary game theory.

It will be easier to formalize the intuition from the above discussion with some further definitions. Define the neighborhood $\mathbf{N}_S \subset \mathbf{S}$ of $S \in \mathbf{S}$ as:

$$\mathbf{N}_S = \left\{S' : S' = (g_{ij}, S_{-ij}), i \neq j\right\} \bigcup \left\{S' : S' = (a_i, S_{-i})\right\}$$

Define a state S as a *long-run* equilibrium of the network formation model if for any sequence of vanishing preference shocks, the stationary distribution π places a positive probability on S.[28] With this definition the following result obtains (its proof can be found in the appendix on p. 48).

Theorem 3 *Suppose assumptions 1-5 hold:*

1. *A state $S \in \mathbf{S}$ is 2-player Nash stable if and only if it receives the highest probability in its neighborhood $\mathbf{N}_S$.*

2. *The most likely network state $S^{mode} \in \mathbf{S}$ (the one where the network spends most of its time) is a Nash equilibrium of the one-shot game i.e. $S^{mode} \in S^*_{n-NS}$.*

3. *There is a unique long-run equilibrium of the underlying theoretical model that is given by S^{mode}, which need not be Pareto efficient.*

Meeting process of random dimension

Now consider what appears to be a very unrestrictive meeting process, where every period a random individual meets a set of potential friends of random size and composition. Let κ be a discrete process with support $2, \ldots, n$ and augment the meeting process with an additional initialization step with respect to the dimension of μ. In particular, at each period first κ is realized

[28] For more details on this, see Kandori et al. (1993)

and then μ^k is drawn just as before. It is relatively straightforward to establish, without any assumptions on the process κ, that this augmented process has the same stationary distribution π as the one from theorem 2.[29] The point that I want to make is that alternative modeling techniques of the meeting process result in observationally equivalent models. This is important in that it demonstrates that the model predictions do not hinge on ad hoc assumptions.

Quantal response equilibrium

The notion of a quantal response equilibrium (QRE) is based on a fixed point of the quantal-response functions (QRFs), very much like Nash equilibrium is a fixed point of the best response functions. The QRF of player i is a smoothed best response function, where the strictly rational choice of player i (i.e., the best response) is replaced by an approximately rational response. A (regular) quantal-response function satisfies the following axioms:

1. Interiority: every strategy in i's strategy space receives a strictly positive probability.

2. Continuity: the probability of player i choosing pure strategy s is a continuously differentiable function of i's expected payoff from choosing s.

3. Responsiveness: the first derivative of the above probability is strictly positive for all players on the support of the expected payoffs.

4. Monotonicity: strategies with higher expected payoff receive a higher probability.

The existence of a (regular) quantile response equilibrium of a finite-player finite-strategy space normal form game trivially follows from Brower's fixed point theorem. Any such equilibrium induces a probability distribution π^{QRE} over $S = S_1 \times S_2 \times ... \times S_n$ where S_i is the set of pure strategies of player i. Note that π^{QRE} is the Cartesian product of the equilibrium quantile responses and whence inherits their properties - i.e., the conditional distributions satisfy the axioms above. By way of comparison, the stationary distribution π of the network formation model bears some similarities to the axioms of QRF inherited in π^{QRE}. However, there are important differences. I discuss each in turn.

Proposition 4 *The conditional distribution of player i's choices (i.e., $a_i, \{g_{ij}\}_{j \neq i}$) on the choices of the rest of the players – $\pi_{(i)|(-i)} \in \Delta(\{0,1\}^n)$ – induced by the optimal play in the network formation model satisfies all properties of a (regular) quantile response function - interiority, continuity, responsiveness, and monotonicity. However, a QRE cannot induce π.*

[29] I omit here the formal statement and the proof as it essentially follows the one from 2.

The first part follows trivially from the expression for π, given by theorem 2. Since the discrete network formation game is a game with potential, all payoffs can be represented by a single function of the network state - the potential $\mathcal{P}$. Importantly, the potential ranks the network states consistently with individual preferences and, at the same time, is linked to the probability of observing a given state. Thus, it is not surprising that π exhibits the intuitive properties of π^{QRE}.

For the second part, note that the sequential decision process induces a correlation between players' decisions. In particular, the distribution π in theorem 2 cannot necessarily be factored as a product of marginal probability distributions over each player choice set $\mathbf{S}_{(i)}$. (for a formal example see page 48) It then follows that QRE cannot induce π. For the same reason, π cannot be induced by a mixed strategy profile of the unperturbed n player one-shot network formation game. In essence, the outcome of the sequential network play is bears similarities to the notion of a correlated equilibrium in Aumann (1974).

4 Structural Estimation

This section describes the data, the estimation method and the structural parameterization of the model. To estimate the model, I adapt an MCMC algorithm developed by Geyer and Thompson (1992) to simulate the likelihood. I use sample information on adolescent friendship patterns by gender, grade, and race to parsimoniously parameterize the general utility specifications (2) and (3) in a way that is consistent with prominent features of the data. I also discuss a re-parameterization technique, which relates the model's parameters to the conditional choice probabilities in the statistical structure. The discussion of estimation results and model fit is postponed to the next section.

4.1 Monte Carlo maximum likelihood estimator

Within the empirical model of network formation, theorem 1 demonstrates that the network state evolves according to a Markov chain with a unique stationary distribution π on the set of all network states $\mathbf{S}$. Because no information is available on when the network process started or on its initial state, the best prediction about the network state is given by π. Thus, for estimation purposes, the expression for the stationary distribution from theorem 2 can be treated as the likelihood. Given a single network observation $S \in \mathbf{S}$, the log-likelihood is given by:[30]

$$l(\theta) = \ln \frac{\exp\{\mathcal{P}_\theta(S)\}}{H_\theta}$$

where $\mathcal{P}_\theta$ is the potential (at parameters θ) and $H_\theta = \sum_{S\in\mathbf{S}} \exp\{S\}$ is an (intractable) normalizing constant. The size of $\mathbf{S}$ and the summation in calculating H_θ are so large, even for small networks, that the value of $l(\theta)$ cannot be calculated directly for practical purposes.[31] Fortunately, it is possible to approximate the likelihood l via an MCMC algorithm and proceed with a maximum likelihood estimation.

The idea behind the estimation method is to transform the likelihood in a way that makes it possible to approximate the intractable H_θ via a suitably generated Markov chain on $\mathbf{S}$. Then, one can maximize the approximated likelihood rather than the likelihood itself and obtain an approximation of the maximum likelihood estimator. The approximation is valid as the Monte Carlo effort grows. The general approach was originally developed by Geyer and Thompson (1992), and in this setting, it reduces to Gibbs sampling (Geman and Geman (1984)). I briefly describe the approach below.

Fix $\tilde{\theta} \in \Theta$ and an observation $S \in \mathbf{S}$. Instead of maximizing l one can maximize l plus a conveniently chosen constant $l(\tilde{\theta})$, where $\tilde{\theta} \in \Theta$ is a point in the parameter space.

$$\begin{aligned} l^*(\theta) &= l(\theta) - l(\tilde{\theta}) \\ &= \ln \frac{\exp\{\mathcal{P}_\theta(S)\}}{\exp\{\mathcal{P}_{\tilde{\theta}}(S)\}} - \ln \frac{H_\theta}{H_{\tilde{\theta}}} \end{aligned}$$

Now observe that

$$\begin{aligned} \frac{H_\theta}{H_{\tilde{\theta}}} &= \frac{1}{H_{\tilde{\theta}}} \int \exp\{\mathcal{P}_\theta(S)\} d\nu(S) \\ &= \int \frac{\exp\{\mathcal{P}_\theta(S)\}}{\exp\{\mathcal{P}_{\tilde{\theta}}(S)\}} \frac{\exp\{\mathcal{P}_{\tilde{\theta}}(S)\}}{H_{\tilde{\theta}}} d\nu(S) \\ &= \int \frac{\exp\{\mathcal{P}_\theta(S)\}}{\exp\{\mathcal{P}_{\tilde{\theta}}(S)\}} d\mathbb{P}_{\tilde{\theta}}(S) \end{aligned}$$

where ν is the discrete measure on $\mathbf{S}$ and $\mathbb{P}_{\tilde{\theta}}$ is the probability measure on $\mathbf{S}$ induced by $\tilde{\theta}$. The integral on the last row can be simulated with an iid sample from $\mathbb{P}_{\tilde{\theta}}(.)$. It is not clear, however, how to obtain such a sample. Fortunately, if $\mathbb{P}_{\tilde{\theta}}$ is the stationary distribution of the Markov

[30] Note that in the expression for π from theorem 2, the coefficient $\underline{\beta}$ is normalized to equal 1. This is needed because $\mathcal{P}$ is linear in the utility parameters so that one cannot separately identify $\underline{\beta}$.

[31] For $n = 10$ the summation includes 2^{100} terms.

chain $(\tilde{S}_t)$, by a law of large numbers for Markov chains, for

$$\hat{l}_T^*(\theta) \quad = \ln \frac{\exp\{\mathcal{P}_\theta(S)\}}{\exp\{\mathcal{P}_{\tilde{\theta}}(S)\}} - \ln \left(\frac{1}{T} \sum_{t=1}^{T} \frac{\exp\{\mathcal{P}_\theta(\tilde{S}_t)\}}{\exp\{\mathcal{P}_{\tilde{\theta}}(\tilde{S}_t)\}} \right)$$

we have that

$$\arg\max \hat{l}_T^*(\theta) \xrightarrow{a.s.} \arg\max l^*(\theta) = \arg\max l(\theta)$$

In our particular settings $\mathcal{P}$ is linear in θ (and S) so that $\frac{\exp\{\mathcal{P}_\theta(S)\}}{\exp\{\mathcal{P}_{\tilde{\theta}}(S)\}} = \exp\{\mathcal{P}_{\theta-\tilde{\theta}}(S)\}$ and the regularity conditions guaranteeing the above conclusion are satisfied.[32] The Monte Carlo simulated likelihood of observing sample $\{S_j\}_{j=1}^m$ is:

$$\hat{l}_{m,T}^*(\theta) = \frac{1}{m} \sum_{j=1}^{m} \mathcal{P}_{\theta-\tilde{\theta}}(S_j) - \ln \left(\frac{1}{T} \sum_{t=1}^{T} \exp\{\mathcal{P}_{\theta-\tilde{\theta}}(\tilde{S}_t)\} \right)$$

Here m is the size of the sample and T is the Monte Carlo effort, i.e., $(\tilde{S}_t)$ has stationary distribution $\mathbb{P}_{\tilde{\theta}}$.

4.2 The Add Health data

The National Longitudinal Study of Adolescent Health (Add Health) is a longitudinal study of a nationally representative sample of adolescents in grades 7–12 in the United States in the 1994–95 school year. In total, 80 high schools were selected together with their "feeder" schools. The sample is representative of US schools with respect to region of country, urbanicity, school size, school type, and ethnicity. The students were first surveyed in-school and then at home in four follow-up waves conducted in 1994–95, 1996, 2001–02, and 2007–08. This paper makes use of the data from Wave I of the in-home interviews, which contain rich data on individual behaviors, home environment, and friendship networks.[33]

To provide unbiased and complete coverage of the social network, *all* enrolled students in the schools from the so-called saturated sample were eligible for in-home interviews. These were 16 schools of which 2 large schools (with a total combined enrollment exceeding 3,300) and 14 small schools (with enrollments of fewer than 300). One of the large schools is predominantly

[32] For details, see Robert and Casella (2010).

[33] In addition to the in-home interview from Wave I, data on friendship are available from the in-school and Wave III interviews. However, the in-school questionnaire itself does not provide information on important dimensions of an individual's socio-economic and home environment, such as student allowances, parental education, and parental smoking behaviors. On the other hand, during the collection of the Wave III data, the respondents were not in high school any more. For more details on Add Health research design, see `www.cpc.unc.edu/projects/addhealth/design`.

white and is located in a mid-sized town. The other is ethnically heterogeneous and is located in a major metropolitan area. The 14 small schools, some public and some private, are located in both rural and urban areas.

In addition, Add Health data have been merged with existing databases with information about respondents' neighborhoods and communities. For example, the American Chamber of Commerce Research Association (ACCRA) compiles cost of living index, which is linked to the Add Health data on the basis of state and county FIPS codes for the year in which the data were collected. From the ACCRA, I use administrative data on the average price of a carton of cigarettes.[34]

Sample selection and sample statistics

The Wave I in-home questionnaire contains 44 sections collecting a wide array of information about adolescents. In particular, the data contain information about adolescents' friendship networks. Each respondent is asked to nominate up to five of her best male and female friends. If individual A nominates individual B as a friend, this does not imply that B nominates A. Indeed, only 30% of the friendships in the saturated sample are mutual.[35]

In addition to the friendship network data, I use demographic data for the adolescents (age, gender, grade, and race), for their home environments (presence of smoker in the household, pupil's income and allowances, and mother's education), and data for their smoking behavior. The adolescent's smoking status is deduced from the question, "During the past 30 days, on how many days did you smoke cigarettes?" and if the answer was one or more days, the student's smoking status is set to positive. Because all of the students in the saturated sample were eligible for in-home interview, I have detailed information about student friends as well.

Table 1 shows selected descriptive statistics for the estimation sample of $1,125$ students in 14 schools.[36] The overall smoking prevalence averages to 20.8% and varies substantially between schools as the minimum and maximum statistics of 3.8% and 44.7% suggest. Students of all grades are reasonably well represented in the sample, with proportions ranging between 11.2% and 25.9%. This is not the case for students from different races, however. In particular, Asians

[34] For details see the Council for Community and Economic Research `www.c2er.org`, formerly the American Chamber of Commerce Research Association.

[35] Also, both the in-school and the in-home questionnaires contain data about best 5 male and best 5 female friendships. For about 26% of the cases in the in-home sample, however, the interviewer asked only about best 1 male and best 1 female friends. In these cases, to prevent the friendship network from being truncated, I use friendship nominations data from the in-school sample whenever available.

[36] The two largest schools from the saturated sample, with 806 and $1,664$ students, respectively, are excluded (the third largest school has 159 students). With such large networks, estimation is currently infeasible. The computation of the likelihood becomes very costly and the convergence properties of the MCMC algorithm could be very poor.

Table 1: Descriptive Statistics for the Final Sample

	Overall	Min	Max
Students	1125	20	159
Smoking	0.208	0.038	0.447
Male	0.487	0.346	0.581
Grade 7	0.249	-	0.538
Grade 8	0.259	-	0.600
Grade 9	0.121	-	0.214
Grade 10	0.133	-	0.273
Grade 11	0.126	-	0.265
Grade 12	0.112	-	0.214
Whites	0.779	-	1.000
Blacks	0.146	-	0.975
As-Hi-Ot	0.076	-	0.364

Note: The final sample contains students from 14 out of the 16 schools in the saturated sample.

and Hispanic are overall under-represented. For this reason, I merge them into a combined group of Asians-Hispanic-Other (consisting of about 7.6% of the sample).

The saturated sample of Add Health features other prominent characteristics with respect to smoking prevalence and friendship networks. In particular, there is substantial variation in smoking rates by grades and races. In addition, the mixing by grades is not symmetric. That is, adolescents in 7th and 8th grade tend to avoid friends from different grades much more than those in grades 9 to 12. I discuss these dimensions of the network data in connection with the model parametrization in the next section, while I postpone the presentation of the variation in the smoking prevalence rates to the section on the model fit.

4.3 Parametrization

Recall the incremental payoff of changing one's smoking status $\Delta_{a_i} u_i(S, X) = u_i(a_i = 1, S_{-i}, X) - u_i(a_i = 0, S_{-i}, X)$ given in (2)

$$\Delta_{a_i} u_i(S, X) \quad = \quad v(X_i) + \sum_{j \neq i} g_{ij} c(X_i, X_j) + h \sum_{j \neq i} a_j + \phi \sum_{j \neq i} a_j g_{ij} g_{ji} + \epsilon_{a,i}$$

I specify the baseline utility of smoking $v(.)$ as:

$$v(X_i) \quad = \quad v_0 + v_1 \ln(y) + v_2 price + \sum_{.} v_. d_.(X_i)$$

where the indicator functions $d_{.}(X_i)$ include:

1. smoker in the household
2. mother's education (high school or some college)
3. sex
4. grades 9-12 (1 dummy)
5. dummy race (only for) blacks.

I parametrize the contextual effect term $c(X_i, X_j)$ in a way that will enable the model to capture the observed regularity in the data that individuals involved in a cross-gender friendship where the male is older than the female are more likely to smoke. In particular $c(X_i, X_j) = \zeta(d_1(X_i, X_j) + d_2(X_i, X_j))$ where

$$\begin{aligned} d_1(X_i, X_j) &= 1 \text{ if } sex_i = male,\ sex_j = female, \text{ and } grade_i > grade_j \\ d_2(X_i, X_j) &= 1 \text{ if } sex_i = female,\ sex_j = male, \text{ and } grade_i < grade_j \end{aligned}$$

Next, I parametrize the marginal utility of befriending an individual j - $\Delta_{g_{ij}} u_i(S, X) = u_i(g_{ij} = 1, S_{-ij}, X) - u_i(g_{ij} = 0, S_{-ij}, X)$, as (equation (3)):

$$\begin{aligned} \Delta_{g_{ij}} u_i(S, X) &= w(X_i, X_j) + m g_{ji} + \phi g_{ji} a_i a_j \\ &\quad + \sum_k (g_{jk} + g_{kj})(g_{ki} + g_{ik}) q + \epsilon_{g,ij} \end{aligned}$$

where X_i and X_j are individuals i and j exogenous characteristics. The first term $w(X_i, X_j)$, which is the baseline utility of an individual with attributes X_i of befriending someone with attributes X_j, can be flexibly parametrized as

$$\begin{aligned} w(X_i, X_j) &= w_0 + d(\tau_i, \tau_j) \\ &\quad + \sum_{k \in \mathbf{K}^{sex}} \sum_{k' \in \mathbf{K}^{sex}} w^{sex}_{k,k'} d^{sex}_{k,k'}(X_i, X_j) \\ &\quad + \sum_{k \in \mathbf{K}^{grade}} \sum_{k' \in \mathbf{K}^{grade}} w^{grade}_{k,k'} d^{grade}_{k,k'}(X_i, X_j) \\ &\quad + \sum_{k \in \mathbf{K}^{race}} \sum_{k' \in \mathbf{K}^{race}} w^{race}_{k,k'} d^{race}_{k,k'}(X_i, X_j) \end{aligned} \tag{9}$$

where w^{sex}, w^{grade}, and w^{race} are matrices of coefficients, $\tau_i \in \mathbb{R}$ is the type of individual i, and $d(.,.)$ is the Euclidean distance. The type variable is meant to capture the unobserved aspects of an individual, which, although relevant in a friendship, are not measured, such as personality, friendliness, etc. The type is common knowledge to all players and is assumed to be $\tau_i \sim N(0, \sigma_\tau^2)$, iid across individuals.

Expression (9) allows for a very rich specification of the utility of smoking and friendships. In particular, one can estimate $1 + 3 + 35 + 24 = 63$ parameters only from the dummy variables. Given the relatively small sample size of 14 school networks, I first look at the mixing matrices (Table 2). Note that there is little variation between off-race friendship nominations. That is, conditional on students deciding to nominate somebody from a different race, they are relatively indifferent about which race. Following the friendship patterns from the data, I estimate the coefficient on only 1 categorical variable for different races. A similar argument can be made for adolescents in grades 7 and 8 as seen in Table 2 Panel B. Given such considerations, the final parametrization includes indicator functions $d_{k,k'}(X_i, X_j)$ for:

1. different sex
2. different (adjacent) grade
3. different (non-adjacent) grade
4. same grade 9, same grade 10-12
5. different race
6. same race white and black (separately)
7. friendship nomination by Asian-Hispanic-Other of white

Finally, I allow the parameters ϕ_{ij} and q_{ijk} to be different for students in grades 7–8 and 9–12. However, to preserve the identification requirement of assumption 1 (from p. 15), I postulate:

1. If $\max\{grade_i, grade_j\} \leq 8$ then $\hat{\phi}_{ij}$
2. If $\min\{grade_i, grade_j, grade_k\} \leq 8$ then $\hat{q}_{ijk}$

4.4 Identification

To obtain identification, I explore the connection of the model with the family of exponential random graph models (ERGM).[37] These are a broad class of statistical models, capable of incorporating arbitrary dependencies among the links of a network. As a result, ERGM have

[37]See Snijders (2002) for a historical overview of ERGM.

Table 2: Mixing Matrices

PANEL A. RACE

Nominator's race		*Nominee's school grade*		
		Whites	Blacks	As-Hi-Ot
	Whites	**2979 (96%)**	33 (1%)	81 (3%)
	Blacks	35 (8%)	**392 (85%)**	32 (7%)
	As-Hi-Ot	135 (58%)	33 (14%)	**63 (27%)**

PANEL B. SCHOOL GRADE

Nominator's school grade	*Nominee's school grade*					
	7	8	9	10	11	12
7	**833 (82%)**	153 (15%)				
8	96 (9%)	**852 (80%)**	57 (5%)			
9		39 (8%)	**286 (61%)**	63 (14%)	50 (11%)	20 (4%)
10		17 (4%)	57 (12%)	**253 (53%)**	96 (20%)	51 (10%)
11			26 (6%)	89 (20%)	**230 (53%)**	84 (19%)
12				25 (8%)	82 (25%)	**213 (64%)**

Note: In panel B, cells with less than 3% are shown as empty.

been very popular in estimating statistical models of network formation.[38] A powerful corollary of theorem 2 is that the likelihood of the model falls in the family of ERGM. In more general terms, it follows that ERGM are broad enough to incorporate the strategic incentives of the static one-shot play embedded in π.

Corollary 1 *The likelihood of the structural model $l(.)$ can be written as:*

$$l(\theta|S) \propto \exp\left\{\sum_{r=1}^{R} \theta_i w_i(S,X)\right\}$$

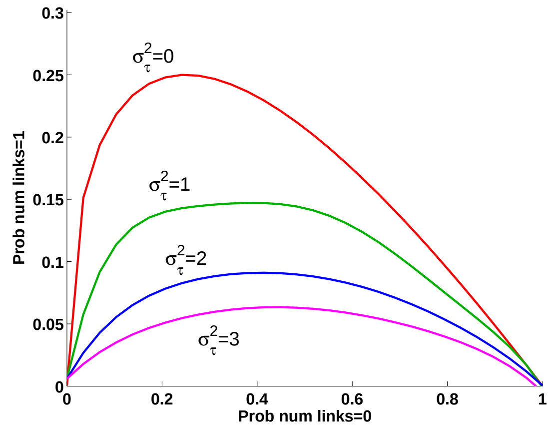
where $w_i : \mathbf{S} \times \mathbf{X} \longrightarrow \mathbb{R}$.

As the number of networks grows to infinity, identification follows from the theory of the exponential family.[39] In particular, it is enough that the sufficient statistics w_i are linearly independent functions on $\mathbf{S} \times \mathbf{X}$. In the structural parametrization of the model above, this

[38] For discussion, see Jackson (2008). A notable exception to the statistical approach is Mele (2010).

[39] See Lehmann and Casella (1998).

Figure 4: Identification of the variance of τ_i



Note: Each curve plots the simulated pair of probabilities $(\Pr(\#links = 0), \Pr(\#links = 1))$ for a fixed σ_τ^2 and a sequence of w_0.

condition is established immediately.[40]

Identification of unobserved heterogeneity in friendship selection

The discussion above secures the identification of all parameters except the variance of the unobserved type σ_τ^2. Recall the type variable τ_i captures individuals' attributes that are relevant for forming friendships, but are unobserved by the econometrician. In particular, individuals who are similar, i.e., whose τ's are closer, are more likely to become friends. Technically the likelihood has to integrate the unknown type:

$$l(\theta, \sigma_\tau | S) = \int \pi(S|\theta, \tau) dP(\tau) \tag{10}$$

where $\tau = (\tau_1, ..., \tau_n)$ and τ_i is i.i.d. $N(0, \sigma_\tau^2)$. While closed-form expression for the likelihood cannot be derived, simulation evidence can be provided that σ_τ is identified from the rest of the parameters.

Heuristically, the larger is σ_τ the fewer friendships will be observed, everything else being equal. This is the case because the realizations of τ_i will vary a substantially from individual to individual and only those with close types will become friends. Note that the intercept of the baseline utility of friendship w_0 has a similar effect on the realized number of friendships. However, I provide simulation evidence, that even in the simplest case with $n = 2$ individuals, w_0

[40] Most of the parameters are identified in the asymptotic frame, where the size of the network grows to infinity (as opposed to the number of networks going to infinity). Further discussion of this point is outside the scope of this work. For more see Xu (2011); Chandrasekhar and Jackson (2012).

is separately identified. In particular, suppose there is no heterogeneity, that the action decision is suppressed, and there are only 2 parameters: w and σ_τ. In this case three outcomes are permissible, i.e., network with 0, 1, and 2 links. Figure 4 provides evidence that the frequency of observing these outcomes will differ for different parameters. In the figure each curve plots the pair of probabilities $(\Pr(\#links = 0), \Pr(\#links = 1))$ for different w_0 and a fixed σ_τ^2. Clearly, curves with different variance parameter σ_τ^2 do not intersect. In other words different pair of parameters (w_0, σ_τ^2) induce different outcome frequencies, i.e., the model is identified.

5 Empirical Results

Table 3 presents the maximum likelihood estimates of the model parameters. The model is parametrized in a way such that each parameter corresponds to marginal probability instead of a coefficient in the alternative specific value function. For example, the parameter on the baseline utility of smoking in fact equals $\frac{e^{v_0}}{1+e^{v_0}}$. The index MP stands for marginal probability and MP% stands for marginal probability in percentages (with respect to the baseline probability).

The estimates suggest a substantial role for friends and the home environment in adolescents' decisions to smoke. In particular, one additional friend who is a smoker increases the conditional probability of smoking by 7.2 and 6.1 ppt for grades $7-8$ and $9-12$, respectively (rows $24-25$ of Table 3).[41] If 50% of the students in a school smoke, all other things being equal, then an individual is 15.2 ppt more likely to smoke (row 11). In addition, from the estimates of the parameters of the home environment, it can be seen that the presence of a smoker in the household increases the likelihood of smoking by 18.5 ppt. Moreover, rows $5-6$ demonstrate that mother education matters. Adolescents whose mothers have completed high school (college) are 5.7 (7.9) ppt less likely to smoke.

5.1 Model fit

To evaluate model fit, I simulate 10^6 replications of my sample from the sequential model using the parameter estimates and compute more than 180 statistics. Table 8 (on page 49) displays selected statistics from the data and from the simulation. Overall the model fits well the smoking decisions and the network features of the data. In particular, the model tracks the different smoking prevalences across sexes, grades and races (rows $2-16$). Moreover, as figure 5 and the lower part of Table 8 show (on pages 50 and 49), the model generates network patterns

[41] Because both friendships and smoking are choices in the model, this parameter should be interpreted with caution. In particular, the estimate cannot be interpreted as the effect on the likelihood of smoking from a randomly assigned friend who is a smoker. In such a case, the individual who is subject to this random assignment may simply drop this friendship as opposed to start smoking.

Table 3: Parameter Estimates

	Utility of smoking		
	Parameter	Estimate	(S.E.)
1	Baseline probability of smoking	0.162	(0.038)
2	Income	0.025	(0.014)
3	Price	-0.0008	(0.0004)
4	HH smokes^{MP}	0.185	(0.062)
5	Mom high school^{MP}	-0.057	(0.043)
6	Mom college^{MP}	-0.079	(0.037)
7	Males^{MP}	0.023	(0.016)
8	Grades 9-12^{MP}	0.075	(0.027)
9	Blacks^{MP}	-0.101	(0.029)
10	Contextual ζ^{MP}	0.041	(0.018)
11	50% of the school smokes^{MP}	0.151	(0.069)
	Utility of friendships		
	Parameter	Estimate	(S.E.)
12	Baseline number of friends	6.971	(1.549)
13	Different sex$^{MP\%}$	-0.322	(0.265)
14	Different grades adjacent$^{MP\%}$	-0.872	(0.307)
15	Different grades non-adjacent$^{MP\%}$	-0.908	(0.511)
16	Off-grade factor grades 9-12	0.205	(0.108)
17	Different races$^{MP\%}$	-0.552	(0.196)
18	Same race whites$^{MP\%}$	0.452	(0.18)
19	Same race blacks$^{MP\%}$	0.243	(0.092)
20	Same race AHO-WH$^{MP\%}$	0.418	(0.293)
21	Reciprocity^{MP}	0.276	(0.107)
22	$q^{MP\%}$	0.121	(0.075)
23	$q_{78}^{MP\%}$	0.057	(0.041)
24	ϕ^{MP}	0.061	(0.027)
25	ϕ_{78}^{MP}	0.072	(0.039)
26	σ_τ^2	0.197	(0.076)

Note: MP stands for the estimated marginal probability in percentage points and MP% for estimated marginal probability in percent, relative to the baseline probability. Asymptotic S.E. are in parenthesis.

Table 4: Predicted Change in Adolescent Smoking following Price Increases (in ppt)

Price increase	Model	Fixed friendships	Fixed social environment
20 cents	-2.6	-2.2	-1.4
40 cents	-4.9	-4.2	-2.3
60 cents	-7.1	-6.2	-3.5
80 cents	-8.5	-7.7	-4.4

Note: For purposes of comparison, in the data, the overall smoking prevalence is 20.8% and the mean price of a pack of cigarettes is $1.75.

that are reasonably similar to the ones observed in the data.

5.2 Policy experiments

A. Changes in the price of tobacco

I use the estimated model to perform a number of counterfactual policy experiments.[42] First, I simulate adolescent reaction to increases in tobacco prices in the amount of 20, 40, 60, and 80 cents. Table 4 presents the results. The first column reports the model's prediction for the change in overall smoking prevalence in percentage points. The second column does so in the situation where friendship decisions are suppressed, i.e., individuals are allowed to alter their smoking statuses but not their friendships. The third column displays the model's prediction whenever all dimensions of the social environment are kept fixed. Note that this entails stronger restrictions on individuals' social environments than just keeping friendships fixed. In particular, when simulating individuals' smoking decisions, I keep constant their friendship choices, their friends' smoking statuses, and the average smoking behavior of the population overall (i.e., the number of smokers in the population).

Table 4 shows that adolescents are sensitive to tobacco prices (the mean price of a pack of cigarettes is $1.75 in the sample), and that the social interactions magnify the tax effects substantially. In particular, the predicted decrease in smoking prevalence is about double because of the additional incentives working through the change in the social environment (compare the first and the third columns). In addition, neglecting the endogeneity of the friendship network leads to a downward bias of 10 to 15% on the predicted effect (compare the first and the second columns). Finally, I also simulate a 10% price increase, which gives an approximate estimate of the elasticity of demand for tobacco of -0.98. As before, if one conditions on the existing friendship nominations, i.e., restricts adolescents' choices in the model only to smoking, the

[42]To assess the effect of each proposed policy, I consider 10^6 draws from the stationary distribution of the model.

estimated elasticity is -0.87.

B. Changes in the racial composition of schools

I use the estimated model to analyze how changes in the racial composition of schools affect adolescent smoking. In particular, I am interested in the following thought experiment. Suppose that in city X there are two schools: the "White School" and the "Black School." As their names suggest, both schools are racially segregated: the White School consists of only Whites, and the Black School consists of only Blacks. Now consider a policy aiming to promote racial desegregation, which prevents schools from enrolling more than x percent of students of the same race. Will this policy impact the overall prevalence of smoking?

To simulate the effect of the proposed policy I consider one of the racially balanced schools in my sample.[43] The Whites and the Blacks from this school serve as prototypes for the White School and Black School respectively.[44] To implement the proposed policy I randomly select a set of students from the White School and a set of students from the Black School and swap them. Then I use the estimated model to simulate the smoking (and friendship) decisions in the newly formed schools.

Table 5 presents the simulation results, which suggest that racial composition affects the overall smoking prevalence.[45] The first column shows the size of the set of students which is being swapped. For example to simulate the effect of a 70% cap on the same-race students in a school, I need to simulate a swap of 30%. The second, third, and forth columns show the simulated smoking prevalences in the White School, Black School, and both, respectively. The table suggest that the overall smoking prevalence in racially segregated schools (the first three rows of the last column) is higher than that of racially desegregated schools (the bottom three rows of the last column).[46] Such a finding provides empirical support for policies promoting racial integration in the context of fighting high smoking rates.

[43] The school has 150 students of which 40% are Whites and 42% are Blacks. It incorporates students from grades 7 to 12.

[44] As an alternative to splitting one school into two racially segregated schools, one could consider two schools from the data that are already racially segregated. However, the only school with a high ratio of Blacks in the sample incorporates students from grades 7 and 8. If this school serves as a prototype for the Black School, then I am faced with two options for the choice of the White School. If the White School incorporates only 7th and 8th graders, then smoking prevalence will be low regardless, since these grades are mostly nonsmoking. Alternatively, if the White School incorporates higher grades, the simulation results will be driven in part by the asymmetry in the population (7th and 8th graders do not make friends with students from older grades; recall the mixing matrix in Table 2). Consequently, the school that incorporates black students in grades 7 and 8 only cannot properly serve as a prototype for the Black School.

[45] The policy is simulated 100 times, where each time new (same-size) sets of students from the White School and the Black School are drawn.

[46] However, an equal proportion of races in schools (i.e., cap $x = 50\%$) does not seem to lead to the lowest overall smoking prevalence.

Table 5: Predicted Smoking Prevalence following Same-race Students Cap

% swapped	White School	Black School	Overall
0	0.322	0.035	0.173
10	0.295	0.041	0.163
20	0.281	0.061	0.167
30	0.224	0.069	0.144
40	0.223	0.095	0.157
50	0.175	0.162	0.168

Note: A cap of $x\%$ same-race students is implemented with a swap of $(100 - x)\%$ students.

C. Separation of middle school and high school grades

There are 8 schools in the estimation sample that incorporate grades 7 through 12. For these schools, I simulate the effect on smoking choices of separating middle school students (grades 7-9) from high school students (grades 10-12) in different schools. This scenario involves severing the existing friendships between middle school and high school students and restricting the model to not allow the formation of new ones across grades that are not in the same school. The overall smoking prevalence for these schools is 25.3%. If the proposed policy is implemented, the model predicts that in middle schools, smoking decreases by 2.4 ppt, while in high schools it increases by 0.3 ppt. (overall smoking drops by 0.9 ppt) The difference in the effect of this policy in middle schools versus high schools can be explained mainly by the difference in the change in the aggregate peer effects. While in middle schools the separation will likely decrease the base of students who are likely to smoke (without any peer norm considerations), the opposite is true for high schools.

D. Cascade effects of an anti-smoking campaign

In the school with the highest smoking prevalence in the sample, 44.7% of its students smoke. For this school, I consider the effects of an anti-smoking campaign that can prevent with certainty a fixed number of students from smoking. For example, suppose that a fixed number of students are invited to a weekend-long information camp on the health consequences of smoking. The camp is very effective in terms of preventing students from smoking; however, it is too costly to invite all students. For the sake of the analysis, suppose that all students who attend, independently of whether they are currently smokers, with certainty will not smoke after the camp. I use the estimated model to assess how interactions between students who attended the camp and those who did not impact the overall effect of the campaign.

Table 6 presents the simulation results, which suggest that an anti-smoking campaign may

Table 6: The Effects of an Anti-smoking Campaign

Attendance rate	Change in number of smokers	Spillover effects
3%	-3.3%	1.11
5%	-5.9%	1.18
10%	-14.1%	1.41
20%	-30.2%	1.51
50%	-72.5%	1.45

Note: To illustrate the cascade effects, I assume that all students who attend the anti-smoking camp, independently of whether they currently smoke, with certainty will not smoke after.

have a large impact on the overall prevalence of smoking, without necessarily being able to directly engage a large part of the student population.[47] The first column lists the alternative attendance rates I consider. The second column displays the mean change in the number of smokers after the policy is in place. The last column displays the ratio between the percentage change in the number of smokers and the attendance rate. Observe that, in the absence of spillover effects,[48] the percentage decrease in the smoking rate will follow the attendance rates, i.e., the spillover effects will be constant and equal to 1. Table 6 shows that this is not the case. Whenever the attendance rates are large enough, there are substantial spillover effects from those who attended the camp to the rest of the school.

E. Effects of home environment

Finally, I use the model to analyze how changes in the family environment affect the prevalence of smoking. In 37.7% of the adolescent families, there is a smoker at home. The model predicts that if the family members who smoke at home quit, adolescent smoking will drop by 8.7 ppt. On the other hand, 77% of adolescent mothers in the sample have at least a high school education (20% of them have completed college). The model predicts that if all mothers acquire at least a high school education, adolescent smoking will drop by 2.7 ppt. The magnitude of these results suggests that the root of high smoking rates may indeed be adolescents' family environment.

6 Concluding remarks

In this paper, I propose a framework to study social interactions and use it to analyze the determinants of teen smoking. In contrast to previous literature, the framework allows both

[47]The policy is simulated 100 times, where each time a new random draw of attendees is being considered.

[48]I also assume that attendance is random with respect to the smoking status of the students. Observe that if the campaign is able to target only students who are currently smokers, the spillover effects will be even larger, i.e., the support for the conclusion I draw will be even stronger.

friendships and actions to be jointly chosen. I structurally estimate the model with data from the National Longitudinal Study of Adolescent Health and demonstrate that it replicates prominent features of social networks, such as the tendency of agents who make similar choices to be friends and the fact that individuals who share a common friend are usually friends themselves. The proposed framework enables the analysis of novel policies targeting teen smoking, such as school racial desegregation, separating middle from high school grades, and anti-smoking campaigns and their cascade effects. The estimated model is used to shed light on the importance of modeling the joint decisions of friendships and cigarette smoking. I find that neglecting the endogeneity of the friendship network leads to downward bias in the model's predictions of 10% to 15%.

This research addresses existing identification concerns in the literature on social interactions and opens up exciting avenues for future research. With more granular data on friendship networks, possibly including historical data on how friendships evolve over time, the strategic foundations of this paper can be extended to address new set of questions. In the spirit of Currarini et al. (2009), one can ask how friendship searches and information frictions shape an individual's social environment. More broadly, this line of work enables researchers to examine individual behaviors and to evaluate feasible policy alternatives in situations where the mechanism for intervention is likely to operate through existing social norms and, more generally, through the structure of the social network.

References

Ali, Mir M and Debra S Dwyer, "Estimating peer effects in adolescent smoking behavior: a longitudinal analysis.," *Journal of Adolescent Health*, 2009, *45* (4), 402–408.

Amemiya, Takeshi, "Qualitative response models: A survey," *Journal of Economics Literature*, 1981, *19* (4), 1483–1536.

Aumann, Robert, "Subjectivity and correlation in randomized strategies," *Journal of Mathematical Economics*, 1974, *1*, 67–96.

Bajari, Patrick, Han Hong, and Stephen P Ryan, "Identification and Estimation of a Discrete Game of Complete Information," *Econometrica*, 2010, *78* (5), 1529–1568.

Bala, Venkatesh and Sanjeev Goyal, "A Noncooperative Model of Network Formation," *Econometrica*, 2000, *68* (5), 1181–1229.

Beckmann, Martin, C. B. McGuire, and Christopher B. Winsten, *Studies in Economics of Transportation*, Yale University Press, 1956.

Berry, Steven T, "Estimation of a Model of Entry in the Airline Industry," *Econometrica*, 1992, *60* (4), 889–917.

Bisin, Alberto, Andrea Moro, and Giorgio Topa, "The Empirical Content of Models with Multiple Equilibria in Economies with Social Interactions," *NYU mimeo*, 2011.

Blume, Lawrence E., "The Statistical Mechanics of Strategic Interaction," *Games and Economic Behavior*, 1993, *5*, 387–424.

_ , William A. Brock, Steven N. Durlauf, and Yannis M. Ioannides, "Identification of Social Interactions," *Handbook of Social Economics*, 2010, *1* (260), 213–228.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin, "Identification of peer effects through social networks," *Journal of Econometrics*, May 2009, *150* (1), 41–55.

Bresnahan, Timothy F and Peter C Reiss, "Entry in Monopoly Markets," *The Review of Economic Studies*, 1990, *57* (4), 531–553.

Brock, William A. and Steven N. Durlauf, "Chapter 54 Interactions-based models," *Handbook of Econometrics*, 2001, *5*, 3297–3380.

_ and _ , "Identification of binary choice models with social interactions," *Journal of Econometrics*, September 2007, *140* (1), 52–75.

Calvó-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou, "Peer Effects and Social Networks in Education," *Review of Economic Studies*, 2009, *76* (4), 1239–1267.

Card, David and Laura Giuliano, "Peer Effects and Multiple Equilibria in the Risky Behavior of Friends," *Health San Francisco*, 2011, *No. 17088.*

Case, Anne C. and Lawrence F. Katz, "The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youths," *NBER Working Paper No. 3705*, 1991.

Chandrasekhar, Arun and Matthew O. Jackson, "Tractable and Consistent Random Graph Models," *Stanford Working Paper*, 2012.

Ciliberto, Federico and Elie Tamer, "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 2009, *77* (6), 1791–1828.

Cournot, Augustin, *Researches into the Mathematical Principles of the Theory of Wealth*, The Macmillan Company, 1838.

Currarini, Sergio, Matthew O. Jackson, and Paolo Pin, "An Economic Model of Friendship: Homophily, Minorities, and Segregation," *Econometrica*, 2009, *77*, 1003–1045.

De Giorgi, Giacomo, Michele Pellizzari, and Silvia Redaelli, "Identification of social interactions through partially overlapping peer groups," *American Economic Journal: Applied Economics*, 2010, *2*, 241–75.

Durlauf, Steven N and Yannis M Ioannides, "Social Interactions," *Annual Review of Economics*, 2010, *2* (1), 451–478.

Dutta, Bhaskar, Sayantan Ghosal, and Debraj Ray, "Farsighted network formation," *Journal of Economic Theory*, June 2005, *122* (2), 143–164.

Foster, Dean and Peyton Young, "Stochastic evolutionary game dynamics," *Theoret. Population Biol*, 1990.

Geman, Stuart and Donald Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1984, *PAMI-6* (6), 721–741.

Geyer, Charles J. and Elizabeth A. Thompson, "Constrained Monte Carlo Maximum Likelihood for Dependent Data," *Journal of the Royal Statistical Society*, 1992, *Series B* (54), 657–699.

Gilles, Robert P. and Sudipta Sarangi, "Social Network Formaiton with Consent," *Discussion Paper, Tilburg University*, 2004.

Gilleskie, Donna B. and Yun S. Zhang, "Friendship Formation and Smoking Initiation Among Teens," *UNC-Chapel Hill*, 2010.

Giorgi, Giacomo De and Michele Pellizzari, "Understanding Social Interactions: Evidence from the Classroom," Working Papers 385, IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University 2011.

Goldsmith-Pinkham, Paul and Guido Imbens, "Social Networks and the Identification of Peer Effects," *Working paper*, 2011.

Goyal, Sanjeev and Fernando Vega-Redondo, "Network formation and social coordination," *Games and Economic Behavior*, February 2005, *50* (2), 178–207.

Gumbel, Emil Julius, *Statistics of Extremes*, Columbia University Press, New York, 1958.

Hartmann, Wesley R, "Demand Estimation with Social Interactions and the Implications for Targeted Marketing," *Marketing Science*, 2010, *29* (4), 585–601.

Heckman, James J., "Dummy Endogenous Variables in a Simultaneous Equation System," *Econometrica*, 1978, *46* (4), 931–959.

_ **and Edward Vytlacil**, "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 05 2005, *73* (3), 669–738.

Hsieh, Chih-Sheng and Lung-Fei Lee, "A Social Interactions Model with Endogenous Friendship Formation and Selectivity," *Working paper*, 2011.

Jackson, Matthew O., *Social and Economics Networks*, Princeton, N.J., 2008.

_ **and Alison Watts**, "On the formation of interaction networks in social coordination games," *Games and Economic Behavior*, November 2002, *41* (2), 265–291.

_ **and Asher Wolinsky**, "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory*, 1996, *71* (0108), 44–74.

Kandori, Michihiro, George J. Mailath, and Rafael Rob, "Learning, Mutation, and Long Run Equilibria in Games," *Econometrica*, January 1993, *61* (1), 29–56.

Kessler, DA, AM Witt, PS Barnett, MR Zeller, SL Natanblut, JP Wilkenfeld, CC Lorraine, LJ Thompson, and WB Schultz, "The Food and Drug Administration's Regulation of Tobacco Producs," *New England Journal of Medicine*, September 1996, *13* (335), 988–94.

Köenig, Michael, Claudio Tessone, and Yves Zenou, "Nestedness in Networks: A Theoretical Model and Some Applications," Discussion Papers 11-003, Stanford Institute for Economic Policy Research January 2012.

Krauth, Brian, "Structural estimation of peer effects in youth smoking.," *Advances in Health Economics and Health Services Research*, 2005, *16*, 201–211.

Krauth, Brian V, "Peer effects and selection effects on smoking among Canadian youth," *Canadian Journal of EconomicsRevue canadienne dconomique*, 2005, *38* (3), 735–757.

Lavy, Victor and Analía Schlosser, "Mechanisms and Impacts of Gender Peer Effects at School," *American Economic Journal Applied Economics*, 2011, *3* (October 2006), 1–33.

Lehmann, E L and George Casella, *Theory of Point Estimation*, Vol. 31, Springer, 1998.

Leibenstein, Harvey, "Bandwagon, Snob, and Veblen Effects in the Theory of Consumers' Demand," *The Quarterly Journal of Economics*, 1950, *64*, 183–207.

Liang, Lan, Frank Chaloupka, Mark Nichter, and Richard Clayton, "Prices, policies and youth smoking," *Addiction Abingdon England*, 2001, *98 Suppl 1* (May 2001), 105–122.

Manski, Charles F., "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, 1993, *60*, 531–42.

Marschak, Jacob, "Binary choice constraints on random utility indications," in Kenneth Arrow, ed., *Stanford Symposium on Mathematical Methods in the Social Sciences*, 1960, pp. 312–329.

Mazzeo, Michael J, "Product Choice and Oligopoly Market Structure," *The RAND Journal of Economics*, 2002, *33* (2), 221.

McFadden, D, "Conditional logit analysis of qualitative choice behavior," *Frontiers in Econometrics*, 1974, *1* (2), 105–142.

McKelvey, Richard D. and Thomas R. Palfrey, "Quantal Response Equilibrium for Normal Form Games," *Games and Economic Behavior*, 1995, *10.*

Mele, Angelo, "A Structural Model of Segregation in Social Networks," *UIUC-mimeo*, 2010.

Moffitt, Robert A, "Policy Interventions, Low-Level Equilibria, and Social Interactions," in Steven N Durlauf and H Peyton Young, eds., *Social Dynamics*, MIT Press, 2001, pp. 45–82.

Monderer, Dov and Lloid S. Shapley, "Potential Games," *Games and Economic Behavior*, 1996, *14*, 124–143.

Nakajima, Ryo, "Measuring Peer Effects on Youth Smoking Behavior," *The Review of Economic Studies*, 2007, *74*, 897–935.

Narayanan, Sridhar, "Bayesian Estimation of Discrete Games of Complete Information," *conditionally accepted, Quantitative Marketing and Economics*, 2011.

Plantenga, Todd D., "HOPSPACK 2.0 User Manual," Technical Report SAND2009-6265, Sandia National Laboratories, Albuquerque, NM and Livermore, CA October 2009.

Powell, Lisa M, John A Tauras, and Hana Ross, "The importance of peer effects, cigarette prices and tobacco control policies for youth smoking behavior.," *Journal of Health Economics*, 2005, *24* (5), 950–968.

Reka, Albert and Barabási, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, June 2002, *74*, 47–97.

Richards-Shubik, Seth, "Peer Effects in Sexual Initiation: Separating Demand and Supply Mechanisms," *Carnegie Mellon Mimeo*, 2011.

Robert, Christian P. and George Casella, *Monte Carlo Statistical Methods*, Springer, 2010.

Rosenthal, Robert W., "A Class of Games Posessing a Pure-Strategy Nash Equilibria," *International Journal of Game Theory*, 1973, *2* (65-67).

Sacerdote, Bruce, "Peer Effects With Random Assignment: Results For Dartmouth Roommates," *The Quarterly Journal of Economics*, May 2001, *116* (2), 681–704.

Snijders, Tom A. B., "Markov Chain Monte Carlo Estimation of Exponential Random Graph Models," *Journal of Social Structure*, 2002, *3* (2), 1–40.

Staudigl, Mathias, "Potential games in volatile environments," *Games and Economic Behavior*, May 2011, *72* (1), 271–287.

_ , "Stochastic stability in asymmetric binary choice coordination games," *Games and Economic Behavior*, 2012, *75* (1), 372–401.

Tamer, Elie, "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *The Review of Economic Studies*, 2003, *70* (1), 147–165.

Thurstone, L L, "A law of comparative judgment," *Psychological Review*, 1927, *34* (4), 273–286.

Xu, Haiqing, "Social Interactions: A Game Theoretical Approach," *UT-Austin mimeo*, 2011.

Add Health Data

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. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

A Proofs

PROOF (LEMMA 1 (ON P. 14)) For part (i), if S is k-player Nash stable then by definition $S|_{I_r}$ is a Nash equilibrium of $\Gamma|_{I_r}$ for any I_r such that $|I_r| = k$. But it is easy to see that if no player has incentive to deviate in $\Gamma|_{I_r}$ then no player have incentive to deviate in $\Gamma|_{I'_r}$ where $I'_r \subseteq I_r$ and $r' < r$. This in turn establishes that any partition with maximal component of size k is Nash stable. The converse follows mutatis mutandis.

One way to see (ii) is with the result we just established. If a state is partition Nash stable with respect to any partition with maximum component of size k, it will certainly be partition Nash stable with respect to a finer partition with smaller maximum component. ∎

PROOF (PROPOSITION 1 (ON P. 15)) It suffices to verify two conditions.

Condition A. For any i, S_{-i} and X

$$\mathcal{P}(S', X) - \mathcal{P}(S, X) = u_i(S', X) - u_i(S, X) \tag{11}$$

where $S, S' \in \mathbf{S}$ are defined as $S = (a_i = 0, S_{-i})$ and $S' = (a_i = 1, S_{-i})$.

Condition B. For any $i \neq j$, S_{-ij} and X

$$\mathcal{P}(S', X) - \mathcal{P}(S, X) = u_i(S', X) - u_i(S, X) \tag{12}$$

where $S, S' \in \mathbf{S}$ are defined as $S = (g_{ij} = 0, S_{-ij})$ and $S' = (g_{ij} = 1, S_{-ij})$.

With the conditions of assumption 1 these are trivial to verify. Substitute for $\mathcal{P}(.,.)$ in the left-hand-side of condition (11):

$$\begin{aligned}\mathcal{P}(S',X)-\mathcal{P}(S,X) &= v(X_i)+\frac{1}{2}\sum_{j\neq i}a_j g_{ij}g_{ji}(\phi_{ij}+\phi_{ji})+\frac{1}{2}\sum_{j\neq i}a_j(h(X_i,X_j)+h(X_j,X_i)) \\ &= u_i(a_i=1,S_{-i},X)-u_i(a_i=0,S_{-i},X)\end{aligned}$$

The verification of condition (12) proceeds similarly. ■

Proof (Proposition 3 (on p. 18) That any k-player Nash stable state is absorbing follows from the definition of S^*_{k-NS}. The second part follows from the following claim:

Claim 1 *The sequence of the values of the potential $\mathcal{P}_t$ induced by the outcome of a meeting process of dimension k is a submartingale. i.e.*

$$E[\mathcal{P}_{t+1}|S_t]\geq\mathcal{P}_t \tag{13}$$

So that $\{\mathcal{P}_t\}$ converges almost surely. Since the network is of finite size it follows that $\{\mathcal{P}_t\}$ is constant for large t. Because of assumption 2 this can happen only if $S_t\in S^*_{k-CS}$. ■

Proof (Markov property) First note that $\Pr(S_{t+1}|S_t)>0$ only if S_{t+1} differs from S_t by the play of one agent w.r.t. at most $k-1$ of her friendships and her action. For those transitions, the probability equals the product of the probabilities of two events - the probability that appropriate meeting will realize and the probability that the appropriate change in friendships/action is going to be optimal. Both of these probabilities are functions of S_t only (vs $\{S_{t'}\}_{t'=1}^{t}$) and whence the chain exhibits the Markov property. ■

Proof (Theorem 1 (on p. 20) The *existence* of equilibrium (aka stationary, invariant) distribution over the state space $\mathbf{S}$ is trivial algebraic property as the (finite dimensional) transition matrix has a row eigen vector with eigen value of 1. The rest of the properties can be established as a consequence of the following claim.

Claim 2 *The Markov chain of network configurations $(S_t)_{t\geq 0}$ induced by the matching process $\{\mu_t\}$ and the optimal choice of individuals is: (i) irreducible, (ii) positive recurrent, and (iii) aperiodic.*

Recall that the state space is *finite*. Fix the current state to S. Since the preference errors have unbounded support any state S' in the neighborhood of S - $\mathbf{N}_S$, communicates with S. Since all states can communicate with each other the Markov chain has a single class i.e. it is *irreducible*. The class is trivially closed and whence *recurrent* (Every finite closed class is recurrent); that is the probability that every state will be revisited infinitely often equals one.

The chain is trivially aperiodic: there exists N such that $\Pr(S_{t+n} = S'|S_t = S) > 0$ for all $n > N$ and $S, S' \in \mathbf{S}$. This completes the proof of the claim.

The stationary distribution π is unique as the chain is irreducible and (positive) recurrent. The last two properties are related to convergence to the stationary distribution. Since in addition the chain is aperiodic

$$\Pr(S_t = S) \longrightarrow \pi(S)$$

that is S_t converges to the stationary distribution π no matter where the chain starts. Finally since the chain is irreducible and positive recurrent for any (trivially bounded on $\mathbf{S}$) $f : \mathbf{S} \longrightarrow \mathbb{R}$ we have the conclusion of the ergodic theorem:

$$\frac{1}{n}\sum_{t=0}^{T} f(S_t) \xrightarrow{a.s.} \overline{f}$$

where $\overline{f} = \int f(S)\, d\pi$. ∎

Proof (Theorem 2 (p. 21)) The proof uses the so called detailed balance property of Markov process (S): if π and (S) are in detailed balance, then π is invariant for (S). Fix $2 \le k \le n$ and let $S, S' \in \mathbf{S}$. Denote the transition probability $\Pr(S_{t+1} = S'|S_t = S)$ with $P_k(S'|S)$. It suffices to demonstrate that the detailed balance condition holds:

$$P_k(S'|S)\pi(S) = P_k(S|S')\pi(S')$$

for all $S, S' \in \mathbf{S}$. First note that because the dimension of the meeting process is fixed, $P_k(S|S') > 0$ if and only if $P_k(S'|S) > 0$. Indeed, a state S' can be reached from S in one period if it differs by only the play of one agent w.r.t. at most $k-1$ of her friendships and her action. Consequently if $P_k(S|S') = 0$ then $P_k(S'|S) = 0$ and the detailed balance condition trivially holds.

Suppose that $P_k(S'|S) > 0$ and S and S' differ w.r.t. the play of individual i to the set $I^* = \{j \in I : g_{ij} \neq g'_{ij}\}$ of her friends. The transition low $P_k(S|S')$ can be derived as

$$\begin{aligned}
P_k(S'|S) &= \sum_{\tilde{I} \supseteq I^*, |\tilde{I}| \le k} Pr\{\mu_t = (i, \tilde{I})|S\} \frac{\exp\{u_i(S')/\underline{\beta}\}}{\sum_{S_{(i|\tilde{I}_k)} \in \mathbf{S}_{(i|\tilde{I}_k)}} \exp\left\{u_i\left(S_{(i|\tilde{I}_k)}, S \backslash S_{(i|\tilde{I}_k)}\right)/\underline{\beta}\right\}} \\
&= \sum_{\tilde{I} \supseteq I^*, |\tilde{I}| \le k} Pr\{\mu_t = (i, \tilde{I})|S\} \frac{\exp\{\mathcal{P}(S')/\underline{\beta}\}}{\sum_{S_{(i|\tilde{I}_k)} \in \mathbf{S}_{(i|\tilde{I}_k)}} \exp\left\{\mathcal{P}\left(S_{(i|\tilde{I}_k)}, S \backslash S_{(i|\tilde{I}_k)}\right)/\underline{\beta}\right\}}
\end{aligned}$$

The first row above follows from the assumption that the error term follows *Gumbel*$(\underline{\mu}, \underline{\beta})$ distribution, while the second row follows from the properties of the potential function. From

Table 7: Example non-factorizability of π.

	$g_{21}=0$	$g_{21}=1$
$g_{12}=0$	1	1
$g_{12}=1$	1	$\exp\{w\}$

analogous expression for $P_k(S'|S)$ it can be seen that

$$\frac{P_k(S|S')}{P_k(S'|S)} = \frac{\exp\{\mathcal{P}(S)/\underline{\beta}\}}{\exp\{\mathcal{P}(S')/\underline{\beta}\}}$$

which completes the proof because by definition $\frac{\pi(S)}{\pi(S')} = \frac{\exp\{\mathcal{P}(S)/\underline{\beta}\}}{\exp\{\mathcal{P}(S')/\underline{\beta}\}}$. ■

Proof (Theorem 3 (p. 22) The proof follows immediately from the expression for the stationary distribution obtained in theorem 2 ■

Proof (Proposition 4 (p. 23) I will proceed by a way of contradiction. Consider a network with $n = 2$ players and suppose all coefficients except m are set to zero. In addition, consider the subspace $\overline{\mathbf{S}}$ of $\mathbf{S}$ consisting of the product of the linking strategies only:

$$\overline{\mathbf{S}} = \{g_{12} = 0, g_{12} = 1\} \times \{g_{21} = 0, g_{21} = 1\}.$$

Table 8 shows the distribution π conditional on $a_1 = a_2 = 0$ on $\overline{\mathbf{S}}$ up to a normalizing factor. Clearly this matrix is full rank (its determinant is nonzero provided $w \neq 0$) and thus cannot be factored into two independent marginals (i.e., play where individuals 1 and 2 randomize independently). ■

B Additional tables and graphs

C Technical background

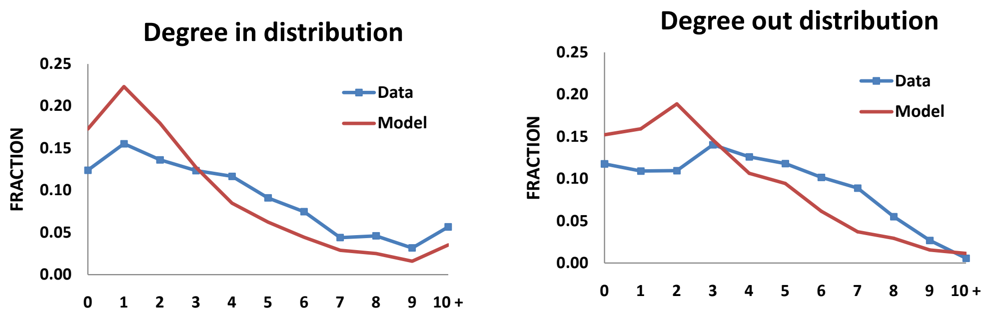
Network metrics

Freeman segregation index (FSI) Freeman (1972) proposed an index to measure the level of segregation between two groups in a social network. To illustrate the basic intuition behind the index, consider a population of n individuals some of which have action status $a_i = 0$ and some of which with action status $a_i = 1$. The index is designed to measure how far are the links between groups 0 and 1 in a network from being drawn at random with respect to group belonging. Let M_{01} be the number of links between individuals from group 0 to individuals

Table 8: Model Fit: Selected Statistics

	Smoking prevalence		
	Statistics	Data	Model
1	Overall	0.208	0.210
2	Middle school (grades 7-9)	0.156	0.163
3	High school (grades 10-12)	0.297	0.288
4	Males	0.234	0.235
5	Females	0.184	0.186
6	Grade 7	0.143	0.127
7	Grade 8	0.120	0.157
8	Grade 9	0.257	0.258
9	Grade 10	0.281	0.282
10	Grade 11	0.268	0.264
11	Grade 12	0.349	0.316
12	Whites	0.239	0.244
13	Blacks	0.055	0.048
14	As-Hi-Ot	0.173	0.164
	Network topology		
	Statistics	Data	Model
15	FSI sex	0.174	0.185
16	FSI grade 7	0.784	0.716
17	FSI grade 12	0.494	0.585
18	FSI Whites	0.115	0.095
19	FSI Blacks	0.137	0.114
20	FSI smoking status	0.092	0.061
21	Clustering	0.411	0.258
22	Nominations per person	3.361	3.027
23	Reciprocity	0.383	0.336

Figure 5: Model Fit: Degree Distribution



Note: For individual i, degree in is defined as the number of friendships for which i is the receiver of a nomination. Similarly, degree out is the number of friendships for which i is the sender of a nomination.

from group 1 and, similarly, let M_{10} be the number of links between individuals from group 1 to individuals from group 0. Freeman proposed to calculate the difference between expected links, if drawn at random, and actual links of the form

$$FSI = 1 - \frac{M_{01} + M_{10}}{\mathbb{E}[M_{01} + M_{10}]}$$

Note that $FSI = 0$ implies that links are drawn at random, while $FSI = 1$ is consistent with no inter-group links (only intra-group links). Also higher FSI implies more segregation.

In practical terms the index can be calculated as following. Consider a network state $S = (\{a_i\}, \{g_{ij}\})$ and define the $n \times n$ adjacency matrix (sociomatrix) $G = (g_{ij})$ and the $n \times 2$ state indicator matrix A, with rows given by $(\mathbb{1}_{\{a_i=0\}}, \mathbb{1}_{\{a_i=1\}})$. Here $\mathbb{1}_{\{a_i=0\}}$ is the indicator function on the action status space for the event $a_i = 0$. More simply any row in A is either $(1, 0)$ (if $a_i = 0$) or $(0, 1)$ (if $a_i = 1$). The mixing matrix for action status is then given by $M = A'GA$ and can be written as:

$$\begin{bmatrix} M_{00} & M_{01} \\ M_{10} & M_{11} \end{bmatrix}$$

Finally, it can be shown that

$$\mathbb{E}[M_{01} + M_{10}] = \frac{(M_{00} + M_{01})(M_{01} + M_{11}) + (M_{10} + M_{11})(M_{00} + M_{10})}{M_{00} + M_{01} + M_{10} + M_{11}} \tag{14}$$

Clustering coefficient In graph theory, a clustering coefficient is a measure of degree to which nodes in a graph (in our cases individuals in a social network) tend to cluster together. The clustering coefficient C_i of an individual i is the number of friendship triangles i is involved in divided by the number of all possible triangles she could be involved, given her number of friends

$$C_i = \frac{\textsf{number of friendships bw i's friends}}{\textsf{maximum number of friendships bw i's friends}} \tag{15}$$

Gumbel distribution

The Gumbel distribution, also known as log-Weibull, double exponential, and type I extreme value, is a two parameter family of continuous probability distributions. Its probability density function $f(.|\mu, \beta)$ and cumulative distribution function $F(.|\mu, \beta)$ are:

$$\begin{aligned} f(x|\mu, \beta) &= \frac{1}{\beta} \exp\{-z - \exp\{-z\}\} \\ F(x|\mu, \beta) &= \exp\{-\exp\{-z\}\} \end{aligned}$$

where $z = \frac{x-\mu}{\beta}$. Its mode, median, and mean are μ, $\mu - \beta \ln(\ln 2)$, and $\mu + \gamma\beta$ respectively ($\gamma \approx 0.5772$ is the Euler-Mascheroni constant); Its variance is $V(X) = \frac{\beta^2\pi^2}{6}$. The following property of the Gumbel distribution makes it particularly attractive one in the analysis of qualitative response (aka quantal, categorical, or discrete) models.

Lemma 2 *Let $V_i \in \mathbb{R}$ for $i = 1, ..., n$ and ϵ_i are i.i.d. Gumbel(μ, β) random variables. For $X_i = V_i + \epsilon_i$,*

$$\Pr(X_i = \max_j X_j) = \frac{\exp\{V_i/\beta\}}{\sum_{j=1}^n \exp\{V_j/\beta\}} \tag{16}$$

The proof proceeds through a sequence of algebraic steps. First note that $\Pr(X_i = \max_j X_j) = \Pr(\epsilon_j < V_i - V_j + \epsilon_i, \quad \forall j \neq i) = \mathbb{E}_{\vec{\epsilon}}\left[\prod_{j\neq i} \chi_{\epsilon_j < V_i - V_j + \epsilon_i}(\epsilon_j)\right]$. Since ϵ_i are independent:

$$
\begin{aligned}
\Pr(X_i = \max_j X_j) &= \mathbb{E}_{\epsilon_i}\left[\mathbb{E}_{\vec{\epsilon}|\epsilon_i}\left(\prod_{j\neq i} \chi_{\{\epsilon_j < V_i - V_j + \epsilon_i\}}(\epsilon_j)\right)\right] \\
&= \int \left[\prod_{j\neq i} F(\epsilon_j < V_i - V_j + \epsilon_i)\right] f(\epsilon_i) d\epsilon_i \\
&= \int \exp\left\{-\sum_{j\neq i} \exp\left\{-\frac{V_i - V_j}{\beta} - z\right\}\right\} \exp\left\{-z - \exp\left\{-z\right\}\right\} dz \\
&= \int \exp\left\{-\exp\left\{-z\right\} \underbrace{\sum_{j\neq i} \exp\left\{-\frac{V_i - V_j}{\beta}\right\}}_{v}\right\} \exp\left\{-z\right\} \exp\left\{-\exp\left\{-z\right\}\right\} dz \\
&= \int_{-\infty}^{0} \exp\left\{z' v\right\} \exp\left\{z'\right\} dz' \\
&= \frac{\exp\{V_i/\beta\}}{\sum_j \exp\{V_j/\beta\}}
\end{aligned}
$$

where $z = \frac{\epsilon_i - \mu}{\beta}$, $z' = -\exp\{-z\}$ and $\chi_{\{A\}}(.)$ is the characteristic function of the set A.