



12-2008

## Modeling social interactions: Identification, empirical methods and policy implications

Wesley R. Hartmann

Puneet Manchanda

Harikesh Nair

Matthew Bothner

Peter Dodds

*See next page for additional authors*

Follow this and additional works at: [https://repository.upenn.edu/oid\\_papers](https://repository.upenn.edu/oid_papers)



Part of the [Interpersonal and Small Group Communication Commons](#), [Other Communication Commons](#), [Other Social and Behavioral Sciences Commons](#), and the [Social Psychology and Interaction Commons](#)

---

### Recommended Citation

Hartmann, W. R., Manchanda, P., Nair, H., Bothner, M., Dodds, P., Godes, D., Hosanagar, k., & Tucker, C. (2008). Modeling social interactions: Identification, empirical methods and policy implications. *Marketing Letters*, 19 (3), 287-304. <http://dx.doi.org/10.1007/s11002-008-9048-z>

This paper is posted at ScholarlyCommons. [https://repository.upenn.edu/oid\\_papers/81](https://repository.upenn.edu/oid_papers/81)  
For more information, please contact [repository@pobox.upenn.edu](mailto:repository@pobox.upenn.edu).

---

# Modeling social interactions: Identification, empirical methods and policy implications

## Abstract

Social interactions occur when agents in a network affect other agents' choices directly, as opposed to via the intermediation of markets. The study of such interactions and the resultant outcomes has long been an area of interest across a wide variety of social sciences. With the advent of electronic media that facilitate and record such interactions, this interest has grown sharply in the business world as well. In this paper, we provide a brief summary of what is known so far, discuss the main challenges for researchers interested in this area, and provide a common vocabulary that will hopefully engender future (cross disciplinary) research. The paper considers the challenges of distinguishing actual causal social interactions from other phenomena that may lead to a false inference of causality. Further, we distinguish between two broadly defined types of social interactions that relate to how strongly interactions spread through a network. We also provide a very selective review of how insights from other disciplines can improve and inform modeling choices. Finally, we discuss how models of social interaction can be used to provide guidelines for marketing policy and conclude with thoughts on future research directions.

## Keywords

social interactions, networking, social multiplier, peer effects

## Disciplines

Interpersonal and Small Group Communication | Other Communication | Other Social and Behavioral Sciences | Social Psychology and Interaction

## Author(s)

Wesley R. Hartmann, Puneet Manchanda, Harikesh Nair, Matthew Bothner, Peter Dodds, David Godes, kartik Hosanagar, and Catherine Tucker

# **Modeling Social Interactions: Identification, Empirical Methods and Policy Implications**

Wesley R. Hartmann (co-chair), Stanford University  
Puneet Manchanda (co-chair), University of Michigan  
Harikesh Nair (co-chair), Stanford University  
Matthew Bothner, University of Chicago  
Peter Dodds, University of Vermont  
David Godes, Harvard University  
Kartik Hosanagar, University of Pennsylvania  
Catherine Tucker, MIT (\*)

September 2007

This version: February 2008

## **Abstract**

Social interactions occur when agents in a network affect other agents' choices directly, as opposed to via the intermediation of markets. The study of such interactions and the resultant outcomes has long been an area of interest across a wide variety of social sciences. With the advent of electronic media that facilitate and record such interactions, this interest has grown sharply in the business world as well. In this paper, we provide a brief summary of what is known so far, discuss the main challenges for researchers interested in this area and provide a common vocabulary that will hopefully engender future (cross-disciplinary) research. The paper considers the challenges of distinguishing actual causal social interactions from other phenomena that may lead to a false inference of causality. Further, we distinguish between two broadly defined types of social interactions that relate to how strongly interactions spread through a network. We also provide a very selective review of how insights from other disciplines can improve and inform modeling choices. Finally, we discuss how models of social interaction can be used to provide guidelines for marketing policy and conclude with thoughts on future research directions.

(\*) This paper is based on a session titled "Interdependent Choices and Social Multipliers: Identification, Empirical Methods and Policy Implications" (with the same participants) that was part of the Seventh Triennial Invitational Choice Symposium hosted by the University of Pennsylvania's Wharton School in Philadelphia during June 13-17, 2007. The authors would like to thank the organizers, Eric Bradlow and Robert Meyer of the Wharton School, for giving them an opportunity to be a part of the symposium. All correspondence may be addressed to the co-chairs at [Hartmann\\_Wesley@gsb.stanford.edu](mailto:Hartmann_Wesley@gsb.stanford.edu), [pmanchan@bus.umich.edu](mailto:pmanchan@bus.umich.edu) and [harikesh.nair@stanford.edu](mailto:harikesh.nair@stanford.edu).

# 1 Introduction

The increasing recognition of the role of social interactions in various domains (e.g. social-networking on sites such as MySpace.com or Facebook.com, and social-search on Yahoo!) has spurred renewed interest in modeling and understanding the implications of interactions among agents. These interactions are of primary importance to firms and policy-makers because they allow a stimulus to one individual to be magnified by its dispersion through the network. In some cases, an intervention such as a marketing message to an individual (e.g., via an advertising exposure) can spillover to others in a network, allowing them to learn about a product despite not being directly exposed to the advertisement. In other cases, network members may not only share information or spread phenomena through the network, but may actually care about whether others make similar choices. In these latter cases, an agent's actions generate a feedback to the agent through the choices of other network members, thereby multiplying the effect of any initial stimulus to the agent. In both cases, social interactions imply that the aggregate level effect of marketing activity to agents becomes much larger than the just the sum of the individual-level effects.

Academic progress in modeling and understanding social interactions is likely to have implications for both policy makers and industry. Policy makers will benefit from the facilitation of improved design and measurement of social interventions. Industry will benefit from the development of rigorous metrics for making informed decisions. While spending on social network advertising is already at \$280 million (about 2% of all online advertising spending) and expected to reach about \$2 billion (6% of all online advertising) in 2010, currently, there is little understanding and consensus on the definition of social networking and the measurement of the effectiveness of advertising on these sites (AMA 2007). Our hope is that this paper will stimulate further progress on this front.

The existing literature in this area comes from a variety of academic disciplines including development economics, industrial organization, sociology, computation and marketing. Each discipline emphasizes varying research questions using differing methodologies. There exist immense economies-of-scope in bridging these disciplines and bringing these differing approaches to bear on solving common problems of interest. We offer a selective discussion of the state-of-the art across these disciplines. Our goal is to summarize the progress made, to highlight the current challenges, and to attempt to provide a common vocabulary for engendering cross-disciplinary research. Since the field is vast, we do not by any means, attempt or claim to provide a comprehensive summary. Rather, we consider examples from the literature to help us articulate a unifying framework that helps identify actual causal social interactions in data. We distinguish whether these interactions involve spillovers through the network or an even greater compounding of return-on-investments of marketing actions through a "social multiplier." We focus heavily on issues of measurement, considering (i) very specific requirements of primary and secondary data, (ii) modeling approaches that can help disentangle the contributors to the multiplier and hence the full effect of a marketing activity, and (iii) experiments which can uncover multipliers.

Empirical analyses of social interactions seek to understand the manifestation of some underlying model of individual behavior in observed data. We structure our summary of empirical analyses of social interactions around whether relationships have the necessary characteristics to generate social multipliers, as opposed to spillovers. A social multiplier arises in a dyadic relationship when both agents' actions (or outcomes) are affected by the *actions* of the other agent *and* the agents recognize this, and internalize these effects, when selecting their actions. We characterize such interactions as *active*, while a *passive* interaction involves at least one agent being unaffected by the actions of the

other. We consider this distinction more formally in the following section, but the critical point is that a passive interaction lacks the feedback which generates the social multiplier.

The relevance of the active vs. passive terminology can be illustrated by considering two historical approaches within the literature. In disease spreading models, an agent's probability of infection is a stochastic function of other agents' infections. This forms the underpinnings of many approaches to social interactions, such as those in marketing derived from the Bass model (1969). These models evolve sequentially: adoption choices in one period are specified as a function of adoption choices in the previous period. However, the lack of a forward component prevents adopters at one point in time from internalizing the impact of their action on future adoption decisions. As we discuss later, this feature makes these models passive because of the absence of a feedback loop.

Active social interactions involve feedback loops that create interdependencies between the choices of individuals. In such situations, modeling approaches typically have to deal with issues of simultaneity and considerations of equilibria. Early considerations of equilibria between socially interacting individuals arose in the sociology literature with the work of Schelling (1971) and Granovetter (1978). More recently, models and approaches that explicitly model agents' payoffs as a function of other agents' choices or states have appeared in economics based models. These models often require unique solutions to evaluate the effect of a stimulus that is subject to a social multiplier and involve a rigorous consideration of identification in data.

Both passive and active interactions are causal in nature and have implications for policy in terms of magnifying the effect of agent-level policy interventions. A challenge for empirical work that tries to uncover social interactions from data on behavior is that both passive and active interactions are confounded with other spurious sources that generate correlation in observed behavior in the data. A key goal of empirical work has been to separate true causal interactions from these sources of spurious correlation.

In terms of the organization of the paper, we first elaborate on passive versus active interactions both from a modeling point of view as well as from a policy outcome point of view. We use examples from the economics and marketing literatures to clarify and illustrate our discussion. The economics based approach to social interactions is valuable in that it allows for an empirically testable or estimable model to be specified, that captures behavioral relationships and specifies causal linkages among variables. However, there are many identification challenges in estimating such causal effects. We then discuss how additional data and experimental approaches (in the field) could help in overcoming these challenges.

There are many types of social interactions too complex to be entirely captured within a simultaneous equations econometric model or simple field experiments. The sociology literature considers many different types of actions taken by an individual that can result in a variety of social structures. We therefore illustrate how researchers could use the richness of this approach to inform and guide econometric models. We then provide examples of marketing policy questions that research on social interactions can address and note how the rise of online media can prove to be a rich source of data. In conclusion we suggest some overall guidelines that could accelerate research in this area.

Throughout, we argue that marketing researchers should seek to use models that distinguish causal effects between agents from other sources of correlated behavior. This will allow marketing researchers to measure policy-relevant effects that inform us about the full effects of marketing mix expenditure and allocation. However, the detail and complexity of the data may in many cases prevent modeling or identification of causal effects. In these cases, a goal of this paper is to guide researchers

in making recommendations that reflect the degree of uncertainty that such data leave with respect to the presence and magnitude of social interactions effects.

## 2 Modeling Social Interactions

Models of social interactions begin with a framework of how an individual is affected by others in his reference group. The reference group is simply the set of other agents whose behavior the focal agent can be affected by. This definition is broad, and can include all other agents in the economy (“macro”-level models), or small social groups (“micro”-level models).

Social interactions are of key interest to marketers and policy makers due to the presence of *social spillovers* and *social multipliers*. A *social spillover* arises when an intervention or marketing action to an agent affects the behavior of others in his reference group via a social interaction. In some cases, social interactions also incorporate a feedback loop such that the changed behavior of the group induced by a marketing action to the focal agent, feeds back to reinforce the focal agent’s action. This in turn changes the behavior of the group, until a steady-state is attained. In such situations, social interactions also engender a *social multiplier* that multiplies the effect of the intervention to the initial agent. From a normative point of view, spillovers and multipliers may vastly increase the return-on-investment to policy interventions, and hence are of significant interest to firms. From a positive point of view, spillovers and multipliers can help explain large observed variation in economic outcomes of interest, even in spite of small changes in underlying primitives.

To clarify the distinction between spillovers and multipliers, we first present a simple linear model of social interactions between two agents. We use the model to explain what underlying primitives of the interaction generate spillovers in some contexts, and multipliers in others. We then use these implications as a basis of developing a useful taxonomy to interpret extant models in the literature.

### 2.1 Spillovers and Multipliers

We model social interactions via a simultaneous equations framework. We consider a group composed of two agents  $i$  and  $j$  taking an action,  $a$ . The actions form the outcome variable of interest. Denote  $a_i$  and  $a_j$  as the actions of agents  $i$  and  $j$  respectively. We write,

$$a_i = \beta_i x_i + \gamma_i a_j + \mu_i z_j + u_i \quad (1.1)$$

$$a_j = \beta_j x_j + \gamma_j a_i + \mu_j z_i + u_j \quad (1.2)$$

where  $x_i$  are characteristics of agent  $i$  and  $z_j$  is a characteristics pertaining to agent  $j$  that affects  $i$ ’s outcome (analogously for  $j$ ). We think of  $z_j$  as exogenous in that it is not chosen by either  $i$  or  $j$ .  $\beta$  and  $\mu$  respectively measure the effect of these variables on outcomes, which are allowed to be agent-specific. The  $u$ ’s are unobservables or errors affecting outcomes. Here, the  $\gamma$ -s are parameters measuring the causal effect of one agent’s actions on another, while the  $\mu$ -s are parameters measuring the causal effect of one agent’s characteristics on one another. For now, we assume that both  $x$  and  $z$  are exogenous.

*Social multiplier*: The system of equations specified above engenders a *social multiplier* if  $\gamma_i \neq 0$  and  $\gamma_j \neq 0$ , and are of the same sign. In this case,  $i$ ’s action affects  $j$ ’s action and vice versa, such that  $i$ ’s action feeds back upon itself through  $\gamma_j$  then  $\gamma_i$ . The multiplier occurs because a small increase in  $a_j$  increases  $a_i$  through  $\gamma_i$ , which in turn increases  $a_j$  even more through  $\gamma_j$ , and so on, until an equilibrium is attained. The term social multiplier is used because a change in  $x_i$ , for example, will have a greater

total effect than  $\beta_i$  on  $a_i$  because of the feedback through  $\gamma_j$  and  $\gamma_i$ . The key for a multiplier to arise therefore is that *actions* of members have a direct, similar effect on each other.

*Social Spillover*: There are two cases in which social interactions exist, but create spillovers instead of multipliers. The first is the case of an asymmetry in which one agent's action does not affect the other, e.g. if  $\gamma_j = 0$ , and  $\gamma_i \neq 0$ . In this case, there is no multiplier because a small change in  $x_j$  will shift  $a_j$  via  $\beta_j$ , and shift  $a_i$  via  $\gamma_i$ . However, since  $\gamma_j = 0$ , there is no feedback loop, and the effect on  $a_j$  is limited to  $\beta_j$ . The second case of a spillover without a multiplier arises when  $\gamma_i = \gamma_j = 0$ , but  $\mu_i \neq 0$  or  $\mu_j \neq 0$ . This social interaction results in a *spillover* since an increase in  $z_j$  caused for instance, by changing marketing effort to  $j$ , also affects  $i$  via  $\mu_i$ . However, since the  $\gamma$ -s are 0, there is no feedback from  $j$  back to  $i$ , and hence, there is no multiplier.<sup>1</sup>

Finally, note that in the presence of either kind of social interactions, whether involving actions or characteristics, the outcome variables  $a_i$  and  $a_j$  will be correlated. The goal of econometric work is to use the observed correlation in actions between members, to identify the causal effects,  $\gamma$ -s and  $\mu$ -s. Note that correlation can also arise if the unobservables  $u_i$  and  $u_j$  are correlated. This spurious correlation does not result in any spillovers or multipliers, and is therefore not policy-relevant.

### An extension: Word-of-mouth

The model above captures social interactions arising through inter-related outcomes  $a_i$  and  $a_j$ . However, many social interactions, such as word-of-mouth (WOM) do not involve a direct relationship between the outcomes of group members. Suppose  $z_j$  represents word-of-mouth from individual  $j$  that affects the outcome of individual  $i$ . Furthermore, to focus on WOM effects, assume for the moment that  $\gamma_i = \gamma_j = 0$ . Without an extension of the model, and under the assumption that  $\mathbf{E}[z_j \mu_i] = 0$ , the  $\mu$ -s are identified and measure the effect of WOM. However, it is possible the WOM is endogenously chosen by agents, and hence viewed as an action, rather than a characteristic. We therefore introduce a more general specification of  $z$ :

$$z_i = \theta w_i + \lambda a_j + \delta a_i + e_i \quad (2.1)$$

$$z_j = \theta w_j + \lambda a_i + \delta a_j + e_j \quad (2.2)$$

where  $w$  may include variables in  $x$ , as well as other variables not contained in  $x$ .<sup>2</sup> Conceptually,  $i$ 's word of mouth to  $j$ ,  $z_i$ , depends on his characteristics  $w_i$ , unobservables  $e_i$ , as well as his actions  $a_i$ , and the actions of member  $j$ ,  $a_j$ . As an example, suppose the actions are the decisions to adopt a new technology. The above model says that  $i$ 's decision to send WOM to  $j$  depends on whether  $i$  himself adopts, as well as how much  $i$  expects  $j$  to adopt.<sup>3</sup> Based on the two sets of inter-related simultaneous

---

<sup>1</sup> Manski (2000) refers to the presence of  $\gamma$  in both interacting agents equations as an endogenous social effect and  $\mu$  as an exogenous social effect.

<sup>2</sup> We do not index the parameters  $\theta$ ,  $\lambda$  and  $\delta$  by the indices  $i$  and  $j$  for expositional simplicity; but conceptually, these parameters can be individual-specific.

<sup>3</sup> We could generalize further without any change in the substantive implications of our argument, that only the *expected* action of  $j$  drives  $i$ 's decision to send WOM: e.g.

$$z_i = \theta w_i + \lambda \mathbf{E}_i a_j + \delta a_i + e_i$$

$$z_j = \theta w_j + \lambda a_i + \delta \mathbf{E}_j a_j + e_j$$

equations models, a WOM-based social multiplier only occurs if both of the  $\delta$ -s are nonzero. In such a case,  $a_i$  affects  $z_i$  (through  $\delta$ ), which in turn affects  $a_j$  (through  $\mu_j$ ), which cycles back to  $a_i$  through  $z_j$ . An example of this might be pre-release anticipation for a movie or new technology product such as the Apple iPhone. If  $a$  measures individual's latent utilities for eventually buying the iPhone, a WOM-based multiplier may exaggerate these latent utilities because an individual is more likely to talk about the iPhone to others the more he likes it ( $\delta$ 's effect on  $z$ ). This WOM can lead to a reinforcement of his latent utility if his talking about the iPhone increases the other agents' latent utility and how much she in turn talks about it.

## Sequential decisions

Finally, note that multipliers inherently arose in the above contexts due to the simultaneous nature of decision-making. We now consider the question of whether multipliers of this sort disappear in interaction contexts where decisions are made sequentially. Let  $t$  denote time, and suppose only one agent can take an action in each period. This may be the case because decision-making in the group is known *a priori* to be sequential or because the analyst formulates the problem in continuous-time with infinitely small time increments.<sup>4</sup> Suppose  $i$  takes an action in period  $t$  conditioning on  $j$ 's action in period  $t-1$ , following which  $j$  takes an action in period  $t+1$ , conditioning on  $i$ 's action in period  $t$ . Suppose the model generating the actions  $a_{it}$  and  $a_{j,t+1}$  is,

$$a_{it} = \arg \max_a \left[ V(a, a_{j,t-1}, x_{it}, z_{jt}, u_{it}) \right]$$

$$a_{j,t+1} = \arg \max_a \left[ V(a, a_{it}, x_{j,t+1}, z_{i,t+1}, u_{j,t+1}) \right]$$

where  $V$  is the agent's present-discounted payoffs given action  $a$ . For the moment, consider one iteration of this process for each agent. A social multiplier arises in this context, if a small change in  $x_{it}$  increases  $a_{it}$  through a feedback effect via the action  $a_{j,t+1}$ . This arises if  $i$  is forward-looking and anticipates that his action  $a_{it}$  would change  $a_{j,t+1}$ . To clarify, let  $v_i(\cdot)$  and  $v_j(\cdot)$  denote  $i$  and  $j$ 's immediate payoff from the action. Then, a social multiplier is generated if one can write,

$$V(a_{it}, a_{j,t-1}, x_{it}, z_{jt}, u_{it}) = v_i(a_{it}, a_{j,t-1}, x_{it}, z_{jt}, u_{it}) + E \left[ V(a_{i,t+2}, a_{j,t+1}, x_{i,t+2}, z_{j,t+2}, u_{i,t+2}) \right] \quad (3.1)$$

$$V(a_{j,t+1}, a_{it}, x_{j,t+1}, z_{i,t+1}, u_{j,t+1}) = v_j(a_{j,t+1}, a_{it}, x_{j,t+1}, z_{i,t+1}, u_{j,t+1}) + E \left[ V(a_{j,t+3}, a_{i,t+2}, x_{j,t+3}, z_{i,t+3}, u_{j,t+3}) \right] \quad (3.2)$$

In the above equation, a small change in  $x_{it}$  changes  $v_i(\cdot)$  and thus affects  $a_{it}$ . This in turn, shifts  $a_{j,t+1}$  via  $v_j(\cdot)$ . However, since the term in expectations in (3.1) above is a function of  $a_{j,t+1}$ , there is an additional feedback effect on  $a_{it}$ . Hence, a social multiplier can still arise in this system. Intuitively, in sequential decision-making, a social multiplier may arise if each agent making a move incorporates that his action today will change the behavior of the agents in his group tomorrow. This is naturally satisfied in most models of dynamic forward-looking interactions.

## Discussion and a Taxonomy

To formalize the source of multipliers in social interactions, we characterize an interaction as *passive* if the feedback loop does not exist. A *passive* social interaction therefore arises when one agent in a dyadic relationship either is not affected by the other or does not recognize the effect of their outcome

<sup>4</sup> As Doraszelski and Judd (2007) point out, a continuous time model does not have simultaneity because only one agent can act at an instantaneous point in time.



on the other when choosing it. In contrast, an *active* social interaction arises when both agents in a dyadic relationship affect one another similarly (i.e. the interaction effects have the same sign) and recognize the effect of their outcome on the other when choosing it. An active social interaction may therefore manifest itself in a simultaneous decision context as in Equation (1) above or in a dynamic decision context in which a future outcome of another agent affects the payoffs of the agent's current action/outcome as in Equation (3) above.

The necessity of modeling agents' actions requires the corresponding models to take a stand on whether to specify formally agent's preferences, information sets and the equilibria being played. *Structural* models of active interactions spell out the incentives and payoffs associated with agent's actions, and explicitly model some combination of agents' preferences, information sets and equilibria. *Reduced form* models of active interactions are consistent with structural models, but are typically agnostic about the explicit definition of these fundamentals, focusing more on consistent estimation of social effects with minimal assumptions.

The distinction made above between active and passive social interactions has important implications for policy. Passive interactions imply that a firm or policy maker's stimulus to a member of a social group will spill over to others within that group. Active interactions have this feature, but also produce a social multiplier. Formally, agents' actions in active social interactions are strategic complements. As is well known, in the presence of strategic complements, small changes in underlying preferences can result in large changes in aggregate outcomes. Finally, the presence of a feedback loop has important implications for the estimation of social interactions because it often implies a simultaneity issue and requires some consideration of equilibrium phenomena. This forms one aspect of the challenge of identifying social interactions from behavioral data. We consider these identification issues in more detail below.

## 2.2 The Identification problem

In confronting data, all non-experimental analyses of social interactions, whether passive or active, using either structural or reduced form models, are subject to an *identification* problem. The identification problem arises due to the challenge of separating correlations in observed behavior from true causal effects of one agent on another. In the remainder of this section, we first describe sources of this identification problem and another that may exaggerate estimated effects of social interactions. Second, we discuss extant models in the literature.

Correlation in observed choices among agents within a reference group is a necessary empirical feature of data in the presence of social interactions. An identification challenge arises because factors other than causal effects of agents' actions on each other can create spurious correlations in behavior. Only causal effects can result in a social interaction effect. Hence, uncovering causal effects accurately is key to formulating marketing and social policy. The primary confounding factors are endogenous group formation, correlated unobservables, and simultaneity. These issues were first formally laid out by Manski (1993) and Moffitt (2001) in the context of a linear model of social interactions. We provide a discussion of these issues, and discuss potential solutions to the identification problem outlined in the literature since then.

### 2.2.1 Endogenous group formation

The endogenous group formation problem arises because agents with similar tastes may tend to form social groups; hence, subsequent correlation in their behavior may reflect these common tastes, and not a causal effect of one's behavior on another. For example, in the application of Nair et al. (2006) to measuring social interactions in physician prescription behavior, many general practitioner physicians

tend to meet specialist physicians at conferences, some hosted by drug companies, which are organized around specific disease-conditions and therapeutic treatment options. Correlation in prescriptions between general practitioners and specialists can thus arise if preferences for treatment-options that make doctors tend to attend these events also relate to their tastes for drug prescription. Alternatively, in the application of Hartmann and Yildiz (2006) to modeling social interactions in golf demand, correlations in observed golfing activity between two partners may simply be driven by common tastes for golf that induced them to form a group in the first place. A researcher cannot therefore conclude directly from observed correlation in behavior that there exists a causal effect of an agent's prescription or golfing behavior on the prescription or golfing decisions respectively, of others in his reference group.

One solution to the endogeneity of group formation is facilitated by the availability of panel data. With panel data one can control for endogenous group formation via agent fixed effects (e.g. Nair et al. 2006), or by including a rich specification for heterogeneity (e.g. Hartmann and Yildiz 2006). Both fixed and random effects here serve the role of picking up common aspects of group tastes. In other contexts, researchers can directly model the process of group formation (Bala and Goyal 2000, Glaeser and Scheinkman 2001, Conley and Udry 2003). Nonparametric identification in the latter contexts is similar to selection models (e.g. Lee 1982), and requires an exclusion restriction, i.e. a variable that affects the propensity to join a group, but not subsequent behavior. Given the data requirements, studies that attempt this approach are rare.

### **2.2.2 Correlated unobservables**

A second source of correlation is *correlated unobservables* (to the econometrician) that drive the actions of all agents in a reference group similarly. The inclusion of fixed or random effects mitigates the correlated unobservables problem to some extent, since these control for time-invariant aspects of unobservables driving agents' behavior. In some contexts, one can use a difference-in-difference strategy of using the behavior of other agents *not* in the focal agent's reference group to control for common unobservables (e.g. Nair et al. 2006.) With large enough groups, one can also feasibly think of differencing out common unobservables by modeling *differences* in behavior within a group, as opposed to levels themselves.

### **2.2.3 Simultaneity**

Finally, a *simultaneity* problem arises due to the potentially simultaneous nature of decisions by the focal agent and others in his reference group. Due to simultaneity, correlation in subsequent actions could simply reflect the fact that the agents' decision affects the group's behavior, and at the same time, the group's behavior affects the agent's behavior. This has been referred to as the "reflection problem" in the literature (c.f. Manski 1993). Exclusion restrictions are the most accessible identification approach to solving the simultaneity problem. The researcher needs an instrumental variable that affects a focal agent's decision, but can be a priori excluded from the decision of others in his reference group. Nair et al. (2006) discuss how this approach may be used to achieve identification. Nam et al. (2006) and Tucker (2006) use this identification strategy in the related context of measuring local network effects among agents in the adoption of new technology. Alternatively, researchers can directly model the equilibrium being played to solve the simultaneity problem. This is the strategy adopted by Hartmann and Yildiz (2006). An alternative exclusion restriction imposes a temporal ordering i.e., the focal individual's behavior in time ( $t+1$ ) is affected by the group behavior up to time ( $t$ ) (see Manchanda et al. 2004 for an example). However, a caveat to this identification strategy is that unobservables should not be correlated over time, and agents must be assumed not to be forward looking (i.e., agents are passive).

Each of the three factors discussed above could generate spurious correlation in observed actions, which when not controlled for, may lead to the misattribution of correlated behavior to causal social

interaction effects or biases in the estimates of the causal effects. We wish to emphasize for the reader that these issues are pervasive in all empirical settings that try to identify social interaction effects using behavioral data, irrespective of model specification or functional form assumptions.

We now discuss the extant models in the literature. We first discuss models with passive social interactions, and then present a discussion of active social interactions. For each, we also discuss how the identification problems discussed above have been addressed.

## **2.3 Models of Passive Social Interactions**

### **2.3.1 Epidemiology/Disease Spreading Models**

Many models of social interactions and contagion have their roots in the epidemiology literature, which has as its goal the forecasting of the rate of growth of diseases in the population. These models typically consider agents as being in one of three states: susceptible (S), infected (I) or recovered (R). The primitives of the model are the disease-specific probabilities by which agents randomly move from one state to another. Of key interest is the reproduction rate,  $R_0 = \beta/\gamma$ , where  $\beta$  is the probability of a movement from state S to state I, and  $\gamma$  is the probability of movement from state I to state R. A disease with  $R_0 > 1$  is termed an epidemic. These models treat agents as passive because agents' states stochastically change based on the state of their reference group (i.e. actions, preferences and tradeoffs are not explicitly modeled) rather than having the potential to feedback on themselves. The basic disease-spreading model does not control for agent heterogeneity and is susceptible to the endogenous group selection problem (the agent may choose to not be in a group or area that has high prevalence of the disease). Correlated unobservables may also be an issue: for example, two agents in a community may have caught the disease because they were both exposed to the germ sequentially and developed symptoms sequentially. However, their visits to the location of the germ is unobserved to the econometrician, and hence, he may conclude from the observed data that the first agent infected the second.

One can generate differing models of contagion from the basic epidemiological framework by specifying various specifications of the probability of state transitions. The framework has been successfully applied in marketing via the Bass (1969) model to studying aggregate patterns of diffusion of new products. A useful distinction for such network models is between a pure mixing model (e.g., Bass 1969 discussed below) and a network spatial model. Network spatial models explicitly model network topology and illustrate how the specific network structure can sometimes generate novel and different insights on the diffusion process (e.g., Durett, 1999). We refer the interested reader to Van den Bulte and Wuyts (2007) for more detail on network models.

### **2.3.2 Bass Model**

The Bass model is a particular instance of an epidemiological model in which the probability by which an individual catches a "social contagion" is linearly related to the number of others in the agents' reference group (in this case, the entire economy), that are in that state (i.e.  $\beta = p + q \cdot n(I)$ ). Diffusion is analogous to the process by which a disease spreads, and an epidemic is a product that has "taken-off". The particular specification of the probability of social contagion implies a positive concave relationship between sales and the installed base, which is also commonly observed in aggregate diffusion data. Hence, the Bass model can provide a good fit to the temporal distribution of many aggregate sales data. The reader should also note that this pattern is potentially consistent with other explanations, including serial correlation in sales-related unobservables over time. In the absence of micro-data, the fact that that current sales is statistically significantly related to the installed base is consistent with, but not evidence of, social interactions at the agent level. Further, as noted by Glaeser et al. (2003), in the presence of social interactions at the individual-level, models estimated on

aggregate data are liable to be subject to large aggregation biases that exaggerate the extent of individual-level social interactions. Hence, “ $q$  coefficient-s” in a Bass model are less suited for an interpretation as causal measures of social contagion, and are more appropriately interpreted as descriptive parameters capturing the dependence of current aggregate sales on the past installed base of the product. In terms of prediction, the Bass model is a simple tool, and has difficulty forecasting temporal heterogeneity and high variability in adoption rates. Watts et al. (2005) develop a model that captures these observed features of adoption data.

### 2.3.3 *Spatial Models*

An alternative class of models seeks to flexibly describe the correlations that exist in the observed choices of agents within a reference group. These papers trace their origins to the spatial econometrics literature (Anselin 1998, 2001), which seeks to flexibly describe correlations in observed data among contiguous spatial units or entities. Models for social interactions in this class specify correlation structures such that responses by individuals near one another in location or attribute space generate similar outcomes. These models are reduced-form representations of a social interaction model that may be either active or passive. Yang and Allenby (2003) consider the decision of consumers to buy foreign or domestic automobiles. The idea is that consumers in certain regions may wish to coordinate on one of these choices. They model the interrelatedness of these decisions through correlations in underlying preferences. Such correlations may be the result of individuals’ desires to buy similar autos, but could also be the result of common unobserved beliefs, such as patriotism. The model thus fits the data of a social interaction, but does not actually distinguish whether an interaction is present.

## 3 **Models of Active Social Interactions**

We now discuss active models of social interaction. These models originated in sociology, and are grounded in sociological and economic theory. We first discuss the linear-in-means model, which is a canonical example from the literature. We then discuss non-linear discrete choice models of social interactions. The latter raises an additional issue related to multiplicity of possible equilibria. We discuss potential approaches to solving this problem. We then discuss the sociology literature that leads the research uncovering the rich nature in which payoffs and actions of agents build observed social structures. We conclude the section with a discussion of the challenges faced by the empirical literature from incorporating such richly defined social structures.

We begin with continuous actions because they can often be represented using a simple linear model. We then consider discrete choice models when the actions or states of others in the reference group enter the choice probabilities of agents.

### 3.1 **Linear Models of Social Interactions**

The linear-in-means model is a specification in which the actions of an agent are linearly related to the characteristics, as well as the mean behavior of others in his reference group. The linear-in-means model can be interpreted as a structural model in which agent’s preferences are such that actions are a linear function of other’s actions, or as a linear approximation to the reduced form implied by a non-linear model. An example of this model in marketing is Nair et al. (2006.) We believe that a discussion of the linear model is instructive because, a) it has been a popular model in empirical work, and b) it serves to illustrate the notion of equilibrium, as well as the associated identification challenges clearly. The linear-in-means framework captures active interactions because it considers the effect of an agent’s action on the actions of others in his reference group.

Simultaneity arises in this model because social interactions are symmetric in the sense that an agents action’s affect the actions of other agents in the group. Manski (1993) defines this as the “*reflection*”

problem - group behavior affects  $i$ 's behavior, which, in turn, affects the group's behavior. Referring back to the equations in section 2.1, the solution to the system of equations is the equilibrium for the group. It can be shown that  $\gamma$  and  $\mu$  — i.e. the causal social interaction effect and the “exogenous social effects” — are separately identified only if the errors of the structural form of the equation are uncorrelated. Moffit (2001) lays out the argument and illustrates that the endogenous social effect,  $\gamma$ , can be recovered from the variance/covariance terms of the reduced form of the model. There are two reasons why the errors of the structural equation may be correlated, hence preventing identification of  $\gamma$ . First, if individuals formed groups based on common tastes (i.e. endogenous group formation), there may be a fixed effect common to both individuals in the group. Second, the individuals may be exposed to common or correlated unobservables. In either case, the convenience of identifying  $\gamma$  in the linear model from the reduced form variance covariance matrix breaks down.

### 3.1.1 *The Importance of Social Network Information*

A useful aspect of the analysis above is that it emphasizes the importance of obtaining exogenously defined social network information. In other words, one cannot use the correlation in observed behavior to bin agents into groups, and then subsequently use the correlation in the actions of the agent and those in these groups to study the effect of social interactions. A solution is to obtain social network information via direct elicitation from agents or from surveys (see for example, Nair et al. 2007.) A related issue is that if information clearly defining the network is not available, correlated unobservables can be even more problematic. For instance, without clear network information, a researcher may define the network based on geographic location (e.g. all persons in a zip code). In this case, it is difficult to separate the causal effect of members within the zip-code on each other from common zip-code level unobservables that affect all members similarly, and tend their behavior to be correlated.

## 3.2 Discrete Choice Models of Social Interactions

While the linear model is appropriate for a wide range of situations, some choice contexts are non-linear by definition. A typical example is decision contexts involving discrete choice. Discrete choice models of social interactions extend the typical latent random utility discrete choice model to include the action (or state) of one or more other agents. Specifying the payoff of an agent (A) to be the function of another agent's (B's) action is complicated by the fact that B's action likely also depends on A's. A typical approach in such cases is to use a game theoretic framework to determine which combinations of actions are possible equilibria. Once equilibria are defined, researchers can begin to define probabilities over observed choices. The empirical game-theoretic literature can be broadly classified as models with private information and models with complete information. The completeness aspect here refers to the nature of random shocks to agents' payoff functions. In private information models, the random payoff shocks are only known by the agent directly affected by them, while in complete information models, they are common knowledge to all agents. In the remainder of this section, we briefly review these two approaches to modeling social interactions.

### 3.2.1 *Models with Incomplete Information*

Incomplete information discrete choice models of social interactions trace their origins to the sociology literature, in particular, to the pioneering work of Shelling (1971) and Grannovetter (1978). Shelling and Grannovetter described *threshold* models of social interactions, in which the marginal utility that some agents obtain from an action is an increasing function of the proportion of the population taking the similar action. Both Shelling's and Grannovetter's models result in a *critical mass* effect, such that once the proportion of the population choosing the action crosses a threshold point, only extreme outcomes are stable equilibria. Hence, around the threshold point, small changes in underlying preferences or small shocks to payoffs can result in large changes in aggregate outcomes. Formally, this is the multiple equilibria phenomena, which are pervasive in models with

strategic complementarities. The threshold effects here are analogous to critical mass effects found in models of network effects (for example, Economides and Himmelberg 1995), and has been referred to as “tipping” by the popular press (Gladwell 2000). The basic threshold models formed the kernel for *cascade* models in the mathematics, physics and computer science literatures that modeled networks as collections of connected agents differentiated by their *vulnerability* (Dodds and Watts 2004, 2005). The vulnerability of an agent is defined as the threshold number of connected agents that should take an action, before the agent himself would. Using both analytical tools and numerical techniques, Watts (2002) demonstrated that random networks with sufficiently connected clusters of vulnerables are susceptible to large-scale cascades. These results extend the notion of critical mass on a network in that activating even a single individual could lead to a cascade throughout the vulnerable cluster (the critical mass) which in turn would lead to a cascade throughout the whole network.

In parallel work, the critical mass model was extended by Brock and Durlauf (2001) who cast the model in terms of discrete choice. In their model, agent’s payoff functions for each choice are shifted by the expected actions of others in their reference group, as well as agent-specific errors similar to standard random-utility models of discrete choice. The random payoff errors are given a private information interpretation, in the sense that they are observable to the focal agent, but not to others in the reference group. An equilibrium in the model is a self-consistent set of probabilities for all agents such that the expected actions of all agents within a group coincide with the actual probabilities of choice. The equilibrium satisfies a fixed point condition which facilitates the solution of the equilibrium probabilities. As expected, the fixed-point is not unique, i.e. there are multiple equilibria. The presence of multiple equilibria creates challenges for empirical work – we return to this point below. Bajari, Hong, Kranier and Nekipelov (2006) discuss identification in models analogous to the Brock and Durlauf-framework. The requirements for nonparametric identification are familiar: analogous to standard discrete choice models, the payoff from one of the actions needs to be normalized. Further, the researcher needs an exclusion restriction such that one of the variables shifting the payoff of a focal agent can be excluded from the payoff functions of other agents in his reference group.

### **3.2.2 Models with Complete Information**

In some settings with small reference groups, complete information may likely be a more reasonable assumption. In such contexts, the model of agents’ actions is a discrete game. Following Bresnahan and Reiss (1991), a discrete game is a generalization of a standard discrete choice model where utility depends on the actions of other players. Empirical models of discrete games have traditionally been defined for and applied to firms competing with one another (see the paper by Draganska et al. (2008) in this issue for a more detailed description and analysis of this literature), but the basic model can also be adapted to individuals interacting with one another. Hartmann and Yildiz (2006) is an example that illustrates how the Bresnahan and Reiss (1991) depictions of equilibria are easily adapted to individuals coordinating decisions. While their model is applied to a problem of small scale (two players and two possible actions), the paper describes many of the social interaction identification issues within the context of structural estimation. For instance, the heterogeneity specification can be used to account for endogenous group formation and the equilibrium model can account for simultaneity biases.

Empirical researchers using either class of models above must confront the problem of multiple equilibria. Note that multiplicity is a feature of the non-linearity of the model (i.e. the linear model has a unique equilibrium). Below, we provide a brief discussion of the multiple equilibria problem, and discuss potential solutions that have been proposed in the recent literature.

### 3.2.3 *Multiple equilibria*

The actions of agents in many of the above social interactions models are strategic complements, and hence, these models are particularly prone to multiple equilibria. Multiple equilibria create complications for empirical work since there may be multiple outcomes that are equilibria for a given set of parameters, observable variables and unobservable variables of the agents. This implies that researchers must know which equilibria would be played before being able to define a likelihood for the data. Unfortunately, theory offers little guidance on which equilibria would actually be played, and hence it is a priori difficult to rule out alternative equilibria. One approach proposed by Bajari, Hong and Ryan (2004) is to estimate an equilibrium selection equation. This would be the reasonable approach in most applications, but it requires an exclusion restriction: the researcher must have available a variable which affects the equilibrium chosen but does not affect the payoffs of the agents involved. In many cases, it may be impossible to find such a variable. In cases where such a variable is not available, researchers have typically defined an equilibrium selection rule which sorts between the equilibria. A common example is to assume the equilibria which maximizes total surplus is always chosen. In the case of social interactions models, Hartmann and Yildiz (2006) note that the equilibria that maximizes total surplus also often is Pareto dominant such that it makes all agents better off. This suggests that such equilibrium selection rules may be more reasonable in the context of social interactions models. Similar approaches have also been suggested in the context of empirical games with network effects (Ackerberg and Gowrisankaran 2006). A final approach is to condition on the equilibrium that was actually played in the data, and to obtain initial estimates of the rules describing agent's actions from the data, given the chosen equilibrium. This is the approach followed in the recent literature on two-step estimation of dynamic games (see Bajari, Benkard and Levin 2007, and the references cited therein.) While this solves the estimation problem in many situations, the researcher is still confronted with the issue of picking an equilibrium when using the estimates to simulate counterfactuals. Addressing these concerns remains an open area for future research.

### 3.2.4 *Models with Forward-Looking Consumers*

While the preceding describes models of static games, or repeated static games, many social interactions involve agents interacting repeatedly in environments where past and future actions and states are also relevant. Dynamic empirical models of social interactions involve all of the estimation issues discussed above, plus the complexities of estimating dynamic interactions. To appreciate the complexity, typical social interaction models with incomplete information involve a nested fixed point algorithm that finds the Bayesian Nash equilibrium of the expected actions of agents. The computational challenge in a typical dynamic choice model involves solving a nested fixed point problem to obtain the value functions measuring the discounted present value associated with states and actions. A dynamic empirical model of social interactions needs to solve both computational challenges. These models broadly fall into the empirical dynamic games literature. Fortunately, the recent two-step estimation approach defined by Bajari, Benkard and Levin (2007) allows for estimation of some types of dynamic games.

There have been a few recent applications of these models in the context of network effects. Ryan and Tucker (2006) consider the adoption of a video-conferencing technology by employees of a given firm. The dynamics arise from the fact that adopting the technology involves incurring an adoption cost, but the adoption allows an employee to use the technology to communicate in all future periods. Implicitly the adoption decision is driven by the utility the individual obtains from communicating with other adopters. An appealing aspect of their application is that they recognize that the calling behavior between groups of individuals informs the model about the desirability of adoption. They consider active agents in terms of the adoption decisions (though not the calling decisions) and

heterogeneous types of agents that can pick up behaviors similar to those defined in the threshold models from the mathematics literature.

Another paper considering dynamic network effects is Dubé, Hitsch and Chintagunta (2007). They consider the indirect network effects arising in markets with platforms and compatible software to analyze how agent's forward looking behavior affects the presence/rate of tipping in the market. An important contribution of this paper is that it estimates the dynamic demand side using the two-step methods mentioned above, but explicitly solves the dynamic equilibrium model on the supply side to evaluate counterfactuals.

These ambitious papers take the first steps into the field of dynamic empirical models of social interactions. In terms of the identification issues discussed throughout this paper, the equilibrium models should help with simultaneity concerns, however both papers do have interaction elements that are not modeled in an equilibrium framework.<sup>5</sup> One limitation of the two-step method is that unobserved heterogeneity is not accommodated. This could create concerns over endogenous group formation, but both papers circumvent this by estimating the models for a single pre-defined group (the firm in Ryan and Tucker 2006, and the market of all potential platform holders in Dube, Hitsch and Chintagunta, 2007). Estimation within groups all also helps with the common unobservables problem, so long as the common unobservables are not systematically varying over time.

### 3.3 Field Experiments

As has been noted till now, the identification of peer effects in social networks is a non-trivial task and is usually obtained by either having very high-quality data or imposing structure and using specific functional forms. One solution may be to exploit a natural experiment where there is random assignment of agents into groups (e.g. Sacerdote 2001 who studies peer effects on scholastic achievement of freshmen who were randomly assigned to roommates at Dartmouth college). The confinement to natural experiments is nevertheless, limiting. One possible way to avoid the necessity of either of the above is to carry out experiments directly in the field. The main idea is to exogenously vary the treatment (e.g., a marketing message) across subjects who have been randomly assigned to some treatment groups, as well as to a control group. Fairly straightforward analyses can then be used to verify the existence of the peer effect as well as to estimate the effect size. For example, Katz, Kling and Liebman (2001) use data from the Moving to Opportunity experiment to test for neighborhood effects on juvenile crime. Duflo and Saez (2003) is another experimental example involving word-of-mouth effects between employees deciding on retirement plans. In general however, there is a lack of studies that employ this approach – a situation that we hope changes in the future.

It is worth noting, however, that while field experiments are conceptually appealing, they can be problematic to implement in practice. Some common problems in these studies is obtaining buy-in from firms to carry out such studies (as almost always one of the experimental groups is worse-off), the actual cost of the study, low response rates, data recording problems and lack of foresight to include enough treatments in order to provide an explanation for the detected effect or effect size. Some researchers have turned successfully to the web as experiments are scalable, it is easy to track behavior and outcomes can also be measured cleanly (e.g. Salganik, Watts and Dodds 2006). However, some online experiments suffer from high attrition rates and operational issues. Generally, online experiments are difficult productions that are best designed to be game-like or fun and provide feedback to be appealing to participants.

---

<sup>5</sup> Ryan and Tucker (2006) model the calling behavior in a reduced-form manner while specifying the equilibrium over the employee adoption decisions and Dubé, Hitsch and Chintagunta (2007) model the software industry in a reduced-form manner while specifying the equilibrium over the decisions of consumers and platform suppliers.



Some other researchers have focused more on field studies, rather than field experiments. In such studies, researchers do not explicitly create treatment and control groups but try and measure (exogenously varying) stimuli and correlate them with outcomes of interest. For example, Godes and Mayzlin (2007), enroll loyal and non-loyal customers of a retail chain and encourage them to generate (positive) word-of-mouth for the chain to “close” peers (family and friends) and “far” peers (acquaintances). They then correlate the sales of the retail chain at the location of the enrollees as a function of the generated word-of-mouth. They find a main effect of word-of-mouth as well as that word-of-mouth seems more persuasive on far peers rather than close peers. In addition, word-of-mouth generated by non-loyal customers is more effective than that generated by loyal customers. Note that while field studies may involve some intervention by the researchers, without an experimental design in place, the identification challenges are usually similar to those using revealed data.

### **3.4 Sociological inquiry into social interactions**

There is of course a very strong tradition of work on social interaction in the sociology literature. The classic work that initiated a vast body of research over the last 40 years is the Medical Innovation study (Coleman et al. 1966). This was followed by other groundbreaking studies such as Schelling (1971), Granovetter (1973, 1978) and Burt (1987) to name a few. The focus of this literature has changed over time from whether people’s behavior was affected by social interaction to who was affected, followed by why and how. There is now a rich and varied set of insights that have emerged from this literature based on the questions of who is affected, why s/he is affected and how this effect propagates. Generally speaking, these insights have yet to be incorporated into econometric models of the type discussed above and are likely to prove a rich source of ideas that can be taken to data and models.

Perhaps the richness of this literature can be best exemplified by an example – the role of “status” and its effect on the extent and nature of social interactions. Status can be defined as an intangible asset that is “possessed” by an individual or organization that is highly regarded by others that are highly regarded.<sup>6</sup> The career-related consequences of status, especially for scientists, were brought into relief by Merton who used narrative data (interviews, writings etc.) to describe how, under a variety of situations, individuals of high status obtained disproportionate rewards relative to individuals of low status for equivalent levels of output or effort (Merton 1968). For instance, if two academics coauthor a paper, the higher status contributor reaps the lion share of the credit. The same is thought to occur when two scientists independently make the same discovery. Merton named this the “Matthew Effect” – an effect that has been documented widely since (for a discussion of influential models of status effects in markets and the boundary conditions of the Matthew effect, see Podolny (2005)). In addition to rewards, the Matthew effect also influences outcomes such as the propagation and establishment of ideas in social networks. An indirect illustration of this effect may be seen in Nair et al. (2007) where opinion leader physicians influence the behavior of physicians but not the other way around.

More recent research in sociology focuses on providing a richer description of status along with metrics that allow us to measure status. For example, Bothner et al (2007), allow for status to be multidimensional. The hypothesized dimensions of status in this work are “primary status” (as above) and “complementary status” (where status is based on being in demand by highly regarded others for an auxiliary role e.g., as a team member rather than a leader). They develop metrics for measuring

---

<sup>6</sup> We use the term possessed carefully because, unlike financial or physical capital, status is a form of social capital and therefore owned only in part by the individual or organization, whose audience--capable of withdrawing at any time the esteem and recognition on which status rests--also acts as a part owner (cf. Burt 1992)

both kinds of status and use them to investigate the performance of venture capital firms. They find that, net of primary status, complementary status has a positive effect on firm performance. Interestingly, they find that the two types of status interact negatively, suggesting that each kind of status corresponds to a distinct role in the market.

From a marketing standpoint, these additional dimensions describing social networks can serve as a useful guideline in terms of allocating marketing resources to different parts of the network based on expected flows and outcomes.

#### **4 Applicability to Marketing and Policy Effects**

There are many potential applications of social interactions to marketing. We have already discussed many examples. Other recent applications include television preference relationships between spouses (Yang et al., 2006), spatial diffusion of the use of an Internet grocer (Bell and Song, 2007), physician adoption of a new drug (Manchanda et al. 2004) and the effects of service quality and word of mouth on acquisition, usage and retention for a video on-demand service (Nam et al., 2006). The last application shows how popular metrics such as Customer Lifetime Value (CLV) can be understated if the social interaction effects (spillovers) are not accounted for. Virtually all these applications have found some evidence for the existence of social interaction effects, while the elaboration of the implications for marketing policy has varied across these studies. Another common factor underlying most of these studies is that measurement of the social network and the flow of interaction within the network has rarely been based on objective data (e.g. Bell and Song 2007 and Nam et al 2006 use geographic proximity to define networks). In terms of the future research, the rise of online media – online social networks, retail sites that explicitly allow and encourage interaction such as reviews and recommendations etc. - is an exciting development as the network, links and the flow of (electronic) information can be captured objectively (e.g., Trusov 2006). This has many advantages. For example, clearly defined network links can be helpful in that some common effects (to rule out correlated unobservables) may be estimated by comparing in-network agents with some individuals that may not be in any network.

The use of such data also helps answer questions that have not received enough attention in the past. For example, in recent research, Narayan (2006) models network formation by fitting a model of the time until a link between individuals within a network will occur. Note that even in this case, the reference group is defined ex-ante despite the fact that the timing of connections within the reference group are modeled endogenously. In another interesting example, Fleder and Hosanagar (2007) examine the effect of recommender systems. Recommendation systems create social interactions indirectly by pooling the purchase behavior of similar agents to make recommendations to agents currently making purchase decisions online. The main question that is addressed in this research is whether such systems lead to a long-tail effect in sales by increasing diversity (via exposing consumers to idiosyncratic products that they may not have been aware of otherwise) or a reduction in diversity via a tipping effect. They find that, recommendation systems lead to a reduction in average sales diversity online. Another related application is customer referrals. Essentially, these involve word-of-mouth that can be measured or captured by the firm to properly incentivize customers to spread the word. In fact, as online social networks develop and the data are more readily accessible to researchers and firms, elaborate compensation schemes may be able to reward influential network members on sites such as Facebook.com for the content and traffic they generate.

Going forward, we believe that researchers should continue to focus on the development of models that distinguish causal effects arising from social interactions between agents from other sources of correlated behavior. This is crucial as policy implications for marketers can only be developed

conditional on finding such causal effects. Much work needs to be done in translating these effects to actual policy recommendations. From a dissemination point of view, another area of research that is likely to be of use to marketers is the development of easy to compute metrics that can be applied to market data quickly and easily. Finally, marketing scholars should be encouraged to make use of the rich insights on social interactions and networks developed in other disciplines (e.g., sociology and information systems) to develop methodology and provide policy implications.

## 5 Conclusion

We hope that the discussion so far has made it clear that there are many interesting questions related to social interactions in marketing. These range from tests of theories describing whether, why and how consumers and firms interact to network formation to applied questions about marketing resource allocation in the presence of social interactions. In conclusion, we would like to highlight two key points.

First, the distinction between passive and active social interactions has very significant implications for the identification and magnitude of the effects of marketing policies. While the spillover effect of a passive interaction may help marketing promotions, information, or more generally, policies, to spread through a network of individuals, an active type of social interaction magnifies effects even more. The reinforcing of agents' behaviors under active interactions lead to large multipliers that result in significant aggregate effects of even small stimuli to a single agent. While such interaction effects may be particularly appealing from a policy-makers' point-of-view, active interactions involve simultaneity issues and potential multiple equilibria that affects both measurement and prediction.

Second, whether interactions are expected to be passive or active, empirical applications should explicitly consider how identification problems such as correlated unobservables and endogenous group formation are addressed in the context of their data. While we have been rigorous in drawing upon the literature in listing all the identification challenges, our intention is not to discourage researchers. Rather, our intention is to encourage research to continue to work in this area, while recognizing the boundaries of their analysis e.g. by pointing out potential confounds. Sources of exogeneity, well-defined groups and panel structure determine the scope of questions that can be addressed. With limited natural sources of exogenous variation in the data and modeling which may be too complex to adapt to the extensiveness of the data, it may be best in many applications to avoid questions relying on cleanly identified causal effects. In other words, statistical modeling which provides valuable data descriptors and predictors may be most useful. If causal effects in these social interaction environments are sought, the statistical patterns may help uncover networks and potential interactions that can be tested more explicitly using experimentation. In summary, a wish list of social interactions data that we encourage practitioners to collect includes data that (i) relays the set of relationships between individuals so networks structures can be analyzed, (ii) measures actionable marketing variables such as purchases or page views for network members so profitability can be analyzed, and (iii) uses experimentation to allow researchers to at least measure a "treatment" effect of a stimulus such as advertising to the social network.

## 6 References

- Ackerberg, Dan, and Gawtam Gowrisankaran. (2006) "Quantifying Equilibrium Network Externalities in the ACH Banking Industry," *RAND Journal of Economics*, forthcoming.
- American Marketing Association (2007), "Social Network Ad Spending Up," Report available at <http://www.marketingpower.com/content142914.php>

- Anselin, Luc. (1988). "Spatial Econometrics: Methods and Models," Dordrecht: Kluwer Academic Publishers.
- Anselin, Luc. (2001). "Spatial Econometrics," in B. Baltagi (ed.), *A Companion to Theoretical Econometrics*, Oxford: Basil Blackwell, 310–330.
- Bajari, Pat, Lanier Benkard and Jon Levin (2007) "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, Vol. 75, No. 5, 1331–1370.
- Bajari, Patrick, Han Hong, and Stephen Ryan (2006). "Identification and Estimation of a Discrete Game of Complete Information," working paper, University of Minnesota.
- Bajari, Patrick, Han Hong, John Krainer and Denis Nekipelov, (2006). "Estimating Static Models of Strategic Interactions," working paper, University of Minnesota.
- Bala, V. and S. Goyal (2000) "A Non-Cooperative Model of Network Formation," *Econometrica*, 68, 1181-1229.
- Bell, David and Sangyoung Song (2007), "Neighborhood Effects and Trial on the Internet: Evidence from Online Grocery Retailing," *Quantitative Marketing and Economics*, 5(4), 361-400.
- Bothner, M. (2003) "Competition and Social Influence: The Diffusion of the Sixth-Generation Processor in the Global Computer Industry," *American Journal of Sociology*, Volume 108 Number 6 (May): 1175–1210
- Bothner, Matthew S., Young-Kyu Kim and Wonjae Lee (2006), "Primary Status, Complementary Status, and Capital Acquisition in the U.S. Venture Capital Industry," Working Paper, University of Chicago.
- Burt, Ronald S. (1987), "Social Contagion and Innovation: Cohesion versus Structural Equivalence," *American Journal of Sociology*, 92 (May), 1287-1335.
- Coleman, James S., Elihu Katz and Herbert Menzel (1966), *Medical Innovation*, New York: Bobbs-Merrill.
- Conley, Timothy, and Christopher Udry (2003). "Learning About a New Technology: Pineapple in Ghana," Working Paper, University of Chicago.
- Dodds, P. S. and D. J. Watts (2005) "A Generalized Model of Social and Biological Contagion," *Journal of Theoretical Biology*, Vol. 232, pp. 587–604.
- Dodds, P. S. and D. J. Watts, (2004) "Universal Behavior in a Generalized Model of Contagion," *Phys. Rev. Lett.*, Vol. 92, article 218701.
- Doraszelski, U. and K. Judd, (2007) "Avoiding the Curse of Dimensionality in Dynamic Stochastic Games," working paper.
- Draganska, M., et al. (2008), "Discrete Choice Models of Firms' Strategic Decisions," *Marketing Letters*.
- Dubé, Jean-Pierre, Günter J. Hitsch and Pradeep Chintagunta (2007), "Dynamic Standards Competition and Tipping: The Case of 32/64 Bit Video Game Consoles," Working Paper, University of Chicago.
- Durrett, R., "Stochastic Spatial Models," *SIAM Review*, 1999.
- Economides, Nicholas and Charles Himmelberg (1995). "Critical Mass and Network Size with Application to the US FAX Market," Working Paper, New York University.
- Fleder, Daniel M. and Hosanagar, Kartik, (2007). "Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity," Working Paper, University of Pennsylvania.

- Gladwell, Malcolm (2000), *The Tipping Point: How Little Things Can Make a Big Difference*, Little Brown Publishers.
- Glaeser, E. and J. Scheinkman (2001) "Measuring Social Interactions," in *Social Dynamics*, ed. by S. Durlauf and P. Young. Cambridge: MIT Press.
- Glaeser, Edward L., Bruce I. Sacerdote, Jose A. Scheinkman, (2003), "The Social Multiplier," *Journal of the European Economic Association*, April-May.
- Godes, David, and Dina Mayzlin. (2004). "Firm-Created Word-of-Mouth Communication: A Field-Based Quasi-Experiment," Working Paper, Harvard University.
- Granovetter, M (1973), "The Strength of Weak Ties," *American Journal of Sociology*, 78 (May), pp 1360-1380.
- Granovetter, Mark (1978). "Threshold Models of Collective Behavior," *American Journal of Sociology*, Vol. 83, No. 6, pp 1420-1443.
- Hartmann, Wesley and Taylan Yildiz (2006), "Preferences for Joint Consumption and the Implications for Targeted Marketing," Working Paper, Stanford University.
- Katz, L., A. Kling, and J. Liebman (2001), "Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment," *Quarterly Journal of Economics*, CXVI, 607 - 654.
- Lee, L.F. (1982). "Some Approaches to the Correction of Selectivity Bias," *The Review of Economic Studies* XLIX, 355-372.
- Manchanda, Puneet, Ying Xie and Nara Youn (2004), "The Role of Targeted Communication and Contagion in New Product Adoption," *Marketing Science*, forthcoming.
- Manski, C. F., (1993), "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies* 60, pp. 531-542.
- Manski, C.F. (2000). "Economic Analysis of Social Interactions," *Journal of Economic Perspectives*, vol. 14(3), pages 115-136.
- Moffitt, R., (2001). "Policy Interventions, Low-Level Equilibria, and Social Interactions," in Durlauf, S. and Young, P. (Ed.) *Social Dynamics*, Brookings Institution Press and MIT Press 45-82
- Nair, Harikesh, Puneet Manchanda and Tulikaa Bhatia (2006). "Asymmetric Peer Effects in Prescription Behavior: The Role of Opinion Leaders," Working Paper, Stanford University.
- Nam, Sungjoon, Puneet Manchanda and Pradeep K. Chintagunta (2006), "The Effects of Service Quality and Word-of-Mouth on Customer Acquisition, Retention and Usage," Working Paper, University of Chicago.
- Narayan, Vishal and Sha Yang (2006). "Trust between Consumers in Online Communities: Modeling the Formation of Dyadic Relationships," Working Paper, NYU.
- Podolny, Joel (2005), *Status Signals: A Sociological Study of Market Competition*, Princeton University Press.
- Ryan, Stephen and Catherine Tucker (2006). "Heterogeneity and the Dynamics of Technology Adoption," Working Paper, MIT.
- Sacerdote, Bruce. (2001), "Peer Effects With Random Assignment: Results For Dartmouth Roommates," *Quarterly Journal of Economics*, 2001, v116(2,May), 681-704.
- Salganik, M. J., P. S. Dodds, and D. J. Watts (2006) "An Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market," *Science*, Vol. 311, pp. 854–856.
- Schelling, T. (1971). "Dynamic Models of Segregation," *Journal of Mathematical Sociology*, 1, 143-186.

- Trusov, M. (2006) "Your Agents Are Also Your Customers: Marketing for Internet Social Networks," Working Paper, University of Maryland.
- Tucker, C. (2006), "Interactive, Option-Value and Domino Network Effects in Technology Adoption," working paper, MIT.
- Van Den Bulte, Christophe and Gary L. Lilien. (2001), "Medical Innovation Revisited: Social Contagion versus Marketing Effort," *American Journal of Sociology*, 106(5), 1409-35.
- Van Den Bulte, Christophe and Stephan Wuyts (2007), *Social Networks and Marketing*, Marketing Science Institute, Cambridge, MA.
- Watts, D. J. (2002), "A Simple Model of Global Cascades on Random Networks" *Proc. Natl. Acad. Sci.*, Vol. 99, pp. 5766–5571.
- Watts, D. J., et.al. (2005), "Multiscale, Resurgent Epidemics in a Hierarchical Metapopulation Model" *Proc. Natl. Acad. Sci.*, Vol. 102, No. 32, pp. 11157-11162.
- Watts, D.J. and P.S. Dodds, (2007) "Influentials, Networks, and Public Opinion Formation," forthcoming *Journal of Consumer Research.*, Vol. 34.
- Yang, Sha and Greg M. Allenby (2003), "Modeling Interdependent Consumer Preferences," *Journal of Marketing Research*, 40(3), 282-294.
- Yang, Sha, Vishal Narayan and Henry Assael (2006), "Estimating the Interdependence of Television Program Viewership between Spouses: A Bayesian Simultaneous Equation Model," *Marketing Science*, 3, 336-349.