ON THE IMPACT OF SOCIAL NETWORKS ON CHARITABLE BEHAVIOUR: THEORY AND EVIDENCE

ROGAYEH DASTRANJ*
WORKING PAPER
OCTOBER 2015

Abstract. I study the direct and spillover effects of social interactions using in a network of volunteers from Engineers Without Borders (EWB), Canada. I model social interactions as a network game in which agents simultaneously choose their effort levels, taking the network and their friends' efforts as given. The effects from social interactions are introduced through two separate channels: a strategic interaction term which affects the marginal benefit from supplying effort and a direct spillover term affecting the level of an agent's payoff. I construct three different categories of online and offline networks and estimate the model using instrumental variables and system GMM. The identification strategy relies on the yearly variation in the location of the EWB national conference and new members' participation levels in this event each year. The estimates demonstrate different patterns for engagement versus fundraising activities. Large significant levels of strategic complementarities are always present in fundraising activities regardless of the definition of links, however, in engagement activities, strategic complementarities are only significant in online networks. Additionally, engagement activities exhibit positive significant levels of direct spillovers for all networks. In contrast, in fundraising campaigns, the direct spillover effect is only significant in large offline networks.

1. Introduction

Social networks have value: the whole is greater than the sum of its parts. New properties emerge because of individuals' embeddedness in social networks, and these properties inhere in the structure of the networks, not just in the individuals within them. The recognition that social connections play an essential role in explaining various social and economic phenomena has sparked a growing literature that studies social structures that emerge in different situations and how they impact individual behaviours and outcomes.

*Department of Economics, Simon Fraser University.
I would like to thank my advisors Alex Karaivanov, Brian Krauth and Krishna Pendakur for detailed comments and suggestions and for extensive advice and support; and thank Matthew Jackson, Arun Chandrasekhar, Aureo de Paula, Anke Kessler, Chris Muris, Fernando Aragon, Lucas Herrenbrueck for helpful conversations, comments and feedback.

Studying social interactions and identifying peer effects has been the focus of many papers which have tried to shed light on the channels whereby an individual's peer group impacts their own outcomes. In contrast to most of the existing literature, in this paper I study two separate channels through which social interactions could impact an individual's decisions and outcomes. To do so, I model the interaction among individuals as a simultaneous move game in a given network, in which individuals benefit from their own effort as well as the average of their friends' efforts through two channels: (1) *a strategic interaction term* which impacts the marginal benefit from supplying the effort, and (2) *a direct spillover term* affecting the level of the benefit. Direct spillover term measure the effect of one's peers on own outcome, while strategic interaction term indicates one's peers' effect on one's behaviour or effort choice. For example, a positive coefficient of the strategic interaction term indicates presence of strategic complementarities among members, hence, in organizing a conference for instance, the harder one's friends work, the harder the individual would work; in addition, extra effort from one's friends could directly contribute to a better outcome and turnout for the conference (direct positive spillover). Note that the identification of the effort best response equation suffers from the standard endogeneity issues and the "reflection problem"[1] that need to be addressed appropriately. However, properly estimating the best response equation at best recovers the reduced form parameter which is the ratio of the coefficients of strategic interaction term to marginal cost[2] and provides no information about the direct spillover effect. To address this issue, using raw data from Engineers Without Borders (EWB)[3], I manually construct a dataset

[1]Manski (1993) characterizes three separate channels for peer effects: (1) endogenous effects, (2) exogenous or contextual effects, and (3) correlated effects. Endogenous effects are present when individuals simultaneously influence each others' decisions, whereas the contextual effects arise when an individual's decisions are impacted by his friends' exogenous characteristics such their parents' socio-economic status. Finally, correlated effects are present when a potentially unobserved factor influences both the individual and his peers. Identification of these peer effects is specially problematic (even in the absence of the correlated effects) in linear-in-mean models, because of the perfect collinearity between the peer group's mean characteristics and their mean outcome. This is the well-known *reflection problem*. This problem is especially pronounced when all the individuals are assumed to have the same peer group.

[2]One should use caution in interpreting the coefficients that are obtained from estimating the best response equation only, since there could be several structural models that lead to the same best response equation. For more detailed discussion see: Boucher, Vincent, and Bernard Fortin. "Some Challenges in the Empirics of the Effects of Networks." (2015): 15-04.

[3]EWB is a Canadian non-governmental organization (NGO), founded in 2001 that is engaged in international development work both in Canada and in a selection of countries in Africa. Their work in Canada focuses on policy, education, and advocacy initiatives while their work in Africa comprises capacity building initiatives in sectors including agriculture, water and sanitation, and governance. EWB's Canadian membership is comprised of volunteers organized into local chapters. The chapters are a loosely hierarchical structure led by an elected president and an executive team comprising 10-15 highly involved members. Chapters are typically associated with a university, and range in size from tens to hundreds of members. More information about EWB can be found at http://www.ewb.ca.

that contains information on several distinct *social networks*, individuals' *choices of effort*, as well as *associated outcomes*. This is different from most other datasets used in studying peer effects which usually do not have detail information on the network structure, and also only contain *either* the effort choice *or* the outcome variables. Using this unique dataset, I am able to jointly estimate the best response equation with the corresponding equilibrium benefit equation, and separately recover the impact of the two network effects – the strategic interaction and the direct spillover term in the structural model.

The aspects of my dataset that makes the estimation of the full model possible are three fold. First, it provides information on several dimensions of members' interactions within the organization. Specifically, it distinguishes between several online and offline (in person) measures of social links, making it possible to study how different kinds of interactions influence both individual choices and outcomes. In particular, I construct three types of networks based on: (1) direct messages between members, (2) post and reply relationships in online forums, and (3) membership in common offline groups[4]. Second, the data includes information not only on individual characteristics and choices of effort, but also on the outcomes of these activities, enabling me to jointly estimate the best response and the equilibrium benefit equations to recover the two distinct network effects. Finally, the dataset enables me to estimate the effect of own and peers' decisions on two different measures of charitable activities: (1) engagement activities such as member learning, public engagement and advocacy, etc. that mostly involves donating time, and (2) peer to peer online fundraising campaigns. The first category includes activities that require in person interactions among individuals whereas the second is a online campaign through which individuals invite their family and friends to contribute to their campaign[5].

I estimate the structural model via IV and system GMM and find that (a) in engagement activities the network game among members is one of *strategic complementarities* (positive strategic interaction term), especially when the peer groups are defined based on the *online interactions*, whereas strategic complementarities are always present in the fundraising activities regardless of the type of the network. Observing strategic complementarity as opposed to strategic substitutability implies that an increase in a member's own effort, positively impacts his friends. In particular, I find no evidence for free riding behaviour. This is in contrast to the common view that free-riding is a

Most of the data used in this research originates from a purpose-built website (myewb.ca) used by EWB members to communicate and coordinate their efforts.

[4]For more details see Section 3.

[5]All the contributions are fully donated to EWB.

problem, especially in the context of charitable behaviour and private provision of public goods. Note that if the free-riding problem were present, the estimated coefficients on the interaction between own effort and the efforts of others should exhibit strategic substitution, which is not the case in any of the empirical specifications. Furthermore, I find that (b) direct spillovers play an important role, especially in engagement activities, but also in fundraising campaigns when the individual specific peer group is based on *offline interactions*. In engagement activities, the direct network spillover is present even when no strategic complementarity (or substitution) is detected.

In terms of magnitudes, in the fundraising activities, the parameter estimates imply that if, for example, an individual's peer group increase their effort by sending 100 more emails, he would send up to 45 extra emails (strategic complementarity) depending on whether peer group is defined based on his online or offline interactions. Additionally, 100 extra emails sent on average by his friends, contributes, on average, 550$ directly to his total fundraising amount (direct spillover effect). In engagement activities such as a public outreach event, 10 hours increase in the average of one's peers' volunteering time results in 3 extra hours of volunteering, when his peer group is defined based on online interaction among members. In contrast, I do not find evidence of strategic complementarities among members in offline networks. The direct spillover effect, however, is always significant and positive regardless of the definition of the peer group. One standard deviation increase in the average of peer's efforts, contributed between 0.2 to 0.5 standard deviation to the outcome from a specific event or activities.

To address the reflection problem and the endogeneity of individual effort levels, I take advantage of the exogenous yearly variations in the location of the national conference of Engineers Without Borders, and its impact on the participation levels of new members. The national conference is the most important event of the year and takes place every year in mid January, four months after most of new members join the organization. The location of the national conference is decided by the national office and changes every year to provide equal opportunities for all members across Canada, and is announced around late October or early November when registration opens. Therefore, the location of this conference can be considered exogenous to the characteristics of new members who join EWB in September, as the school year begins. Since the vast majority of new members do not know much about the organization and the importance of the national conference, their decision to attend is mostly dictated by the travel costs. I provide evidence that attending the national conference in the first year of joining EWB has a significant positive impact on members' engagement and participation levels through out their involvement with EWB. Therefore, I use the location and distance to the national conference in the first year of joining EWB as instruments for the endogenous effort variables, and use standard IV approaches as well as system GMM to estimate the parameters of the model and discuss how the reduced-form parameters map into the structural ones.

Finally, to establish the robustness of my results, I use different definitions for the networks that assume a looser or a more stringent characterization of links. I also take advantage of variations present in the network structure by integrating ideas discussed in Bramoulle, Djeebari, and Fortin (2009), and use "friends of friends'" and "friend of friends of friends'"covariates as instruments for the endogenous variables. All of these specification demonstrate that my baseline estimates are robust. Also the results are unchanged even when only a subset of instruments are used.

1.1. **Literature Review.** This paper is related to several literatures. First, I build on and extend the peer effects literature. As mentioned earlier Manski's paper sets out much of the basic terminology and concerns of this literature and shows that in the context of the linear-in-means model the three types of effects discussed earlier in footnote (1) are not separately identified, but that the policy implications of the three types are different. A large literature have tackled this problem through various approaches such as using random assignments to groups (e.g. Sacerdote (2001), Hoxby (2000), Hanushek et al. (2003), Burke and Sass (2011)), an exclusion restriction (e.g. Gaviria and Raphael (2001), Fletcher (2012), Duflo and Saez (2003)), conditional variances (e.g. Glaeser et al (1996), Graham (2008), Friesen and Krauth (2007)), or using other approaches for identification in nonlinear models (e.g. Brock and Durlauf (1995), Card and Giuliano (2012), and Krauth (2006))[6].

More closely related papers to this study, are the identification approaches discussed in Bramoulle et al (2009), Calvo-Armengol et al. (2009) and De Giorgi et al. (2010). Note that Manski's paper assumes the social network takes the form of a set of disconnected groups (i.e. treating them as complete networks). In contrast, the above papers consider more complex social networks, and exploit a simple idea to get the identification: endogenous effects imply that an individual's behaviour is influenced indirectly by the friends of his friends, while contextual effects do not. More specifically, Bramoulle et al (2009) shows that in the absence of correlated effects and under certain conditions on the network structure, the covariates of friends of friends could be use to instrument the average of friends' choices in the reduced form linear in mean model. Unfortunately, the assumption of the no correlated effects may not hold in many different contexts. Therefore, I only include this method as a robustness check

[6]For a comprehensive overview of applied literature on on the identification of social interactions see Blume and coauthors (2010) and de Paula (2015).

in this paper, and use the exogenous location and distance to the national conference instruments in my baseline specifications.[7]

Another area of related research is the literature on network games.[8] The theoretical model used in this study is an adaptation of Bramoulle, Kranton and D'Amours (2012) and Dastranj, Karaivanov and Easton (2015), allowing for the presence of strategic complementarities as well as substitutability. Acemoglu et al (2015) use a similar methodology to study the direct and spillover effects of local state capacity using the network of Colombian municipalities. To my knowledge, this paper is the only other paper that uses the benefit equation equation jointly with the best response function to estimate the parameters of the structural model.

Finally, this paper relates to the research on charitable giving. There is a large literature on this topic that dates back to Adam Smith (1759)[9]. In particular, there are several papers that discuss how charitable behaviour might be motivated by a desire to receive social acclaim or prestige, as well as potentially conforming to some social norms[10]. In a recent paper, Karlan and McConnell (2014) conduct a field study with donors to Yale University to test the impact of a promise of public recognition on giving, and show that charitable gifts increase in response to the promise of public recognition primarily because of individuals' desire to improve their social image[11]. Other papers that look at the charitable behaviour and peer pressure (e.g. DellaVigna et al. (2010) and Carman (2004)) and find strong evidence for the presence of social pressure, similar to the strategic complementarities that I found in several networks. In addition, using data from a university, Meer (2010) analyzes whether alumni are more likely to give, and give larger amounts when they are solicited by someone with whom they have social ties. They show that individuals are much more likely to donate, and donate larger amounts when asked by an acquaintance with whom they share similar characteristics[12].

The rest of this paper is organized as follows. The structural model is presented in section 2 followed by introducing the dataset in section 3 and discussing specific factors used in defining the networks and the measures for the parameters of the model. In section 4, I discuss the empirical strategy and the exclusion restriction assumptions in

[7]More specifically, I use own, friends', and friends of friends' location and distance to the national conference as exogenous instruments, for more information see Section 4.

[8]For a comprehensive review of literature see Jackson (2008).

[9]See Andreoni (1989 and 1990), Glazer and Konrad (1996), Andreoni (1998), Benabou and Tirole (2005), Mayo and Tinsley (2009), Lilley and Slonim (2014), Edwards and List (2014), and many more. For two thorough surveys of this literature see Vesterlund (2006) or Andreoni (2006).

[10]See Becker (1974), Bernheim (1994) and Harbaugh (1998).

[11] Karlan et al (2014)

[12]For earlier studies on the effect of own giving by one's reference groups see Feldstein and Clotfelter (1976) and Andreoni and Scholz (1998).

detail and present the results of estimating the model in section 5. Finally, the outcomes of the robustness checks are demonstrated in section 6 followed by the conclusion.

2. The Model

In this section, I present a theoretical model that builds on public goods games in networks, in which individuals benefit from their own efforts/activity as well as their friends'. The latter component of the benefit function is captured through two separate additive terms: (1) an interaction term that represents strategic complementaries or substitutabilities affecting the marginal product of effort, and (2) an aggregate term that represents any spillovers or externalities and impacts the level of the benefit function directly. There are also costs associated with own effort.

Formally, let $\mathbf{G}$ represent the network of members with the following payoff functions:

$$\begin{aligned}
&\Pi_i(e_i, e_{-i}, \mathbf{G}) = B_i - C_i \\
&B_i(e_i, e_{-i}, \mathbf{G}) = e_i(\alpha x_i + \eta_i) + \gamma e_i \sum_{j=1}^{N} g_{ij} e_j + \lambda \sum_{j=1}^{N} g_{ij} e_j + \delta x_i + \epsilon_i \\
&C_i(e_i, e_{-i}, \mathbf{G}) = \frac{\pi}{2} e_i^2
\end{aligned}$$

Above, x_i captures the strength of the observable component of own effort on the total benefit function, B_i, and η_i captures the strength of the unobserved component. γ represents the effect of the strategic interaction term between own effort and friends' effort (strategic complementarity if $\gamma > 0$), whereas λ captures the direct effect of friends' effort on i's outcome (spill overs). ϵ_i are mean-zero unobservables both to individuals and the modeller. Finally, π is the marginal cost parameter of own effort. Additionally, the network $\mathbf{G}$ is defined as:

$$\begin{cases} g_{ij} = \frac{1}{d_i} & \text{if } i \text{ and } j \text{ are linked} \\ g_{ji} = \frac{1}{d_j} & \text{if } i \text{ and } j \text{ are linked} \\ g_{ij} = g_{ji} = 0 & \text{if } i \text{ and } j \text{ are not linked} \\ g_{ii} = 0 & \end{cases} \tag{2.1}$$

where d_i is the degree (total number of links) of agent i.[13] As a result, the first order condition for optimal effort choice is as follows:

$$(\alpha x_i + \eta_i) + \gamma \sum_{j=1}^{N} g_{ij} e_j - \pi e_i \begin{cases} < 0 & \text{if } e_i \leq 0 \\ = 0 & \text{if } e_i > 0 \end{cases} \tag{2.2}$$

[13]Note that dividing by d_i results in considering the average of peers' effect in the payoff function.

Existence and uniqueness conditions for this game is provided by Bramoulle, Kranton and D'Amours (2012) and Corbo, Calvo-Armengol and Parkes (2007) and depends on the minimum eigenvalue of the network:

LEMMA **1.** *(Bramoulle, Kranton, and D'Amours (2012)). If $|\alpha_{min}(\mathbf{G})| < \frac{\pi}{\gamma}$, then there exists a unique interior equilibrium*[14].

From now on I assume that $|\alpha_{\min}(\mathbf{G})| < \frac{\pi}{\gamma}$ condition is satisfied, so we get $e_i = f_i^{BR}(\mathbf{e}, \mathbf{G})$, where:

$$f_i^{BR}(\mathbf{e}, \mathbf{G}) = \max\{0,\ \frac{(\alpha x_i + \eta_i)}{\pi} + \frac{\gamma}{\pi}\sum_{j=1}^{N} g_{ij}e_j\} \tag{2.3}$$

And if the parameters are such that the solution is interior, then:

$$e_i = \frac{\alpha x_i + \eta_i}{\pi} + \frac{\gamma}{\pi}\sum_{j=1}^{N} g_{ij}e_j$$

$$e_i = \bar{\alpha} x_i + \bar{\gamma}\sum_{j=1}^{N} g_{ij}e_j + \bar{\eta}_i \tag{2.4}$$

where $\bar{t} = \frac{t}{\pi}$. Equation (2.4) presents the source of the endogeneity problem: an agent's effort choice depends on his own characteristics as well as the average of his friends' choice of effort. Note that, even if this equation is identified properly, the estimated coefficient ($\bar{\gamma}$) are not the parameters of the structural model since they are normalized by the marginal cost parameter, π. However, for the interior solution, by substituting

[14]Note that a potential function associated with the payoff function is

$$\Phi(\mathbf{e}, \mathbf{G}) = \sum_{i=1}^{N} e_i(\alpha x_i + \eta_i + \frac{\gamma}{2}\sum_{j=1}^{N} g_{ij}e_j) + \delta\sum_{i}^{N} x_i - \frac{\pi}{2}\sum_{i}^{N} e_i^2$$

$$\text{where} \quad \frac{\partial\Phi}{\partial e_i} = \frac{\partial\Pi_i}{\partial e_i}$$

In the matrix notation, this potential function can be written as:

$$\Phi(\mathbf{e}, \mathbf{G}) = \mathbf{e}'(\alpha\mathbf{x} + \eta) - \frac{\pi}{2}\mathbf{e}'(\mathbf{I} - \frac{\gamma}{\pi}\mathbf{G})\mathbf{e} + \delta\mathbf{x1}$$

$$\Rightarrow \quad \nabla^2\Phi = -\frac{\pi}{2}(\mathbf{I} - \frac{\gamma}{\pi}\mathbf{G})$$

Therefore, the potential function is strictly concave when $(\mathbf{I} - \frac{\gamma}{\pi}\mathbf{G})$ is positive definite. Corbo, Calvo-Armengol and Parkes (2007) show that in games with pure strategic complementarities, this is the case when $\lambda_{\max}(-\mathbf{G}) < \frac{\pi}{\gamma}$ (where $\lambda_{\max}$ is the maximum eigenvalue of matrix $-\mathbf{G}$). They also show that the existence of an interior equilibrium is guaranteed by the same condition. However, note that $\lambda_{\max}(-\mathbf{G}) = -\lambda_{\min}(\mathbf{G})$. Therefore, this condition coincides with Bramoulle, Kranton and D'Amours (2012) condition for the uniqueness of equilibrium.

equation (2.4) in the benefit function $B(.)$, we get:

$$\begin{aligned}B_i(e_i, e_{-i}, \mathbf{G}) &= e_i(\alpha x_i + \eta_i) + \gamma e_i \sum_{j=1}^{N} g_{ij}e_j + \lambda \sum_{j=1}^{N} g_{ij}e_j + \delta x_i + \epsilon_i \\ &= e_i[(\alpha x_i + \eta_i) + \gamma \sum_{j=1}^{N} g_{ij}e_j] + \lambda \sum_{j=1}^{N} g_{ij}e_j + \delta x_i + \epsilon_i \\ &= e_i[\pi e_i] + \lambda \sum_{j=1}^{N} g_{ij}e_j + \delta x_i + \epsilon_i\end{aligned}$$

$$\Rightarrow \qquad B_i(e_i, e_{-i}, \mathbf{G}) = \pi e_i^2 + \lambda \sum_{j=1}^{N} g_{ij}e_j + \delta x_i + \epsilon_i \tag{2.5}$$

Therefore, by taking advantage of the appropriate sources of data and simultaneously estimating Equations (2.5) and (2.4), one could recover all the parameters of the model and the channels through which social interactions impacts behaviour and outcomes. To estimate the following equations empirically,

$$e_i = \frac{\alpha x_i + \eta_i}{\pi} + \frac{\gamma}{\pi} \sum_{j=1}^{N} g_{ij}e_j$$

$$B_i(e_i, e_{-i}, \mathbf{G}) = \pi e_i^2 + \lambda \sum_{j=1}^{N} g_{ij}e_j + \delta x_i + \epsilon_i$$

I assume

$$\mathbb{E}[\epsilon_i | x_i, \mathbf{G}] = 0 \qquad \mathbb{E}[\eta_i | x_i, \mathbf{G}] = 0 \tag{2.6}$$

Note that the effort level in the best response equation (2.4) is endogenous, i.e. own effort depends on friends' efforts. Additionally, there might be unobservables that are correlated with friends' efforts in both equations, i.e. $\text{cov}(\sum_{j=1}^{N} g_{ij}e_j, \eta_i) \neq 0$ and $\text{cov}(\sum_{j=1}^{N} g_{ij}e_j, \epsilon_i) \neq 0$ (i.e. $\mathbb{E}[\epsilon_i\epsilon_j | x_i, \mathbf{G}] \neq 0$ and $\mathbb{E}[\eta_i\eta_j | x_i, \mathbf{G}] \neq 0$)[15]. Also, since the effort level is endogenous, in the benefit equation (2.5), we may have $\text{cov}(e_i, \epsilon_i) \neq 0$ (i.e $\mathbb{E}[\eta_i\epsilon_j | x_i, \mathbf{G}] \neq 0$). Therefore, the endogeneity problem and the correlated effects need to be appropriately addressed potentially through suitable instruments that are orthogonal to the omitted own and friends' effort levels. In other words, the instruments, $\mathbf{z}$, need to have the following properties:

$$\mathbb{E}[\epsilon | \mathbf{z}, \mathbf{G}] = 0 \qquad \mathbb{E}[\eta | \mathbf{z}, \mathbf{G}] = 0 \tag{2.7}$$

Having this, π (the marginal cost parameter) and λ (the spill over effect) could be estimated from the benefit equation (2.5), and using π, I recover γ (the strength of

[15]Correlated effects in Manki's terminology.

the interaction effect) from the estimate of the peer effect term in the best response equation (2.4), and therefore recover all the parameters of the structural model.

3. Data

Data for this study has been mostly constructed using the Engineers Without Borders' online platform called myewb.ca. This website was launched in late 2003 and its purpose was to facilitate communication across different chapters. It was also used for planning activities and inter-chapter communications. In addition, myewb was set up to allow users share information and exchange ideas with other members across the organization through forums, threaded conversations, wiki pages, whiteboards and more. Using this website, I have constructed several online and offline networks as well as the effort and outcome variables. The details are provided in the following paragraphs.

Upon joining EWB, members were encouraged to open a myewb account to make sure that they received communications about events and a wide variety of activities that are planned by chapters. By creating a profile on myewb, individuals could participate in different forum and online conversations, share information, and also join groups or start groups based on their (offline) activities. Overall, this environment provides a rich dataset to study the interaction of individuals and their outcomes through several channels. Table 1 provides summary of the descriptive statistics of individual characteristics available in the dataset.

Table 1. Descriptive Statistics of Individual Characteristics

Statistic	Mean	St. Dev.	Min	Max
Age	25.61	8.84	15	65
Chapter Size	97.54	64.86	1	281
Male dummy	0.59	0.49	0	1
English dummy	0.91	0.28	0	1
Student dummy	0.19	0.39	0	1
Work dummy	0.07	0.26	0	1
Both student & work dummy	0.02	0.14	0	1
Distance to National Conference	2,232.88	1,884.75	0	7,512
Total Login	12.81	90.61	1	4,858
Total group membership	7.25	15.44	1	225
Total Conferences	0.19	0.74	0	10

In order to estimate the structural model presented in the previous section, three key components are needed: (1) networks that characterize social ties, (2) proxies for individual effort, and (3) proxies for individual benefit or outcome from exerting that effort.

3.1. **Networks.** To build the undirected networks used in this study, 3 different criteria are used: (1) *direct messages*: a link between two individuals indicates that at least one of them sent a message to the other person through myewb. (2) *threaded comments*: there is a link between two agents if there exists a post and reply relationship between these individuals (i.e. one individual writes a post and the other replies to that post). Note that there could be more than one reply to a post, hence, it is assumed that all the agents who are part of the same thread are connected. Also, there could be posts that no one replied to which could result in having isolated agents with no connections. (3) *group membership*: individuals who are members of the same groups are assumed to be connected. Note that an agent could be part of many different groups which vary in size from just a handful of members to groups with hundreds of members. Therefore, in the default group membership network only groups with less than 50 members are considered. Later in the robustness checks, I use a broader as well as a more stringent requirement for the definition of these links. Table 2 summarizes the characteristics of the three different networks I construct.

The degree distribution and the average of individuals' neighbour degrees versus individual's degree are plotted in figures (4) and (5). These plot show that there is a fairly linear decay in the log-frequency as a function of log-degree, and also suggest that while there is a tendency for individuals of higher degrees to link with similar individuals, people with lower degree tend to link with individuals of both lower and higher degrees. It is reassuring to observe these characteristics in the constructed networks as these properties are present in many other social networks and distinguish these networks from randomly generated ones.[16]

Table 2. Network Characteristics

Network	Diameter	Avg Path Length	Density	Transitivity
Messages	22	3.94	0.002	0.091
Threaded Comments	69	2.77	0.013	0.386
Group Membership	6	2.56	0.017	0.44

Note that the threaded comments and messages networks are mostly based on *online relationship* between individuals who may not even live in the same city. In contrast, the networks made using the group membership data is mostly based on *offline and in person relationships*.

The correlation between several centrality measures and effort and outcome variables (discussed in the following section) is presented in tables 19 and 20. In all categories,

[16]For more information see: Kolaczyk, Eric D., and Gabor Csardi. Statistical analysis of network data with R. Vol. 65. Springer, 2014.

the correlation between degree and eigenvector centrality with the effort and outcome variables are larger than the betweenness and closeness centralities. This is interesting, since, with strategic complementarities, own effort is increasing in the number of friends (i.e positive correlation with degree centrality and eigenvector centrality).

3.2. **Effort And Outcome Variables.** I also construct three separate categories of activities that provide proxies for the effort and benefit variables of the structural model: (1) *member engagement activities* recoded through CHAMP, (2) *online peer to peer fundraising campaigns*, and (3) *general activities.* In this section, I provide a brief overview of these activities and discuss the variables that are used in this study.

3.2.1. *Engagement Activities (CHAMP).* In 2005, a new data collection system was introduced across EWB called CHAMP. Data available in CHAMP is comprised of 9 categories: Member learning, fundraising, school outreach, workplace outreach, advocacy letters, curriculum enhancement, functioning, public engagement and publications. The variables that capture members input or effort are the "preparation hours" and the "execution hours". For each individual in the CHAMP dataset, the preparation and execution hours are added to calculate "total hours" spent volunteering that is used to proxy variable e_i in estimating the best response and the benefit equations[17].

In contrast, the outcome variables, B_i, are different for each of the above categories. For example, for political advocacy, it is number of letters sent to MPs (member of parliament), whereas for member learning or outreach categories, it is the number of participants. In order to make these categories comparable, the outcome variables are scaled between $[1, 10]$ using a scaling function that retains the rank order and the relative size of separation between any two values. More precisely, the minimum and maximum values are set equal to 1 and 10 respectively and all the other values are scaled based on their distance from this maximum and minimum, so that the resulting scaled outcome is still a continuous variable.

[17]The purpose of CHAMP was to replace monthly chapter reporting. This information sharing and how it was used was critical in supporting chapter performance. Using CHAMP, chapter Presidents could have their entire teams contribute activities, which would only need to be reviewed by Presidents with a quick skim and click to approve. It also allowed archiving, tracking against plans and goals, etc. In a reply to a question about CHAMP, George Roter (co-CEO of EWB) wrote:

> "The result [of using CHAMP] was improved chapter support, and certainly meant staff were spending their time better. I would say CHAMP was one important reason for chapters having reached their peak in activity levels and capacity."

Unfortunately, due to several challenges such as rigidity in terms of indicators and activities, lack of user-friendly input environment, and fundamental internal changes in strategy and direction of EWB, CHAMP was abandoned toward the end of 2010 by National Office, although some chapters kept using it even 2 years after it was abandoned.

3.2.2. *Fundraising Activities.* The other sources of data, are several websites that have been used for running EWB's fundraising campaigns since 2010. By participating in several campaigns such as Run to End Poverty, Dream campaign or "Everyday Innovation" Holiday Campaign, EWB members have raised funds to support both Canadian and African programs. Members may sign up for these campaigns individually or as part of team, and depending on the nature of the activity, they may create a page with their personal statements of why they support EWB. To raise funds, they send several emails inviting their friends and family to contribute to their campaign. The information from these websites has been aggregated through a platform that tracks the number of emails individuals sent for each campaign, how many visits their page got, how many people donated and how much and much more. In order to estimate the parameters of the structural model using this data, the "total number of emails" is used as a proxy for individuals' effort, and the "total money raised" is used as a measure of outcome or benefit from this activity and effort.

The characteristics of the networks used in estimating the parameters of the model are listed in table 3 for both the Engagement and the Fundraising datasets.

TABLE 3. Characteristics of Networks used in Engagement and Campaign datasets

	Netwok Variables:				
Network	Avg Degree	Diameter	Avg Path Length	Density	Transitivity
	Engagement Dataset				
Messages	7.748	7	3.08	0.01	0.20
Threaded Comments	46.38	6	2.23	0.07	0.40
Group Membership	75.41	4	2.04	0.10	0.41
	Campaign Dataset				
Messages	6.44	9	3.4	0.01	0.20
Threaded Comments	33.64	5	2.33	0.06	0.40
Group Membership	69.93	4	2.08	0.10	0.46

3.2.3. *General Activities.* Additionally, by combining several data files, three additional proxies of effort variable are constructed: (1) *group membership*, (2) *total number of logins into myewb*, as well as (3) *total number of national conferences* that individuals have participated in. These variables are not associated with a direct benefit variable, but still could be used to estimate the best response equation and shed light on how different friendship networks impact members effort and participation levels measured using these variables. The following table summarizes the variables used to estimate the structural model.

Table 4. Effort and Outcome Variables

Activity	Effort	Outcome
Engagement Activities	Total Hours Volunteered	Scaled Outcomes
Online Fundraising Campaigns	Total Emails Sent	Total Amount Raised
General Activities	Total Group Membership Total Login in myEWB Total Conferences Attended	N.A.

4. Empirical Strategy and Exclusion Restriction

Previously, I argued that in order to separately recover the impact of the two network effects – the strategic interaction and the direct spillover term in the structural model, the best response equation needs to be estimated simultaneously with the benefit equation. However, the main problems in estimating the parameters of the following model:

$$e_i = \frac{\alpha}{\pi}x_i + \frac{\gamma}{\pi}\sum_{j=1}^{N} g_{ij}e_j + \frac{\eta_i}{\pi} \tag{4.1}$$

$$B_i(e_i, e_{-i}, g) = \pi e_i^2 + \lambda \sum_{j=1}^{N} g_{ij}e_j + \delta x_i + \epsilon_i \tag{4.2}$$

assuming:

$$\mathbb{E}[\epsilon_i | x_i, \mathbf{G}] = 0 \qquad \mathbb{E}[\eta_i | x_i, \mathbf{G}] = 0$$

is the endogeneity of the effort variable in the best response equation as well as the potential correlation between the unobservables and the friends' efforts in both equations (i.e. $\mathrm{cov}(\sum_{j=1}^{N} g_{ij}e_j, \eta_i) \neq 0$, $\mathrm{cov}(\sum_{j=1}^{N} g_{ij}e_j, \epsilon_i) \neq 0$, and $\mathrm{cov}(e_i, \epsilon_i) \neq 0$). Therefore, the instruments need to satisfy:

$$\mathbb{E}[\epsilon | \mathbf{z}, \mathbf{G}] = 0 \qquad \mathbb{E}[\eta | \mathbf{z}, \mathbf{G}] = 0 \tag{4.3}$$

One novel approach to address this problem is proposed by Bramoulle et al (2009), who use the exogenous network structure and the variation in the reference groups of individuals to find valid instruments for the endogenous variables (e.g. effort variable here). They show that in the absence of correlated effects, under certain conditions on the network structure[18], covariates of friends of friends, friends of friends of friends, etc could be used as valid instruments for the endogenous variable. The intuition is that friends of friends who are not directly friends with an individual, indirectly influence him through their impact on his friends' decisions. However, the main problem with this

[18] Bramoulle et al (2009) show that as long as I, G, G^2 are linearly independent matrices, $\{G^2X, G^3X, ...\}$ can be used as instruments for Gy, where X is a vector of individual characteristics and y is the choice or outcome variable.

strategy is the potential presence of unobserved correlated effects between individuals and their friends' and friends of friends' characteristics. Therefore, to credibly estimate the structural model, another source of exogenous variation that is not correlated with the error terms is needed to be used as an instrument for the endogenous variable.

The main source of exogenous variation which I use in this paper is the variation in the location of EWB's national conference each year, and how the distance from the national conference venue impacts the participation of new members in this event. National conference of Engineers Without Borders is the biggest and most important event of the year that between 500 to 1000 members, featuring several world class speakers and many parallel sessions and workshops. Although new members could join EWB at any time, most of them join the student chapters in September, as there is a big member recruitment effort at beginning of the school year. However, the location of the national conference (that changes every year) is decided by the national office of EWB and is announced late October, or early November. Therefore, it is reasonable to assume that the location of the national conference and individuals' distance to this event is orthogonal to members' unobserved characteristics. To show that notice that there is no correlation between a new member's month of joining EWB and his distance to the national conference in the first year, as presented in figure 4(C). In addition, the pattern of the month joined EWB is the same across people who attended the national conference in the first year and those who did not as seen in figure 4(A) and 4(B).

Attending the national conference and being exposed to such a high energy environments, specially in the first year of joining the organization, has two main impacts on individuals; on the one hand, it educates and motivates individuals, which could have a direct effect on their future choice of effort and participation in different activities. Furthermore, through meeting hundreds of other members from across the country, their social connections and networks become vastly different from individuals who did not attend the conference, creating exogenous variation in their social connections that goes beyond individuals they initially met at their university chapters. Therefore, since the location of national conference is exogenously determined and is uncorrelated with the characteristics of new members (*exclusion restriction*), the distance to the conference in the first year of joining EWB is used as an instrument for the endogenous effort variable.

Table 5, provides evidence that the distance to the national conference venue is significantly and negatively correlated with the dummy variable for attending the national conference in the first year of joining EWB even after controlling for several individual characteristics. Additionally, national conference has a direct impact on individuals' choice of effort and participation levels. This is shown in figures (2) and (3) which present the relationship between various effort variables such as group membership, total numbers of login, and total national conference attendance for individuals who attended the national conference in the first year, and those who did not.

Table 5. Probit regression estimation of the binary variable for attending the conference in the first year of joining EWB on the distance to the national conference.

	Dependent variable:	
	Binary variable for attending National Conference in the first year	
	(1)	(2)
Distance to National Conference	-0.222^{***}	-0.392^{***}
	(0.064)	(0.094)
Controls	no	yes
Observations	21,067	8,679
Log Likelihood	−6,681.275	−3,270.066
Note:	*p<0.05; **p<0.01; ***p<0.001	

Although attending the national conference at any time could have an impact on the individual, I only use the distance to the conference in the first year of joining EWB in this study. The intuition is that since, in the first year, an individual does not know much about the organization or the national conference, one of the main deciding factors in attending the conference is the distance and the financial cost of attending this event, specifically because of huge variations in the travel cost across Canada. However, later on, they may attend the conference because they have heard a lot about it, or that they have learned more about EWB and want to participate more, or simply because their friends are attending the conference. Therefore, their reason for attending the conference in the subsequent years of being a member, might be correlated with their unobserved characteristics[19].

In addition to the distance from the national conference, the size and the characteristics of the chapter that organized the conference in a particular year can have a direct impact on the quality and the size of the conference. A bigger chapter has more resources to fundraise or might have connections to better speakers. This could result in a higher quality conference, which in turn could indirectly impact the engagement and

[19] Furthermore, there is a small number of subsidies that each chapter can allocate to individuals who intend to attend the conference from chapters further from the conference location. These subsidies cover a portion of the travel costs and are distributed mostly on a first come first served manner. These subsidies effectively reduce the distance to the conference and could result in an under-estimation of the parameters of the model when distance is used as an instrument. Unfortunately, I do not have the list of members who received these subsidies in different years, but as mentioned, this information if available would only strengthen the results, since I am basically assuming that the distance is larger than it actually is.

participation levels of the attendees after the conference and in future years. Therefore, in addition to the distance from the national conference in the first year of joining EWB, I use dummies for the location of the national conference as instruments for the endogenous variable. The location dummy for a national conference city is equal to one if the individual attended the conference in his/her first year in a particular city, and zero otherwise[20].

To summarize, the key exclusion restriction assumption is that the distance from the national conference venue as well as dummies for the location of the national conference[21] in the first year of joining EWB are uncorrelated with own and friends' unobservables (i.e. the error term in the best response equation, η_i, and benefit equation, ϵ_i) and any other correlated effects and could be used as valid instruments for own effort level. More precisely, for a given exogenous variable z, such that $\mathbb{E}[\epsilon|\mathbf{z},\mathbf{G}]=0$, $\mathbb{E}[\eta|\mathbf{z},\mathbf{G}]=0$, this exclusion restriction implies that:

$$\text{cov}(\mathbf{Gz},\eta)=0, \qquad \text{cov}(\mathbf{G}^2\mathbf{z},\eta)=0 \tag{4.4}$$

$$\text{cov}(\mathbf{Gz},\epsilon)=0, \qquad \text{cov}(\mathbf{G}^2\mathbf{z},\epsilon)=0 \tag{4.5}$$

where $\mathbf{Gz}$ is the average of i's friends' z (i.e. the national conference distance or the location dummies in the first year), and $\mathbf{G}^2\mathbf{z}$ is the average of i's friends of friends' z. Therefore, I use **friends' distance and national conference location dummies** ($\mathbf{Gz}$), as well as **friends of friends' distance and national conference location dummies in their first year** ($\mathbf{G}^2\mathbf{z}$) as instruments for friends' effort[22]. The results from estimating the best response equation as well as the benefit equation for various online and offline networks are presented in the next section.

5. Results

In this section, I present the results of estimating the structural model. Initially, I discuss the results of estimating the best response equation for the three general effort variables for which there is no associated outcome variable (i.e. the general activities in table 3). Estimating the model using these three variables only recovers the reduced form parameters that are mostly talked about in the educational peer effect literature.

[20]The locations dummies for an agent may not add up to one. That is only the case when the individual attended the national conference in the first year of joining the organization.
[21]There are 6 national conference location dummies: (1) Toronto, (2) Montreal, (3) Vancouver, (4) St. Johns, (5) Calgary, (6) Ottawa.
[22]The exogenous instruments (i.e. distance and location dummies) are combined with the approach from Bramoulle at al (2009). This is because the exogeneity assumption of these instruments is consistent with their assumption of the absence of the correlated effects across individuals' covariates.

Then I present the results of estimating the full structural model using engagement (CHAMP) activities as well as the online peer to peer fundraising campaigns discussed in the data section.

5.1. **Estimates of Best Response Equation.** In this section the best response equation

$$e_i = \frac{\alpha}{\pi} x_i + \frac{\gamma}{\pi} \sum_{j=1}^{N} g_{ij} e_j + \frac{\eta_i}{\pi}$$

is estimated using three different effort variables: (1) total group membership, (2) total number of logins, and (3) total conferences that individuals have attended, and three different networks, namely, messages network, threaded comments network, group membership networks. These estimations rely on the exclusion restriction described in the previous section. In particular, I use subsets of the following instruments for the term $\sum_{j=1}^{N} g_{ij} e_j$: (1) friend's distance ($\mathbf{Gs}$), (2) friends' of friends distance ($\mathbf{G^2 s}$), (3) friends' of friends' friends' distance ($\mathbf{G^3 s}$) to the national conference in the first year of joining EWB, as well as (4) friend's ($\mathbf{Gl}$), (5) friends' of friends($\mathbf{G^2 l}$), (6) friends' of friends' friends' city ($\mathbf{G^3 l}$) of national conference in their first year in the organization. This model is over identified, enabling me to perform over identification test to verify the internal validity of instruments.

Table 7, reports the results of estimation of Equation (2.4) for the messages network. Column 1-3 are the OLS (without instruments), 2SLS and GMM results for the group membership effort variable, while column 4-6 and 7-9 present estimated coefficients for the total number of logins and total conferences effort variables respectively. For ease of comparison across three variables, the coefficient of peer effect is standardized. The results show a positive and significant relationship between the average of friends' effort and own effort in each category, controlling for observed individual characteristics, such as age, gender, language, etc as well as city fixed effects. The standard error term, indicated in parenthesis, are also clustered at the city level. Notice that all estimates are between 0.28 and 0.41 and are significant at the 0.001 level

The coefficients of the average of friends' effort are always positive, indicating that the game between individuals is one of strategic complementarities. In other words, a standard deviation in the average of an individuals' peers' effort level increases own effort levels by 0.44, 0.52 and 0.39 standard deviation for group membership, total number of logins and total conferences respectively, if his peer group is defined based on the direct messages he has sent or received from other members. The mean and standard deviation of these variables are given in table 8. For example, if one's friends join 40 new groups on average, he would join 15 new groups.

In all specifications in table 7, age is negatively and significantly correlated with own effort. Being male is only positive and significant in the OLS and IV estimations when the effort variable of interest is total number of logins. Additionally, the coefficients of being student are usually negative and significant expect when the variable of interest is total conference attendance. Whereas, both studying and working is positively and significantly correlated with own effort (exempt for total login variable). The reason behind this could that individuals who both study and work have been members of the organization for a long time, i.e. during university, and after graduation, hence, they are member of more groups and may have attended more conferences. In addition, the number of new members in the first year of joining EWB is positively and significantly correlated with own effort (except for the total conference variable). Lastly, note that the coefficient of the distance to the national conference in the first year of joining EWB is not significant after controlling for average of friends' effort and including city fixed effects.

Table 9 summarizes the estimation of the best response equation using messages, threaded comments and group membership networks. The estimated coefficients of the average friends' effort are larger, and still significant, for the group membership network for all effort variables. Remember, the definition of social ties in the group membership network is based on common offline activities, suggesting that friendships and connections based on activities that require in person interactions could have a larger impact on individuals choices and effort levels.

The first stage F-statistic and R^2 as well as the p-value of the J-test for all the specifications are reported at the bottom of each panel in table 9. In all estimations, I reject the null hypothesis of weak instruments. Additionally, since the p-value of the J-test from the optimal GMM estimator is always greater than 0.1, the null hypothesis that all instruments are valid is never rejected. Finally, since there are no specific outcome variables associated with these three effort variables, the benefit equation cannot be estimated. However, in the next section, using data from engagement (CHAMP) and campaign datasets, I estimate both equations simultaneously to recover all the parameters of the structural model.

5.2. **Estimating the effort choice and outcome equations simultaneously.** In this section, using the effort and outcome variables in two different categories of activities, i.e. engagement and fundraising activities, I estimate the full structural model.

$$e_i = \frac{\alpha}{\pi} x_i + \frac{\gamma}{\pi} \sum_{j=1}^{N} g_{ij} e_j + \frac{\eta_i}{\pi} \tag{5.1}$$

$$B_i(e_i, e_{-i}, g) = \pi e_i^2 + \lambda \sum_{j=1}^{N} g_{ij} e_j + \delta x_i + \epsilon_i \tag{5.2}$$

In addition to the variables previously stated (i.e. $\mathbf{Gs}$, $\mathbf{G}^2\mathbf{s}$, $\mathbf{G}^3\mathbf{s}$ and $\mathbf{Gl}$, $\mathbf{G}^2\mathbf{l}$, $\mathbf{G}^3\mathbf{l}$) to instrument for $\mathbf{Ge}$, I used elements of $x_i'x_i$ as instruments for e_i^2. Note that this relies on assumptions on the functional form of this structural model.

5.2.1. *Engagement Activities.* The CHAMP dataset contains information on off-line member engagement activities that individuals organize. For each of activity, the person(s) in charge of organizing the event has(have) collected total hours they spend organizing the event, as well as the outcome of that event. Since these activities range from member learning to functioning to fundraising events, I define different outcomes variables for each category[23]. For example, the outcome for a member learning activity is defined to be total number of participants, while the outcome of a fundraising event is total amount raised. A member may be engaged in several activities, hence, to get an aggregate measure of outcomes from several activities, I scale all outcome variables between 1 and 10, while retaining rank order and the relative size of separation between these values[24].

Table 10 presents the jointly estimated standardized coefficients[25] of the best response and the benefit equations for the three different networks. The first 3 columns show the OLS, 2SLS and GMM results for the best response equation where the coefficient on the average of peers' hours is $\hat{\bar{\gamma}}$ [26]. Additionally, columns (4)-(6) present the OLS, 2SLS and GMM estimated coefficients of the benefit equation 5.2. The GMM coefficients reported in column (3) and (6) are estimated simultaneously as a system, which increases the efficiency since this system imposes several cross-equation restrictions dues to their join dependence on the parameters of the model. As a result, the system GMM estimates in table 10 are usually more precisely estimated than the 2SLS coefficients. This is due to the fact that GMM makes use of the orthogonality

[23]These categories are: (1) Member learning, (2) Functioning, (3) Fundraising, (4) Curriculum enhancement (5) Publication (6) Public Engagement (7) School outreach (8) Workplace outreach.

[24]The scaling function that has been used is : $\frac{(b-a)(x-min(x))}{max(x)-min(x)} + a$, where $a = 1, b = 10$.

[25]i.e. the variables are standardized to have mean of 0 and standard devotion of 1.

[26]Recall $\bar{\gamma} = \frac{\gamma}{\pi}$.

conditions to allow for efficient estimation in the presence of heteroskedasticity of unknown form[27]. All specifications include individual characteristics such as gender, age, language, etc as controls as well as city fixed effects and clustered errors.

The estimated coefficients of $\bar{\gamma}$ in Equation (2.4) using 2SLS and GMM are significant and rather precisely estimated for the messages and threaded comments network, and since they are positive, once again we can infer the peer effect game is one of strategic complementarities: an increase in the total hours of volunteered by one's peers on average, increases his own time spent volunteering. For instant, a 10-hour increase in the average of peers' total hours volunteers corresponds to 2.5 hours of extra volunteering by an individual when his peer group is defined based on the threaded comments network. Interestingly however, the corresponding coefficient is not significant for the group membership network (the definition of links is based on membership in offline groups). In other words, there is no evidence that individuals' choice of action is strategically influenced by their peers in this network when we control for individual characteristics and city fixed effects. Only when the definition of a link is based on participation in different threaded conversations or the online messages that they send to each other, members seem to be moderately influenced by their peers. This could partly be due to the fact that individuals decide, in the threaded comments, what topics of conversations to join and share common interest, and might therefore be more similar, whereas membership in offline groups and participation in other in-person activities mostly depend on their chapter's overall agenda as apposed to their own interest[28]. Additionally, in the messages network, the links are based on the messages that members sent to each other, who might work together on several member engagement activities, and hence their choice of effort is influenced by how much their friends contribute.

In contrast, the estimated values of λ and π, i.e. the spillover effect in equation 5.2 and the coefficient of own effort squared (the marginal cost parameter) respectively, are always positive and significant, indicating that both the average of peers' effort and own effort has a positive direct spillover effect on own outcome. The GMM estimate of the direct spillover term is the largest for the threaded comments network and smallest for the group membership network. The larger positive spillover in the online networks

[27]Although the consistency of the IV coefficient estimates is not affected by the presence of heteroskedasticity, the standard IV estimates of the standard errors are inconsistent, preventing valid inference. For more detail discussion see: Baum, Christopher F., Mark E. Schaffer, and Steven Stillman. "Instrumental variables and GMM: Estimation and testing." Stata Journal 3.1 (2003): 1-31.

[28]In a very influential and extensive study by Katz and Lazarsfeld (1955), they found evidence that individuals were mostly influenced by others who were similar to themselves in terms of social status and other characteristics. Additionally, Meer (2010) provides evidence that peer pressure is stronger when friends share similar characteristics.

indicates, once more, that the relations defined based on online interactions among members have a higher effect on their outcome in addition to their own choices. In the threaded comments network, a standard deviation increase in the average of peers' effort causes 0.5 standard deviation increase in the outcome from a member engagement activity. Note that π is reassuringly positive and significant for all networks, indicating a positive marginal cost of exerting effort by individuals. Finally, note that the F-statistics from the first stage rejects the weak instruments null for both estimated $\bar{\gamma}$ and λ. The p-value of the system GMM also indicates that the over-identification restriction is valid.

5.2.2. *Online Fundraising Activities.* The second dataset that makes possible the estimation of the best response and the benefit equations simultaneously, is the online peer to peer fundraising campaign dataset. Note that the information on *online* activities differs from the engagement dataset that is based on *off-line* activities. This dataset contains information on the number of emails that individuals participating in these campaigns have sent to their friends and family, inviting them to contribute to their campaign finically and support EWB. It also contains information on the amount of donations they have received. Therefore, I construct standardized variable using the number of emails and total amount raised as proxies for the effort and the outcome variables respectively. As before, I use three different definitions of networks (individual specific peer group) in calculating the average of friends' effort levels. Additionally, individual characteristics such as gender, age, language, etc and city fixed effects are included in all of the following specifications. The errors are also clustered at the city level. The results are summarized in table 11.

The estimation results of the best response equation show positive and significant values for the $\bar{\gamma}$ parameter, indicating, once more, a game of strategic complementaries for all networks. The values of the coefficients ranges between 0.13 for the group membership network, to 0.2 for the threaded comments network. For example, in the threaded comments network, a member would send 10 more emails, if his peers on average send 40 more emails. As in the engagement activities, the coefficient of strategic complementarily is largest for the online networks. However, here the coefficient of strategic interaction is significant for the group membership network. This might be due to the fact that individuals may interact more frequently with others who participate in the same groups and activities, and seeing that their friends are putting more effort and raising more money, encourages them to exert more effort themselves. It could also be related to "social image" or "status", since no one would like to be the person who raised the least amount of money, therefore, seeing their friends exert more effort would encourage them to try harder.

In contrast to the engagement activities (table 10), except for the group membership network, there is no evidence of the direct spillover effect in online fundraising activities: for both of the online networks, the coefficient λ is insignificant for all specifications and sometimes negative. This is perhaps not too surprising: although all the proceeds from the donations goes to Engineers without Borders, each individual is fundraising for their own campaign, therefore, unlike the engagement activities in the CHAMP dataset, their efforts may not contribute to the outcome of their friends' campaigns. That said, since the definition of links and friendships in the group membership network is based on offline and in person interactions, and also since this network exhibits higher link density and transitivity (0.10 and 0.46 respectively) and includes bigger peers groups on average (average degree of 70), the more effort members exert in their own fundraising campaigns may raise more awareness and hype around this massive fundraising event in their community, and therefore positively contribute to their peers's campaigns.

Also, note that the first stage F-statistics and R^2 are listed at the bottom of each panel in Table 11, and reject the weak instruments null for both the average of friends' efforts, $\sum g_{ij}e_j$ and e_i^2. The p-value of the J-test also validates the over identification restrictions.

5.3. **Summary.** As discussed in section 2, using the estimated π, the value of the parameter γ is recovered from the estimated $\bar{\gamma}$. These values, combined with estimated λ, are reported in table 12 for both the engagement and online fundraising activities.

In the engagement activities, we see that the two online networks exhibits medium levels of strategic peer effects ($\bar{\gamma} \approx 0.2$). For instant, in the messages network, when an individual's peers, on average, volunteer 10 more hours, he would spend roughly 3.25 hours more volunteering. In addition, an hour increase in the average of peers' effort, increases the marginal payoff directly by $\gamma = 0.20$ units. Furthermore, we always see a significant and positive direct spillover effects, indicating that a standard deviation increase in volunteer activity by the average of an individual's friends, increases the outcome by 0.2 to 0.5 standard deviation.

In the fundraising activities, the online networks exhibit larger coefficients of strategic complementarities as in the engagement activities with similar values. However, as discussed earlier, only the group membership network exhibits positive spillovers. In the group membership network, if an individual's peers on average send 100 more emails, he would send 45 extra emails (*strategic complementarities*); additionally, these 100 extra emails sent by his peers on average, contributes 556$ directly to his fundraising outcome (*direct spillover*).

6. Robustness Checks

In this section, I present three different robustness checks: (1) based on alternative definition of networks, and (2) using average of friends' of friends' covariates as instruments for average friends' efforts suggested by Bramoulle et al (2009) (3) using random networks among members.

6.1. Alternative Group Membership Networks.

As first set of robustness checks, I use two alternative definitions for the group membership network. Initially, I relax the assumption about the size of the groups that individuals are members of from being smaller than 50 to smaller than 100. This means that any two individuals who are members of at least one common group of size 100 or smaller are defined to be connected in this network. Second, I use a more stringent requirement for being connected in the group membership network: for two individuals to be considered linked, they have to be members of at least 3 common groups of size 50 or less (in the default group membership network, individuals are linked in they at least one group of size 50 or smaller in common). The results of estimating the full model, using both the CHAMP and Campaign dataset with the two new networks are presented in tables 15 and 16. For ease of referring to these alternative network definitions, from now on, "G100" refers to the group membership < 100 network and "G50-3" refers to the group membership < 50 network with at least 3 groups in common.

In engagement activities (CHAMP), the estimated coefficients using both G100 and G50-3 networks present no evidence of strategic complementarities and very little evidence for the presence of the direct spillover effect. This is similar to the baseline group membership results that showed no sign strategic complementarities, and had the smallest coefficients of direct spillover compared to the other two online networks. In the G50-3 network, although insignificant, the coefficient of the interaction term is actually negative, signalling that there might be some substitution among individuals who interact very regularly. Additionally, one reason for the insignificant estimated λ using G50-3 might be because of the smaller average peer group size (average degree is 21 and 75 for the G50-3 and the default network respectively).

In G100 network, the estimated coefficients of direct spillover and strategic network effects are both insignificant. This might be due to several factors, firstly, a member's peer group is defined very loosely in this network, and secondly, the impact from these activities is mostly realized in the long run, therefore, there is no reason to expect that individuals are influenced strategically by members with whom they hardly interact. Also these individuals might be involved in different activities that has very little in common with what they do.

The results of a similar estimation using peer to peer fundraising campaigns is presented in table 16. As in the default group membership network, both the coefficients of strategic complementarity and direct spillover in G100 are significant. Interestingly, in G100, the coefficient of strategic complementarity is greater than the baseline group membership network, supporting the idea that competition to raise more money and "social image and status" among acquaintances could encourage members to put more efforts if they observe others are trying harder[29]. In the G50-3 network, the coefficient of strategic complementarity is also positive and significant and slightly larger than the baseline network. However, there is no evidence for the direct spillover effect.

The characteristics of G100 and G50-3 are listed in table 6 for both datasets.

Table 6. Group Membership Networks Characteristics

	Netwok Variables:				
Network	Avg Degree	Diameter	Avg Path Length	Density	Transitivity
			Champ Dataset		
Default G	75.41	4	2.04	0.10	0.41
G100	252.81	4	1.68	0.33	0.63
G50-3	21.79	5	2.72	0.03	0.43
			Campaign Dataset		
Default G	69.93	4	2.08	0.10	0.46
G100	188.87	4	1.79	0.26	0.41
G50-3	15.93	7	2.90	0.02	0.37

6.2. **Robustness Checks Using Friends' of Friends' Covariates as IV.**

As discussed in the empirical strategy section, Bramoulle et al (2009) present a novel approach to address the reflection problem, specially in the absent of the correlated effects: under certain conditions on the structure of the network, they propose to use the covariates (e.g. age, gender, language, etc) of friend's of friends who are not directly friend with an individual as instruments for his friend's efforts. Since the assumption of no correlated effect is a strong one, these instruments have not been used in the main specifications of this paper, allowing for potential unobserved correlations among individuals' characteristics. However, in this section I use these instruments[30] as a robustness check, *assuming the absence of the correlated effects.* The results for the engagement and fundraising activities are presented in tables 17 and 18 respectively.

[29]This is in contrast to the engagement activities where members are involved in several different activities. In fundraising campaigns, all individuals are raising money. Therefore, competition may play a stronger role here than the engagement activities, specially that individuals who raise more money are recognized by the national office of EWB as champions.

[30]Namely, friends' of friends' and friends' of friends' of friends' average gender, age, language, number of new members in their chapter when they joined and student or work dummies are used as instrument of average of friends' efforts. Note that distance to national conference and location dummies for national conference in the first year are not used here.

In the engagement activities (CHAMP), the estimated coefficients of all the parameters of the model are positive and significant using both IV (except $\bar{\gamma}$ in the messages network) and system GMM. Additionally, both $\bar{\gamma}$ and λ estimated here using these new instruments (that potentially violate the correlated effects assumption) are slightly larger than using the exogenous instruments of distance to and the location of the national conference in the first year of joining EWB. This indicates that these instruments potentially over estimate the parameters of the model, because of the correlation between an individual's own characteristics (that contributes to his choice of effort) and his friends' friends and their friends' covariates. Said differently, it would be hard to argue that the unobserved characteristics of an individuals' friends of friends who are not directly his friends are orthogonal to his, hence, violating the assumption of the exogeneity of these instruments. Note that even the coefficient of strategic complementarities is positive and significant for the group membership network.

Similarly, the predicted coefficients using fundraising campaigns are generally larger here than those in the baseline case, and the system GMM estimator predicts significant values of strategic complementary for all of the networks as well as positive significant coefficient of direct spillover for the threaded comments and group membership networks. This could again be due the fact that using friends of friends' covariates, which in principle could be correlated with own unobservables, as instruments for friends' efforts is picking out the effect of some unobserved correlated factors through friends of friends' covariates as apposed to the effect of the average of friends' efforts on own outcome. Nonetheless, it is reassuring that the predicted coefficient using these instruments have the same sign and relatively similar (larger) values in both the Champ and the Campaign dataset.

7. Conclusion

I present a structural model of social interactions in which individuals' choices depend on their own characteristics as well as their peers' actions both directly and through an interaction term. Using a unique dataset, which I construct from raw data collected by Engineers Without Borders on their membership activities, I provide precise predictions for the coefficient of the model by simultaneously estimating the best response and the equilibrium level benefit equation. To do this, I exploit exogenous variation in the network structure as well as the distance of new members to the national conference of EWB, to instrument for endogenous variables. The network structure was defined using detailed data on members interactions with each other through several online and offline channels, and enabled me to contrast the impact of different networks on individuals choices and outcomes. In addition, the dataset contains information on both time intensive activities as well as peer to peer fundraising campaigns.

The results reveal several important aspects of social interactions and network effects. In member engagement activities, strategic complementarities are only present when the peer group is defined based on individuals (potentially) sharing common interests. A direct and positively significant spillover effect is always present regardless of the definition of networks in these activities. In contrast, in the peer to peer online fundraising campaigns, the coefficient of strategic interactions is always significant and positive, indicating the presence of strategic complementarities, while there is only evidence for direct spillover effects in larger networks when the peer group includes an individual's acquaintances as well as his close friends. Observing large coefficients of strategic complementarity in peer to peer fundraising campaigns points to the fact that an individual is influenced nontrivially by his peer group, which could be due to social image or social recognition. Finally, in no specification I find evidence for strategic substitution among agent's actions. This is in contrast to public good models that exhibit the free-riding problem.

In conclusion, the contributions of this paper are threefold: I (1) construct a unique dataset with both offline and online social network interactions among individuals, (2) separately recover the impact of the two network effects – the strategic interaction and the direct spillover term in the structural model by using two (instead of one) equations due to the richness of the data that includes proxies for outcome variable as well as effort variable, (3) the identification strategy is based on both exogenous instruments and the network structure. This methodology emphasizes the need for a structural model to correctly interpret the estimates within a strategic model of interactions with agent specific peer groups and highlights the importance of the network effects. It also suggests that individuals may not internalize the positive externalities and full network effects when playing best response, leading to under-provision of effort and thus public good. Therefore, more active investment in understanding the networks among members of an organization and their incentives could potentially lead to improvements and efficiency gains.

References

[1] Acemoglu et al, State capacity and economic development: A network approach. American Economic Review 2015, 105(8): 2364-2409

[2] Andreoni, James. Giving with impure altruism: applications to charity and Ricardian equivalence. The Journal of Political Economy (1989): 1447-1458.

[3] Andreoni, James. Impure altruism and donations to public goods: a theory of warm-glow giving. The economic journal (1990): 464-477.

[4] Andreoni, James. Toward a theory of charitable fund-raising. Journal of Political Economy 106.6 (1998): 1186-1213.

[5] Andreoni, James, 2006. Philanthropy. In: Kolm, S.-C., Mercier Ythier, J. (Eds.), Handbook of Giving, Reciprocity and Altruism. North Holland, Amsterdam.

[6] Ballester, Coralio and Antoni Calvo-Armengol (2010). Interactions with Hidden Complementarities, Regional Science and Urban Economics 40, 397-406.

[7] Ballester, Coralio, Antoni Calvo-Armengol, and Yves Zenou (2006). Who's who in Networks: Wanted: The Key Player, Econometrica, 74(5), 1403-1417.

[8] Banerjee, Abhijit, et al. The diffusion of microfinance. Science 341.6144 (2013): 1236498.

[9] Baum, Christopher F., Mark E. Schaffer, and Steven Stillman. Instrumental variables and GMM: Estimation and testing. Stata journal 3.1 (2003): 1-31.

[10] Benabou, Roland, and Jean Tirole. Incentives and prosocial behavior. No. w11535. National Bureau of Economic Research, 2005.

[11] Blume, Lawrence E., et al. Identification of social interactions. Handbook of Social Economics. Available at SSRN 1660002 (2010).

[12] Boucher, Vincent, and Bernard Fortin. Some Challenges in the Empirics of the Effects of Networks. (2015): 15-04.

[13] Bramoulle, Yann, Rachel Kranton, and Martin D'amours. Strategic interaction and networks. The American Economic Review 104.3 (2014): 898-930.

[14] Bramoulle, Yann, Habiba Djebbari, and Bernard Fortin. Identification of peer effects through social networks. Journal of econometrics 150.1 (2009): 41-55.

[15] Brock, W., Durlauf, S., 2001a. Discrete choice with social interactions. Review of Economic Studies 68 (2), 235-260.

[16] Brock, W., Durlauf, S., 2001b. Interaction-based models. In: Heckman, J., Leamer, E. (Eds.), Handbook of Econometrics, vol. 5. North-Holland, Amsterdam.

[17] Brock, W., Durlauf, S., 2003. Multinomial choice with social interactions, SSRI Working Paper No. 2003-1, University of Wisconsin.

[18] Burke, Mary A., and Tim R. Sass. Classroom peer effects and student achievement. Journal of Labor Economics 31.1 (2013): 51-82

[19] Calvo-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou (2009). Peer Effects and Social Networks in Education, Review of Economic Studies, 76, 1239-1267.

[20] Cameron, Adrian Colin, and Pravin K. Trivedi. Microeconometrics using stata. (2010): 166.

[21] Card, David, and Laura Giuliano. Peer effects and multiple equilibria in the risky behavior of friends. Review of Economics and Statistics 95.4 (2013): 1130-1149.

[22] Conley, T., Udry, C., 2005. Learning about a new technology: Pineapple in Ghana, Mimeo, Yale University.

[23] Corbo, Jacomo, Antoni Calvo-Armengol, and David C. Parkes. The importance of network topology in local contribution games. Internet and Network Economics. Springer Berlin Heidelberg, 2007. 388-395.

[24] Dastranj, R. Karaivanov, A. Easton, S. Optimal Crime Networks – Theory and Lessons for Policy, Working paper (2015)

[25] De Georgi, G., Pellizzari, M., Redaelli, S., 2007. Be Careful of the Books you Read as of the Company you Keep:Evidence on Peer Effetcs in Educational Choices, IZA DP No 2833.

[26] DellaVigna, Stefano, John A. List, and Ulrike Malmendier. Testing for altruism and social pressure in charitable giving. No. w15629. National Bureau of Economic Research, 2009.

[27] De Paula, Áureo. Econometrics of network models. Econometric Society World Congress in Montreal, Canada, 2015.

[28] Duflo, Esther, and Emmanuel Saez. The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. No. w8885. National Bureau of Economic Research, 2002.

[29] Edwards, James T., and John A. List. Toward an understanding of why suggestions work in charitable fundraising: Theory and evidence from a natural field experiment. Journal of Public Economics 114 (2014): 1-13.

[30] Erdös, P. and A. Renyi (1959) On random graphs, Publ. Math. Debrecen, 6:290-297.

[31] Erdös, P. and A. Renyi (1960) On the Evolution of Random Graphs, Publication of the Mathematical Institute of the Hungarian Academy of Sciences, 5, 17-61.

[32] Erdös, P. and A. Renyi (1961) On the strength of connectedness of a random graph, Acta Math. Acad. Sci. Hungar., 12:261-267.

[33] Fletcher, Jason M. Peer influences on adolescent alcohol consumption: evidence using an instrumental variables/fixed effect approach. Journal of Population Economics 25.4 (2012): 1265-1286.

[34] Friesen, Jane, and Brian Krauth. Sorting and inequality in Canadian schools. Journal of public Economics 91.11 (2007): 2185-2212.

[35] Gaviria, A., Raphael, S., 2001. School based peer effects and juvenile behavior. Review of Economics and Statistics 83 (2), 257-268.

[36] Goldsmith-Pinkham, Paul, and Guido W. Imbens. Social networks and the identification of peer effects. Journal of Business & Economic Statistics 31.3 (2013): 253-264.

[37] Graham, B.S., 2008. Identifying social interactions through conditional variance restrictions. Econometrica 76 (3), 643-660.

[38] Graham, B.S., Hahn, J., 2005. Identification and estimation of the linear-in-means model of social interactions. Economics Letters 88 (1), 1-6.

[39] Glaeser, E., Sacerdote, B., Scheinkman, J., 1996. Crime and social interaction. Quarterly Journal of Economics 111 (2), 507-548.

[40] Glazer, Amihai, and Kai A. Konrad. A signaling explanation for charity. The American Economic Review (1996): 1019-1028.

[41] Hanushek, Eric A., et al. Does peer ability affect student achievement?. Journal of applied econometrics 18.5 (2003): 527-544.

[42] Hlavac, Marek (2014). stargazer: LaTeX code and ASCII text for well-formatted regression and summary statistics tables.

[43] Hoxby, Caroline. Peer effects in the classroom: Learning from gender and race variation. No. w7867. National Bureau of Economic Research, 2000.

[44] Ioannides, Y., Soetevent, A., 2007. Social networking and individual outcomes beyond the mean field case. Journal of Economic Behavior and Organization.

[45] Ioannides, Y., Zabel, A., 2003. Neighborhood effects and housing demand. Journal of Applied Econometrics 18 (5), 563-584.

[46] Jackson, Matthew O. Social and economic networks. Princeton University Press, 2010.

[47] Jackson, Matthew O., and Asher Wolinsky. A strategic model of social and economic networks. Journal of economic theory 71.1 (1996): 44-74.

[48] Karlan, Dean, and John A. List. Does price matter in charitable giving? Evidence from a large-scale natural field experiment. No. w12338. National Bureau of Economic Research, 2006.

[49] Kolaczyk, Eric D., and Gábor Csárdi. Statistical analysis of network data with R. Vol. 65. Springer, 2014.

[50] Krauth, B., 2006. Simulation-based estimation of peer effects. Journal of Econometrics 133 (1), 243-271.

[51] Lee, L.F., 2003. Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. Econometric Reviews 22 (4), 307-335.

[52] Lee, L.F., 2007. Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. Journal of Econometrics 140 (2), 333-374.

[53] Lilley, Andrew, and Robert Slonim. The price of warm glow. Journal of Public Economics 114 (2014): 58-74.

[54] Lin, X., 2007. Peer effects and student academic achievement: An application of spatial autoregressive model with group unobservables, Mimeo, Department of Economics, Tsinghyua University, Beijing.

[55] Manski, C., 1993. Identification of endogenous social effects: The reflection problem. Review of Economic Studies 60 (3), 531-542.

[56] Mayo, John W., and Catherine H. Tinsley. Warm glow and charitable giving: Why the wealthy do not give more to charity?. Journal of Economic Psychology 30.3 (2009): 490-499.

[57] Manski, C., 1995. Identification Problems in the Social Sciences. Harvard Univ. Press, Cambridge.

[58] Meer, Jonathan. Brother, can you spare a dime? Peer pressure in charitable solicitation. Journal of Public Economics 95.7 (2011): 926-941.

[59] Moffitt, R., 2001. Policy interventions low-level equilibria, and social interactions. In: Durlauf, Steven, Young, Peyton (Eds.), Social Dynamics. MIT Press.

[60] Sacerdote, B., 2001. Peer effects with random assignment: Results for darmouth roommates. Quarterly Journal of Economics 116 (2), 681-704.

[61] Saez, E., Duflo, E., 2003. The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. Quarterly Journal of Economics 118, 2003, 815-842.

[62] Soetevent, Adriaan R. Empirics of the Identification of Social Interactions; An Evaluation of the Approaches and Their Results. Journal of Economic Surveys 20.2 (2006): 193-228.

[63] Smith, Adam. The theory ofmoral sentiments. New York: Garland 1971 (1759).

[64] Topa, Giorgio (2001). Social Interactions, Local Spillovers and Unemployment, Review of Economic Studies, 68(2), 261-295.

[65] Vesterlund, Lise, 2006. Why do people give?, In: Steinberg, Richard, Powell, Walter W. (Eds.), The Nonprofit Sector, 2nd edition. Yale Press.

TABLE 7. Best Response Equation - Summary Results for Messages Network with City FE and Clustered Errors

	Dependent variable:								
	group	group	group	login	login	login	conference	conference	conference
	OLS	*2SLS*	*GMM*	*OLS*	*2SLS*	*GMM*	*OLS*	*2SLS*	*GMM*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Messages Network								
average peer groups	0.278^{**} (0.087)	0.370^{***} (0.076)	0.374^{***} (0.077)						
average peer logins				0.304 (0.163)	0.425^{***} (0.116)	0.356^{***} (0.055)			
average peer conference							0.283^{***} (0.074)	0.339^{***} (0.083)	0.341^{***} (0.086)
gender	1.3526 (2.3728)	1.0190 (2.2741)	0.6309 (2.1643)	69.5379^{**} (20.2030)	64.6047^{***} (18.8504)	55.9171^{***} (15.7743)	0.1383 (0.0925)	0.1317 (0.0882)	0.0783 (0.0851)
language	-1.5184 (3.9353)	-1.3805 (3.8331)	-3.4475 (4.4019)	14.5092 (44.5619)	18.5537 (46.8138)	27.9878 (46.8702)	-0.6027^{**} (0.1721)	-0.5966^{***} (0.1773)	-0.5954^{***} (0.1714)
age	-0.7967^{**} (0.2180)	-0.7129^{***} (0.2054)	-0.7313^{***} (0.2006)	-6.1017^{***} (1.3682)	-5.1696^{***} (1.3951)	-4.8595^{***} (1.3849)	-0.0428^{***} (0.0089)	-0.0408^{***} (0.0098)	-0.0402^{***} (0.0104)
chapter size	0.1451^{***} (0.0237)	0.1391^{***} (0.0211)	0.1403^{***} (0.0219)	0.9432^{***} (0.2218)	0.8630^{***} (0.1912)	0.7716^{**} (0.2465)	0.0033 (0.0019)	0.0032 (0.0018)	0.0033 (0.0019)
student	-7.6090^{*} (3.0828)	-8.0092^{**} (2.9886)	-7.3362^{*} (2.9904)	-110.8734^{**} (37.0657)	-113.0946^{***} (34.2232)	-94.1481^{***} (27.1390)	-0.1604 (0.1514)	-0.1713 (0.1476)	-0.1293 (0.1400)
work	-5.3140 (3.9118)	-4.3788 (3.9828)	-3.3197 (4.0449)	-76.9789 (39.4569)	-68.1902 (40.3486)	-45.1677 (37.9549)	-0.3394 (0.2530)	-0.3136 (0.2376)	-0.2917 (0.2386)
both	14.0103^{**} (3.9186)	13.2499^{***} (3.9708)	13.6362^{***} (3.9583)	70.2281 (64.5667)	55.8167 (60.5180)	69.3003 (52.0661)	1.0237^{***} (0.2251)	0.9973^{***} (0.2261)	1.0427^{***} (0.2202)
distance	-1.1777 (1.3131)	-1.0418 (1.1801)	-0.8679 (1.1349)	-13.0987 (11.2585)	-11.3689 (9.9530)	-8.6676 (9.7639)	-0.0070 (0.0706)	-0.0020 (0.0644)	-0.0074 (0.0620)
N	1071	1071	1071	1071	1071	1071	1071	1071	1071

Standard errors in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
The coefficients of peer effect are reported as Standardized coefficients for ease of comparison across categories.

Table 8. Mean and Standard Deviation for the General Effort Variables

	Model Variable:					
	total group membership	average peer groups	total logins	average peer logins	total conferences	average peer conferences
‘	*Messages Network*					
Mean	30.31	50.88	117.94	286.16	1.55	2.48
Std. Dev.	34.35	39.88	331.62	462.81	1.84	2.01
	Threaded Comments Network					
Mean	36.60	67.47	144.72	488.15	1.75	3.23
Std. Dev.	33.28	27.44	301.65	393.19	1.81	1.27
	Group Membership Network					
Mean	20.55	37.31	63.63	166.17	0.89	1.67
Std. Dev.	25.80	12.72	249.62	132.04	1.41	0.72

TABLE 9. Best Response Equation - Summary Results for all Network with City FE and Clustered Errors with Standardized Coefficients

	Dependent variable:								
	group	group	group	login	login	login	conference	conference	conference
	OLS	*2SLS*	*GMM*	*OLS*	*2SLS*	*GMM*	*OLS*	*2SLS*	*GMM*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Messages Network								
average peer groups	0.278^{**} (0.087)	0.370^{***} (0.076)	0.374^{***} (0.077)						
average peer logins				0.304 (0.163)	0.425^{***} (0.116)	0.356^{***} (0.055)			
average peer conference							0.283^{***} (0.074)	0.339^{***} (0.083)	0.341^{***} (0.086)
Obs	1,071	1,071	1,071	1,071	1,071	1,071	1,071	1,071	1,071
FS F-stat		27.41^{***}			11.10^{***}			63.74^{***}	
FS R^2		0.37			0.14			0.48	
J test			8.91			3.99			9.26
p-value			0.25			0.78			0.23
	Threaded Comments Network								
average peer groups	0.119^{**} (0.039)	0.226^{***} (0.063)	0.217^{***} (0.065)						
average peer logins				0.021 (0.038)	0.201^{***} (0.054)	0.220^{***} (0.055)			
average peer conference							0.196^{***} (0.032)	0.216^{***} (0.050)	0.199^{***} (0.062)
Obs	1,427	1,427	1,427	1,427	1,427	1,427	1,427	1,427	1,427
FS F-stat		20.60^{***}			18.41^{***}			22.67^{***}	
FS R^2		0.14			0.12			0.26	
J test			1.29			1.27			1.37
p-value			0.26			0.26			0.24
	Group Membership Network								
average peer groups	0.259^{***} (0.036)	0.426^{***} (0.060)	0.442^{***} (0.046)						
average peer logins				0.429^{*} (0.203)	0.376^{*} (0.150)	0.403^{***} (0.110)			
average peer conference							0.435^{***} (0.020)	0.511^{***} (0.033)	0.553^{***} (0.030)
Obs	3461	3461	3461	3461	3461	3461	3461	3461	3461
FS F-stat		31.62^{***}			12.76^{***}			81.67^{***}	
FS R^2		0.42			0.30			0.60	
J test			0.19			0.18			0.10
p-value			0.66			0.66			0.75
Note:	*p<0.05; **p<0.01; ***p<0.001								

TABLE 10. Estimation of Structural Model With Engagement Data (CHAMP), with Standardized Coefficients

	Dependent variable:					
	total effort			outcome		
	OLS	*2SLS*	*sys GMM*	*OLS*	*2SLS*	*sys GMM*
	(1)	(2)	(3)	(4)	(5)	(6)
	Messages Network					
total hours squared				$\pi = 0.315^{*}$	0.565	0.543^{***}
				(0.112)	(0.303)	(0.140)
average peer hours	$\bar{\gamma} = 0.112$	0.233^{*}	0.239^{**}	$\lambda = 0.195^{*}$	0.373^{***}	0.325^{***}
	(0.079)	(0.098)	(0.074)	(0.080)	(0.094)	(0.071)
Observations	361	361	361	361	361	361
J test						41.69
p-value						0.014
	First Stage Statistics for total hours squared					
First Stage F Stat					42.950^{**}	
First Stage R^2					0.081	
	First Stage Statistics for average peer hours					
First Stage F Stat		31.236^{***}			46.010^{***}	
First Stage R^2		0.415			0.475	
	Threaded Comments Network					
total hours squared				$\pi = 0.303^{*}$	0.385	0.301^{*}
				(0.115)	(0.221)	(0.118)
average peer hours	$\bar{\gamma} = 0.073$	0.190^{**}	0.188^{***}	$\lambda = 0.225^{***}$	0.508^{***}	0.499^{***}
	(0.038)	(0.060)	(0.037)	(0.039)	(0.091)	(0.083)
Observations	405	405	405	405	405	405
J test						34.86
p-value						0.28
	First Stage Statistics for total hours squared					
First Stage F Stat					11.005^{***}	
First Stage R^2					0.067	
	First Stage Statistics for average peer hours					
First Stage F Stat		95.45^{***}			35.588^{***}	
First Stage R^2		0.288			0.226	
	Group Membership Network					
total hours squared				$\pi = 0.325^{*}$	0.609	0.637^{***}
				(0.120)	(0.325)	(0.157)
average peer hours	$\bar{\gamma} = 0.020$	-0.026	0.022	$\lambda = 0.258^{***}$	0.299^{*}	0.209^{***}
	(0.093)	(0.098)	(0.047)	(0.068)	(0.130)	(0.059)
Observations	499	499	499	499	499	499
J test						17.467
p-value						0.356
	First Stage Statistics for total hours squared					
First Stage F Stat					40.252^{***}	
First Stage R^2					0.075	
	First Stage Statistics for average peer hours					
First Stage F Stat		42.352^{***}			36.475^{***}	
First Stage R^2		0.311			0.313	

Note: *p<0.05; **p<0.01; ***p<0.001

TABLE 11. Estimation of Structural Model With Campaign Data, With Standardized Coefficients

	Dependent variable:					
	total emails			Total Amount Raised		
	OLS	*2SLS*	*sys GMM*	*OLS*	*2SLS*	*sys GMM*
	(1)	(2)	(3)	(4)	(5)	(6)
	Messages Network					
total emails squared				$\pi = 0.670^{***}$	0.570^{***}	0.658^{***}
				(0.050)	(0.154)	(0.139)
average peer emails	$\bar{\gamma} = 0.092$	0.203^{*}	0.170^{*}	$\lambda = -0.048$	0.040	0.149
	(0.115)	(0.082)	(0.079)	(0.075)	(0.179)	(0.078)
Observations	360	360	360	360	360	360
J test						36.05
p-value						0.141
	First Stage Statistics for total emails squared					
First Stage F Stat					182.946^{***}	
First Stage R^2					0.125	
	First Stage Statistics for average peer emails					
First Stage F Stat		11.606^{***}			298.97^{***}	
First Stage R^2		0.153			0.174	
	Threaded Comments Network					
total emails squared				$\pi = 0.719^{***}$	0.520^{***}	0.437^{***}
				(0.040)	(0.183)	(0.108)
average peer emails	$\bar{\gamma} = 0.195^{***}$	0.229^{***}	0.197^{**}	$\lambda = -0.043$	−0.034	0.084
	(0.029)	(0.094)	(0.057)	(0.074)	(0.120)	(0.079)
Observations	334	334	334	334	334	334
J test						37.67
p-value						0.225
	First Stage Statistics for total emails squared					
First Stage F Stat					299.55^{***}	
First Stage R^2					0.085	
	First Stage Statistics for average peer emails					
First Stage F Stat		57.78^{***}			225.28^{***}	
First Stage R^2		0.362			0.330	
	Group Membership Network					
total emails squared				$\pi = 0.674^{***}$	0.643^{***}	0.533^{***}
				(0.049)	(0.180)	(0.086)
average peer emails	$\bar{\gamma} = 0.135^{*}$	0.205^{*}	0.126^{*}	$\lambda = 0.178^{**}$	0.218^{**}	0.193^{**}
	(0.057)	(0.086)	(0.048)	(0.063)	(0.069)	(0.059)
Observations	426	426	426	426	426	426
J test						43.96
p-value						0.061
	First Stage Statistics for total emails squared					
First Stage F Stat					56.427^{***}	
First Stage R^2					0.085	
	First Stage Statistics for average peer emails					
First Stage F Stat		53.941^{***}			4491.49^{***}	
First Stage R^2		0.492			0.489	
Note:						*p<0.05; **p<0.01; ***p<0.001

TABLE 12. Estimated Parameters of the Structural Model, with Standardized Coefficients

Networks	$\bar{\gamma}$	γ Strategic Complementarities	λ Direct Spillover	π MC of Effort
	Engagement Activities:			
Messages Network	0.239**	0.130**	0.325***	0.543**
	(0.074)	(0.049)	(0.071)	(0.140)
Threaded Comments Network	0.188***	0.056*	0.499***	0.301*
	(0.037)	(0.027)	(0.083)	(0.118)
Group Membership Network	0.022	0.014	0.209***	0.637***
	(0.047)	(0.032)	(0.059)	(0.157)
	Fundraising Activities:			
Messages Network	0.170*	0.112	0.149	0.658***
	(0.079)	(0.058)	(0.078)	(0.139)
Threaded Comments Network	0.197**	0.086**	0.084	0.437***
	(0.074)	(0.028)	(0.079)	(0.108)
Group Membership Network	0.126*	0.067**	0.193**	0.533***
	(0.048)	(0.021)	(0.059)	(0.086)

Note: System GMM estimated coefficients are used in this table.

TABLE 13. Mean and Standard Deviation for the model variables in Engagement data

	Outcome	total hours	total hours^2	peers' total hours
	Messages Network			
Mean	5.295	0.378	0.718	0.472
Std. Dev.	5.307	0.759	4.555	0.557
	Threaded Comments Network			
Mean	5.272	0.365	0.662	0.572
Std. Dev.	5.281	0.728	4.259	0.289
	Group Membership Network			
Mean	4.797	0.327	0.600	0.443
Std. Dev.	5.017	0.702	3.992	0.157

Note:Total hours is measured in 1000 hours.

TABLE 14. Mean and Standard Deviation for the model variables in Campaign data

	Model Variable:			
	total amount	total emails	total email^2	peers' total emails
	Messages Network			
Mean	132.18	114.09	118932.2	170.59
Std. Dev.	2593.13	325.73	1022502	222.16
	Threaded Comments Network			
Mean	1421.01	127.53	136420.6	234.96
Std. Dev.	2635.34	346.98	1094850	265.4
	Group Membership Network			
Mean	1237.71	104.65	101541.1	160.19
Std. Dev.	2384.88	301.19	924206.7	85.72

TABLE 15. Robustness Checks with Group Membership Networks, CHAMP data with Standardized Coefficients

	Dependent variable:					
	total effort			outcome		
	OLS	*2SLS*	*sys GMM*	*OLS*	*2SLS*	*sys GMM*
	(1)	(2)	(3)	(4)	(5)	(6)
	Group Membership < 100 Network					
total hours squared				0.307*	0.484	0.377**
				(0.117)	(0.290)	(0.114)
average peer hours	0.028	0.086	0.010	0.171**	0.193*	0.053
	(0.065)	(0.049)	(0.037)	(0.059)	(0.092)	(0.066)
	First Stage Statistics for total hours squared					
First Stage F Stat					140.0***	
First Stage R^2					0.062	
	First Stage Statistics for average peer hours					
First Stage F Stat		36.40***			856.58***	
First Stage R^2		0.315			0.332	
J test						38.025
p-value						0.050
Observations	507	507	507	507	507	507
	Group Membership <50 Network, Common Groups $\geq$ 3					
total hours squared				0.305*	0.492	0.581***
				(0.116)	(0.283)	(0.153)
average peer hours	0.011	−0.027	−0.006	0.096*	0.161	0.148
	(0.039)	(0.102)	(0.061)	(0.043)	(0.114)	(0.085)
	First Stage Statistics for total hours squared					
First Stage F Stat					34.063***	
First Stage R^2					0.072	
	First Stage Statistics for average peer hours					
First Stage F Stat		45.145***			35.998***	
First Stage R^2		0.193			0.176	
J test						44.48
p-value						0.06
Observations	472	472	472	472	472	472

Note: *p<0.05; **p<0.01; ***p<0.001

Table 16. Robustness Checks with Group Membership Networks, Campaign Data with Standardized Coefficients

	Dependent variable:					
	total effort			outcome		
	OLS	*2SLS*	*sys GMM*	*OLS*	*2SLS*	*sys GMM*
	(1)	(2)	(3)	(4)	(5)	(6)
	Group Membership < 100 Network					
total email squared				0.652***	0.626***	0.684***
				(0.054)	(0.175)	(0.118)
average peer email	0.152	0.218	0.229*	0.125	0.115	0.117*
	(0.128)	(0.164)	(0.100)	(0.087)	(0.097)	(0.045)
	First Stage Statistics for total email squared					
First Stage F Stat					167.92	
First Stage R^2					0.077	
	First Stage Statistics for average peer emails					
First Stage F Stat		207.60***			959.96***	
First Stage R^2		0.513			0.446	
J test						48.94
p-value						0.03
Observations	437	437	437	437	437	437
	Group Membership <50 Network, Common Groups ≥ 3					
total email squared				0.691***	0.532**	0.498***
				(0.047)	(0.218)	(0.137)
average peer email	0.192	0.442	0.186*	0.050	0.114	0.048
	(0.153)	(0.236)	(0.094)	(0.038)	(0.101)	(0.080)
	First Stage Statistics for total email squared					
First Stage F Stat					12.770	
First Stage R^2					0.115	
	First Stage Statistics for average peer emails					
First Stage F Stat		58.73***			183.35***	
First Stage R^2		0.347			0.379	
J test						35.28
p-value						0.08
Observations	381	381	381	381	381	381

Note: $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

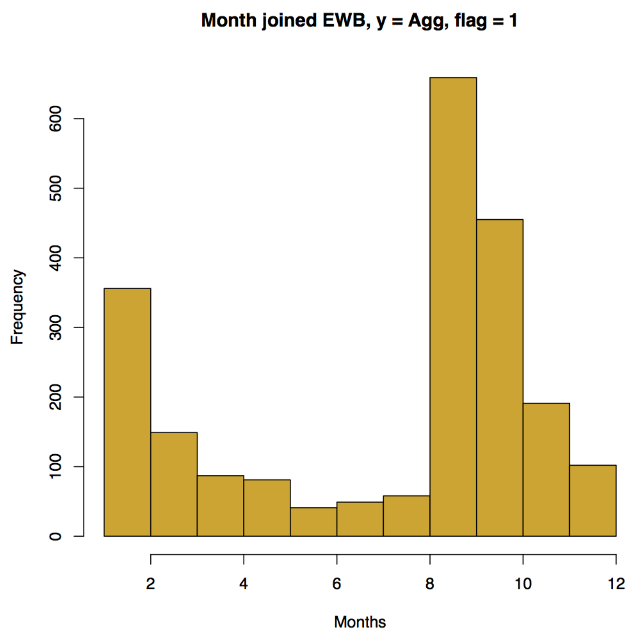
Table 17. Robustness Checks Using Alternative Instruments (friends' covariates), CHAMP Data with Standardized Coefficients

	Dependent variable:					
	total hours			outcome		
	OLS	*2SLS*	*sys GMM*	*OLS*	*2SLS*	*sys GMM*
	(1)	(2)	(3)	(4)	(5)	(6)
Messages Network						
total hours squared				0.315^{*}	0.561^{*}	0.574^{**}
				(0.112)	(0.268)	(0.199)
average peer hours	0.112	0.235^{*}	0.214^{**}	0.195^{*}	0.529^{***}	0.438^{***}
	(0.110)	(0.167)	(0.078)	(0.080)	(0.097)	(0.075)
Observations	361	361	361	361	361	361
J test						58.55
p-value						0.023
First Stage Statistics for total hours squared						
First Stage F Stat					53.471^{**}	
First Stage R^2					0.067	
First Stage Statistics for average peer hours						
First Stage F Stat		19.505^{***}			559.66^{***}	
First Stage R^2		0.251			0.315	
Threaded Comments Network						
total hours squared				0.303^{*}	0.301^{*}	0.349^{**}
				(0.115)	(0.143)	(0.111)
average peer hours	0.073	0.323^{***}	0.278^{***}	0.225^{***}	0.504^{***}	0.471^{***}
	(0.038)	(0.068)	(0.055)	(0.039)	(0.082)	(0.066)
Observations	405	405	405	405	405	405
J test						44.62
p-value						0.07
First Stage Statistics for total hours squared						
First Stage F Stat					13.43^{***}	
First Stage R^2					0.054	
First Stage Statistics for average peer hours						
First Stage F Stat		18.658^{***}			364.57^{***}	
First Stage R^2		0.169			0.308	
Group 50 Network						
total hours squared				0.326^{*}	0.668^{*}	0.808^{***}
				(0.120)	(0.300)	(0.202)
average peer hours	0.020	0.302^{*}	0.256^{**}	0.258^{***}	0.503^{***}	0.405^{***}
	(0.093)	(0.127)	(0.089)	(0.068)	(0.123)	(0.098)
Observations	499	499	499	499	499	499
J test						64.68
p-value						0.01
First Stage Statistics for total hours squared						
First Stage F Stat					894.41^{*}	
First Stage R^2					0.067	
First Stage Statistics for average peer hours						
First Stage F Stat		9.245^{***}			38.74^{***}	
First Stage R^2		0.236			0.315	
Note:					*p<0.05; **p<0.01; ***p<0.001	

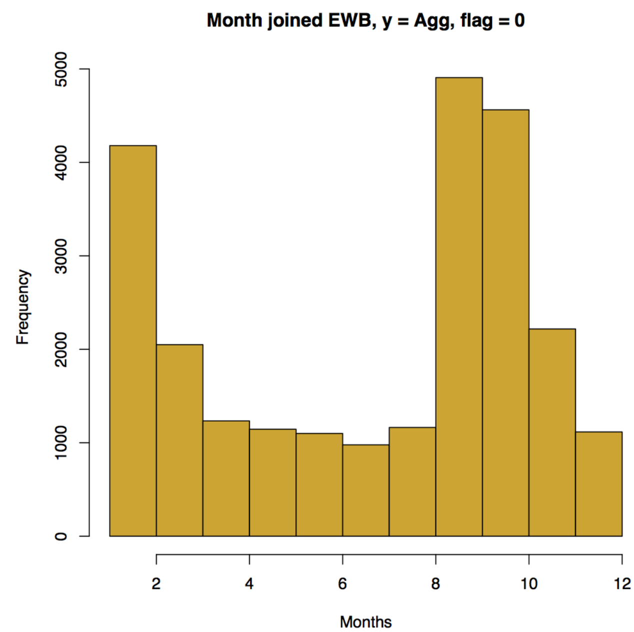
Table 18. Robustness Checks Using Alternative Instruments (friends' covariates), Campaign Data with Standardized Coefficients

	Dependent variable:					
	total email			total amount		
	OLS	*2SLS*	*sys GMM*	*OLS*	*2SLS*	*sys GMM*
	(1)	(2)	(3)	(4)	(5)	(6)
	Messages Network					
total email squared				0.669***	0.557***	0.584***
				(0.049)	(0.110)	(0.073)
average peer emails	0.092	0.235*	0.174*	-0.048	0.043	0.093
	(0.115)	(0.102)	(0.071)	(0.075)	(0.097)	(0.053)
Observations	360	360	360	360	360	360
J test						42.81
p-value						0.60
	First Stage Statistics for total email squared					
First Stage F Stat					12.902**	
First Stage R^2					0.155	
	First Stage Statistics for average peer emails					
First Stage F Stat		91.023***			43.180***	
First Stage R^2		0.226			0.294	
	Threaded Comments Network					
total email squared				0.718***	0.510***	0.518***
				(0.039)	(0.116)	(0.101)
average peer emails	0.195***	0.369**	0.343***	-0.043	0.073	0.209***
	(0.032)	(0.119)	(0.075)	(0.074)	(0.155)	(0.056)
Observations	334	334	334	334	334	334
J test						38.61
p-value						0.70
	First Stage Statistics for total email squared					
First Stage F Stat					1611.79**	
First Stage R^2					0.098	
	First Stage Statistics for average peer emails					
First Stage F Stat		41.586***			130.32***	
First Stage R^2		0.154			0.320	
	Group Membership Network					
total email squared				0.674***	0.547***	0.527***
				(0.048)	(0.124)	(0.059)
average peer emails	0.135*	0.329***	0.223***	0.178**	0.278**	0.235***
	(0.057)	(0.085)	(0.040)	(0.063)	(0.100)	(0.038)
Observations	426	426	426	426	426	426
J test						52.69
p-value						0.14
	First Stage Statistics for total email squared					
First Stage F Stat					677.21**	
First Stage R^2					0.093	
	First Stage Statistics for average peer emails					
First Stage F Stat		26.496***			99.535***	
First Stage R^2		0.343			0.553	
Note:						*p<0.05; **p<0.01; ***p<0.001

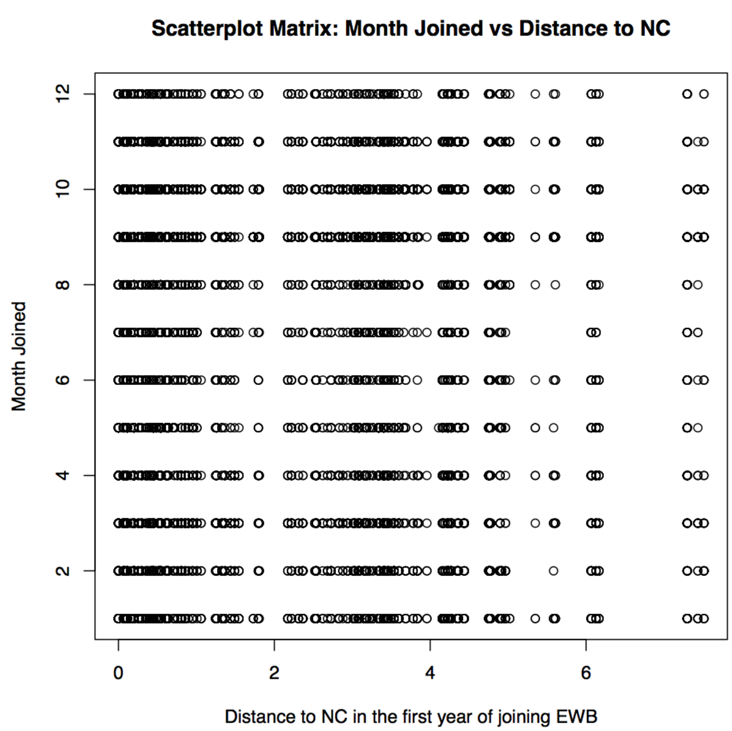
Figure 1. Patterns of month joined EWB among members who attended the national conference in the first year versus those who did not.



(a) Month joined EWB for members who attend the national conference in the first year.

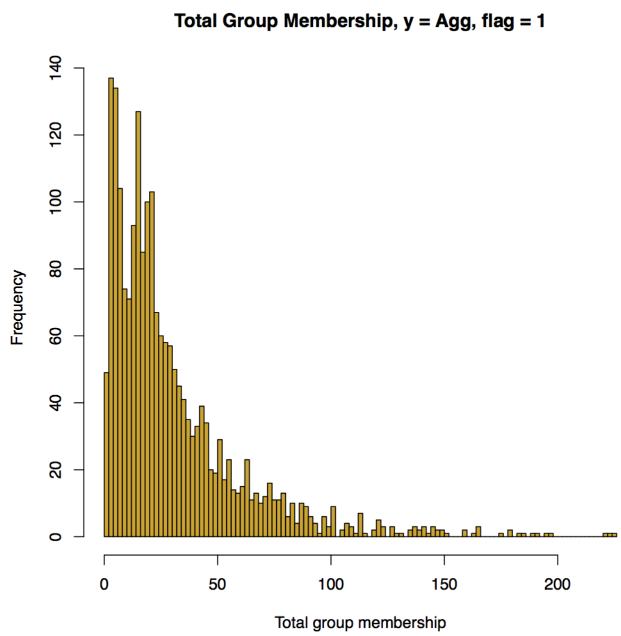


(b) Month joined EWB for members who did not attend the national conference in the first year.

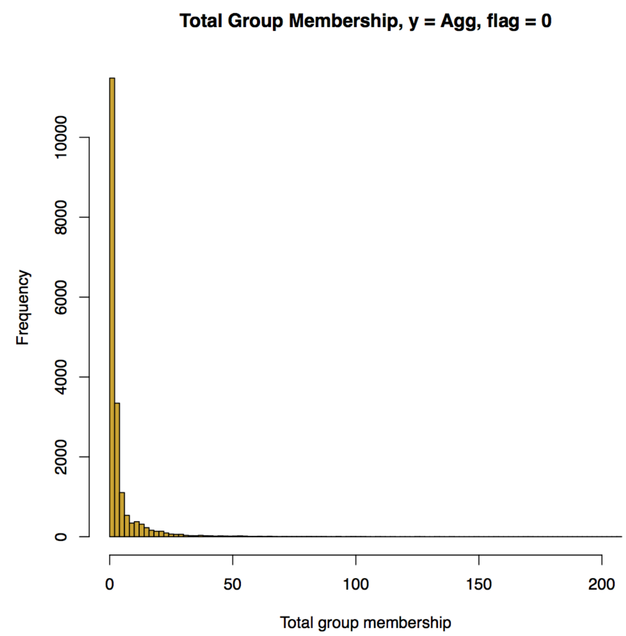


(c) Correlation between month joined EWB and distance to national conference in the first year.

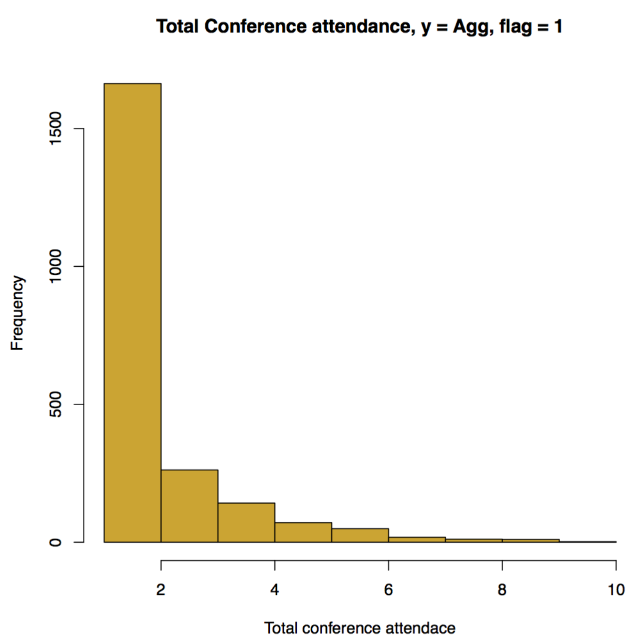
Figure 2. Difference in the effort levels between members who attended the national conference in the first year versus those who did not. Note: Flag = 1 for members who attend the national conference in the first year, and Flag = 0 otherwise.



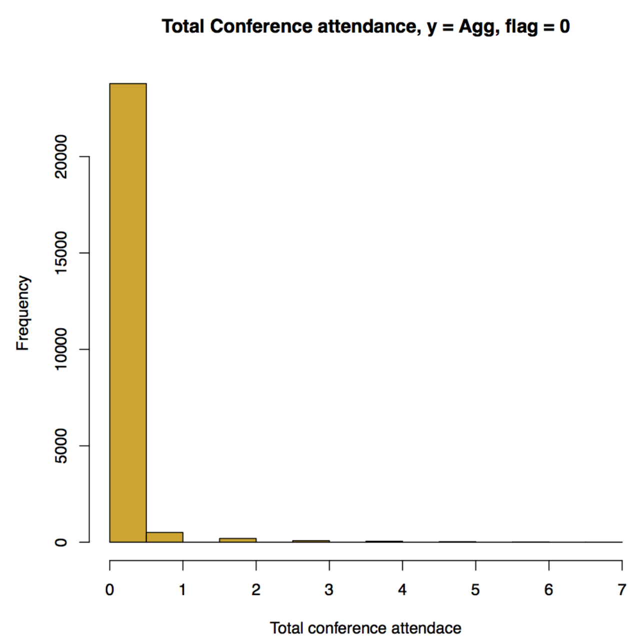
(A) Total group membership, Flag = 1



(B) Total group membership, Flag = 0

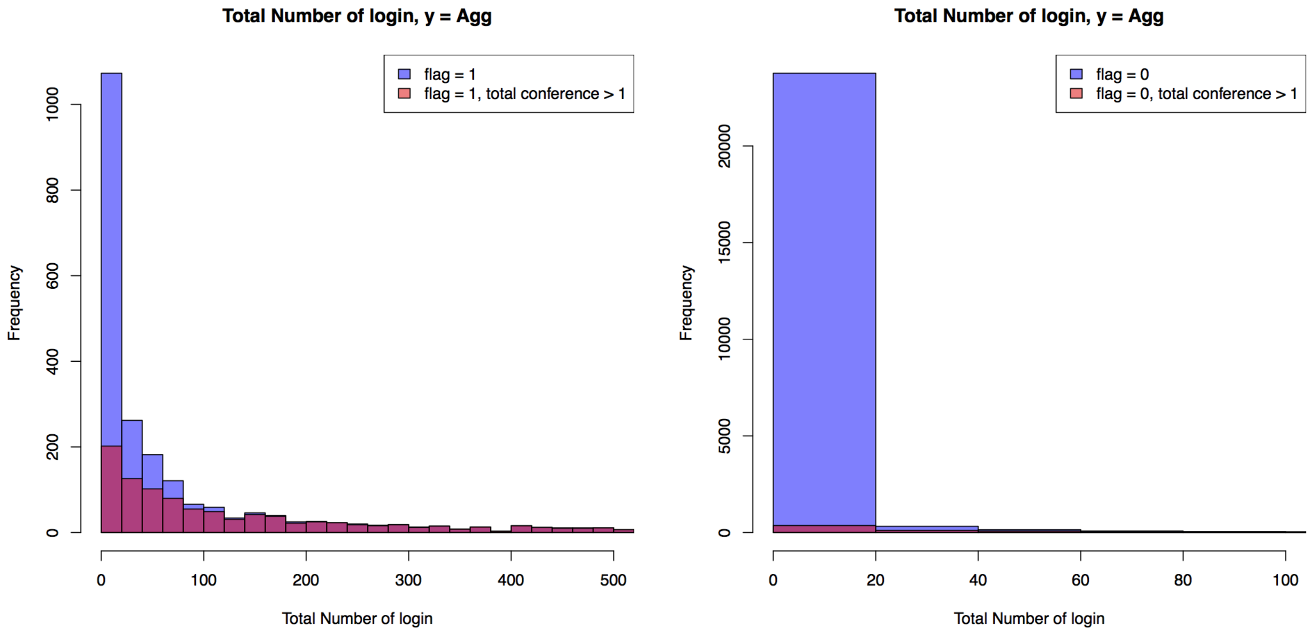


(C) Total number of conferences, Flag = 1



(D) Total number of conferences, Flag = 1

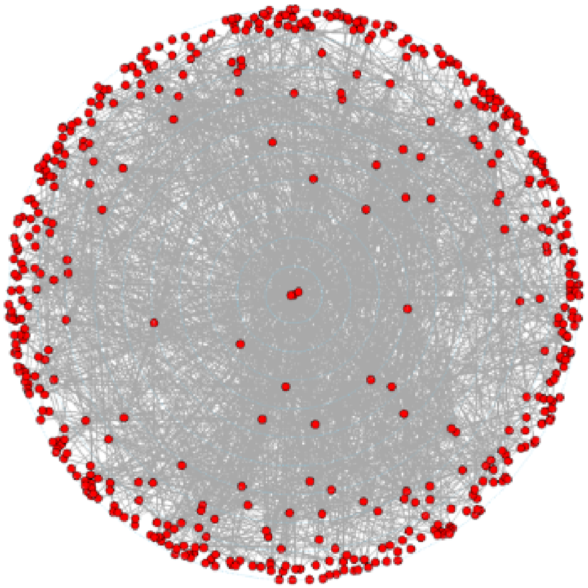
FIGURE 3. Difference in the effort levels between members who attended the national conference in the first year versus those who did not. Note: Flag = 1 for members who attend the national conference in the first year, and Flag = 0 otherwise.



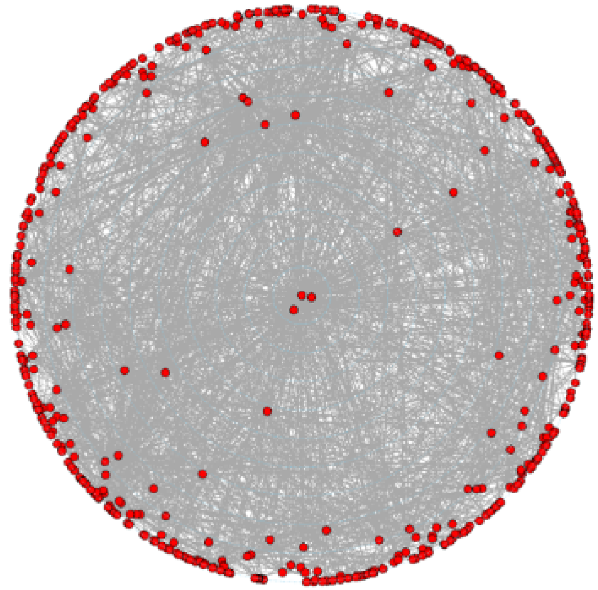
(A) Total number of logins, Flag = 1

(B) Total number of logins, Flag = 1

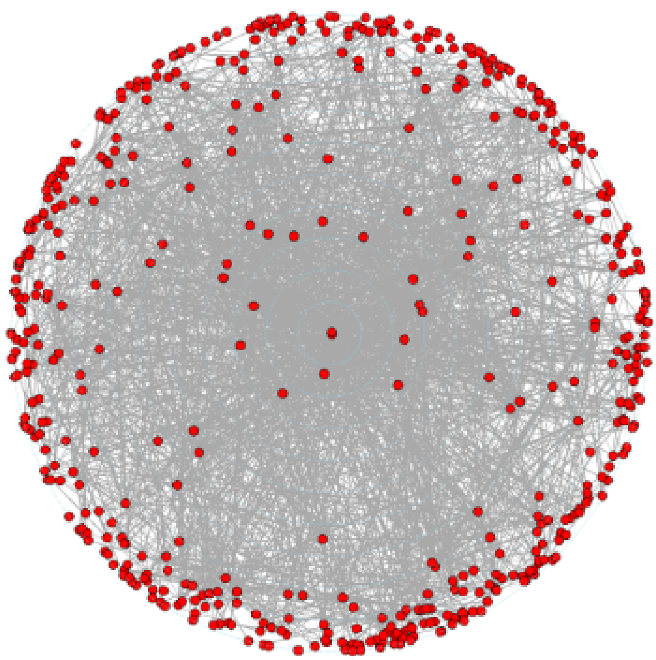
Figure 4. Messages network among individuals who participate in the engagement activities



(a) Network visualization based on degree centrality.

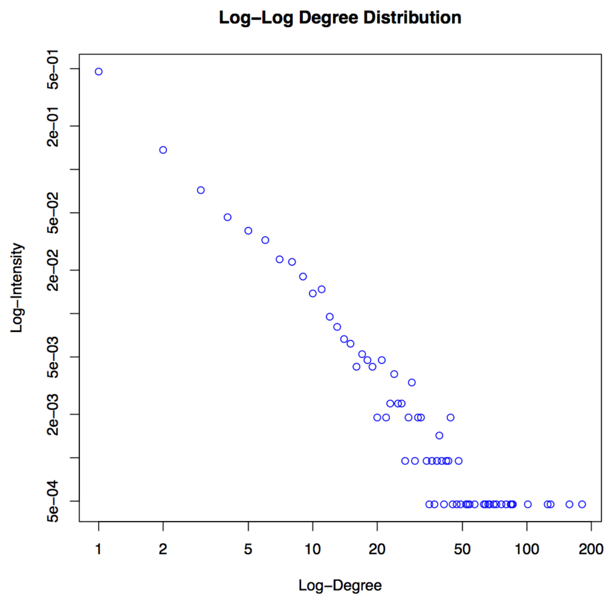


(b) Network visualization based on betweenness centrality.

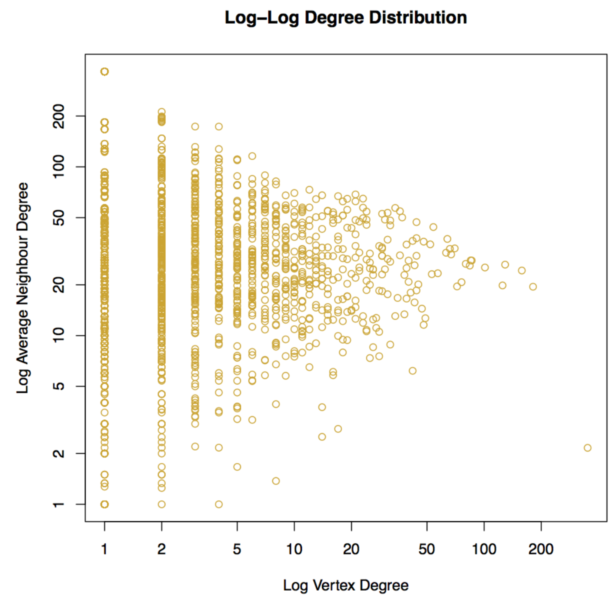


(c) Network visualization based on eigenvector centrality

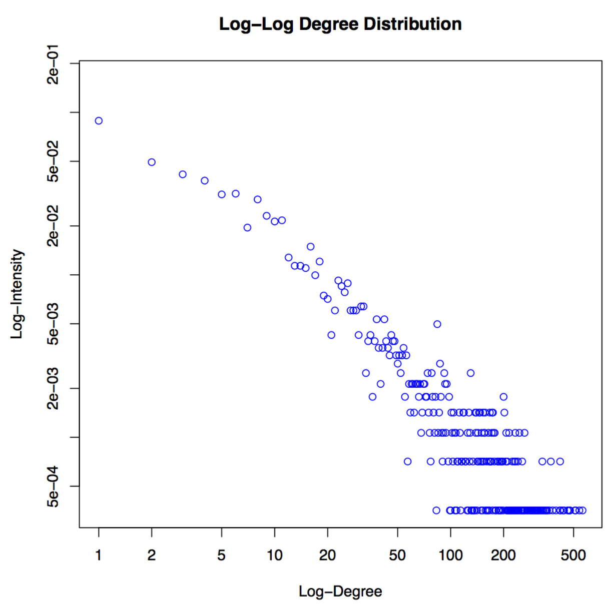
Figure 5. Network Characteristics and Degree Distributions



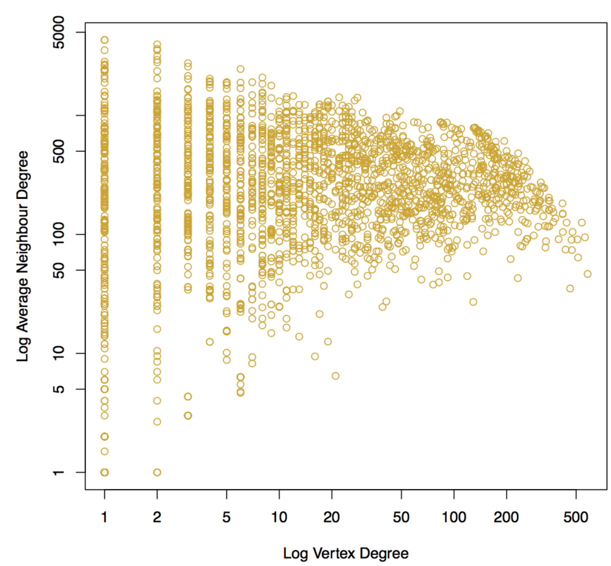
(A) Messages Network: Degree distribution (log-log scale)



(B) Messages Network: Average neighbour degree versus node degree (log-log scale)

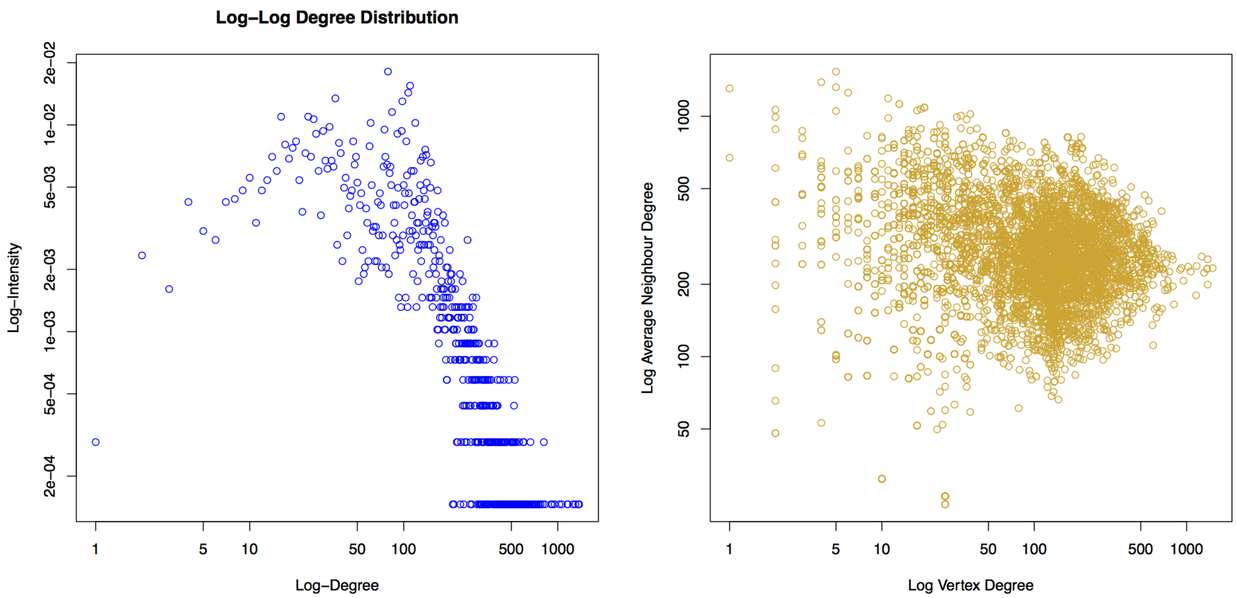


(C) Threaded Comments Network: Degree distribution (log-log scale)



(D) Threaded Comments Network: Average neighbour degree versus node degree (log-log scale)

FIGURE 6. Network Characteristics and Degree Distributions



(A) Group Membership Network: Degree distribution (log-log scale)

(B) Group Membership Network: Average neighbour degree versus node degree (log-log scale)

TABLE 19. Correlation Between Network Centrality Measures and CHAMP Variables

	Messages Network					
	Outcome	Total Effort	Betweenness	Degree	Eigenvector	Closeness
Outcome	1					
Total Effort	0.707	1				
Betweenness	0.202	0.095	1			
Degree	0.271	0.130	0.915	1		
Eigenvector	0.345	0.182	0.757	0.925	1	
Closeness	0.120	0.035	0.107	0.199	0.200	1
	Threaded Comments Network					
Outcome	1					
Total Effort	0.728	1				
Betweenness	0.376	0.254	1			
Degree	0.526	0.342	0.782	1		
Eigenvector	0.526	0.343	0.688	0.981	1	
Closeness	0.133	0.080	0.166	0.300	0.297	1
	Group Membership Network					
Outcome	1					
Total Effort	0.693	1				
Betweenness	0.316	0.205	1			
Degree	0.408	0.269	0.777	1		
Eigenvector	0.390	0.262	0.696	0.982	1	
Closeness	0.324	0.202	0.520	0.815	0.808	1

Table 20. Correlation Between Network Centrality Measures and Fundraising Campaigns Variables

	Messages Network					
	Total Amount	Total Email	Betweenness	Degree	Eigenvector	Closeness
Total Amount	1					
Total Emails	0.685	1				
Betweenness	0.202	0.095	1			
Degree	0.271	0.130	0.915	1		
Eigenvector	0.100	0.146	0.685	0.888	1	
Closeness	0.054	0.091	0.141	0.249	0.196	1
	Threaded Comments Network					
Total Amount	1					
Total Emails	0.712	1				
Betweenness	0.310	0.282	1			
Degree	0.367	0.367	0.761	1		
Eigenvector	0.375	0.381	0.658	0.970	1	
Closeness	0.081	0.066	0.124	0.259	0.247	1
	Group Membership Network					
Outcome	1					
Total Effort	0.694	1				
Betweenness	0.191	0.178	1			
Degree	0.301	0.310	0.740	1		
Eigenvector	0.325	0.340	0.630	0.967	1	
Closeness	0.253	0.255	0.581	0.916	0.896	1