Friend-based targeting

Francis Bloch* Matthew Olckers†

June 18, 2018

Abstract

Each individual i in a community has a private characteristic $\theta_i \in \mathbb{R}$, such as need for a social grant, entrepreneurial ability, or aptitude for a job. Information about the characteristic is ordinal, characteristics can be compared but not measured, and local: individuals only observe the characteristics of their neighbors in the social network. A social planner wants to construct a complete ranking of the characteristics. We show that an incentive compatible and efficient mechanism must take the form of "friend-based" targeting—individuals are asked to report the relative ranking of their neighbors. A complete ranking can only be obtained if for any pair ij there exists an individual $k \neq ij$ which is connected both to i and j. We characterize the sparsest network for which this condition is satisfied as the friendship graph of Erdős-Rényi-Sós (1966). We use data on social networks in India and Indonesia to assess the efficiency of friend-based targeting when the ranking is incomplete.

Keywords: targeting, social network *JEL*: D82, D85, I38.

*Paris School of Economics and Université Paris 1 Panthéon-Sorbonne, francis.bloch@univ-paris1.fr

†Paris School of Economics and Université Paris 1 Panthéon-Sorbonne , matthew.olckers@psemail.eu

1 Introduction

1.1 Introduction

In many social networks, individuals gradually acquire information about their neighbors through repeated interactions. Pupils in a class learn about the ability of other pupils with whom they write joint projects, workers in a firm learn about the productivity of the coworkers in their teams, members of a community in a developing country learn about the wealth of their close friends. This information, which is disseminated in the social network, may be of great use to an outside agent who wants to extract information about members of the community. A teacher wants to learn the ability of his pupils, an employer the productivity of his workers, a funding agency the wealth of villagers in a developing country.

In the classical literature on mechanism design, the principal designs a mechanism which asks agents to report on their own types. However, there is a large number of mechanisms used in practice where agents are asked to report not on their own type but on the type of others. Pupils are asked to assess the performance of other pupils, workers are asked to measure the productivity of their coworkers, villagers are asked to rank other individuals in the community. Our objective of this paper is to analyze these mechanisms, where agents are asked to report about their neighbors in the social network, that we term "friend-based targeting mechanisms". In particular, we want to understand how the architecture of the fixed social network affects the planner's ability to construct a mechanism satisfying desirable properties.

We consider a specific setting, inspired by the study of poverty targeting in villages in developing countries, with two characteristics. First, we suppose that agents only have ordinal information. They lack the ability to measure the characteristics of individuals, and can only assess whether one agent has a higher characteristic than another. All information is thus captured by bilateral comparisons among pairs of agents. In particular, this implies that even by aggregating all local information, the planner may be unable to construct a complete ranking of the characteristics of individuals—an important point that we will discuss later. Second, we assume that the planner only has one instrument at her disposal: she constructs a (complete) ranking of the members of the community. Hence the number of outcomes that the planner can select from is very restricted. The planner cannot use transfers, and cannot punish agents by excluding them from the ranking. In particular, she cannot impose punishments for inconsistent reports, as in the classical literature on implementation with correlated types (Crémer and McLean, 1985).

We require the planner's mechanism to satisfy two properties. First, individuals must have an incentive to report information truthfully. In the ordinal setting we consider, the natural choices for implementation concepts are dominant strategy and ex post implementation. However, we notice that in our setting with a limited number of outcomes, dominant strategy implementation is too strong, leading to impossibility results. We adopt instead ex-post incentive compatibility as the desirable incentive property of the mechanism. Second, we require the mechanism to be ex-post efficient from the point of view of the planner whose objective is to recover the true ranking of individuals in the community. More precisely, the ranking chosen by the planner must match the ranking that society would construct by aggregating all local information. If society can construct a complete ranking of individuals for any realization of types (a situation we label "completely informative"), the ranking chosen by the planner must match the true ranking. Otherwise, we may face different situations according to the realization of types. For some type realizations, even if the information aggregated by society is not complete, transitivity ensures that all agents can be completely ranked. For other type realizations, society will only be able to construct a partial order on the individuals. In the latter case, the complete ranking chosen by the planner will be a completion of the partial order that the community can construct, and this completion will involve arbitrary comparisons across individuals who cannot be compared.

We first analyze mechanism design in completely informative societies. We show that a society is completely informative if any pair of individuals can be compared, either through "self-reports" (the two agents involved in the pair report on each other) or through "friend-based reports" (a third agent observes both agents involved in the pair). Our main Theorem shows that self-reports cannot be used if they are not backed by the report of a third agent. A mechanism satisfying ex-post incentive compatibility and efficiency exists if and only if every pair of agents has a common neighbor. We then characterize the sparsest network which satisfies this property. When the number of agents is odd, this is the "friendship network" of (Erdős, Rényi, and Sós, 1966), the only network in which every pair of agents has a single common friend. This network, also known as the windmill, has an agent in the middle who connects all other agents who form triangles. When the number of agents is even, this is a variant of the windmill where one of the sails contains three agents instead of two. Notice that this in this network, one agent is responsible for a large number of comparisons, so we also investigate networks satisfying this property when the number of comparisons performed by every agent is bounded.

We then turn our attention to societies which are not completely informative. We add another requirement in the mechanism. We guarantee that whenever two agents are incomparable, the mechanism ranks them in the same way independently of (irrelevant) information on other comparisons. Under this independence requirement, we show that any comparison based solely on self-reports must be discarded by the planner, as both agents have an incentive to misreport. Hence the planner can only rely on friend-based comparisons and we construct a "comparison network" by linking two agents if and only if they have a common neighbor. We find that there exist two network architectures for which the planner can construct a mechanism satisfying independence, ex-post incentive compatibility and efficiency. In the first architecture, the social network is bipartite (which is equivalent to the comparison network being disconnected). We use the bipartite structure to partition the set of agents, so that agents in one group rank agents in the other group, and agents across groups are ranked in an arbitrary way. In the second architecture, all links form triangles, and we can construct a mechanism exploiting the fact that any unsupported report is surrounded by supported links. However, we also note that there exist network architectures for which mechanisms satisfying our three properties cannot be constructed. The simplest example is a network with four players, with one triangle and one additional link.

We then illustrate our findings using real-life social networks from two contexts: villages in Karnataka, India, and neighborhoods in Indonesia. To be completed

Finally, we consider different variants and robustness checks of our model. We show that dominant strategy implementation is too strong, leading to an impossibility result in triangles. We analyze the robustness of our mechanism to joint deviations by groups. We study whether coarser rankings are easier or harder to implement than complete rankings. We discuss alternative measures of information contained in networks, beyond a simple count of bilateral comparisons.

In practical terms, our analysis points to two important facts. First, it shows that it may be useful to partition the set of agents into different groups asking agents in one group to rank another. For example, one may want to let men rank women and women rank men in a community. This procedure will result in a truthful and efficient ranking, but the price to be paid is that inter rankings among agents in the two groups will be arbitrary. Second, the analysis highlights the importance of triangles. Truthful and efficient comparisons will be possible if all links form triangles, suggesting that friend-based targeting should only be used in societies with high clustering. As high clustering is also associated to high density and low average distances, we conclude that friend-based targeting should only be used in communities with dense social networks with low diameters.

1.2 Literature review

We first discuss the relation of our paper with the literature on community-based targeting in development economics. Policy makers have long known that individuals living within a community have good access to information about their fellow community members. Community-based targeting is implemented by gathering the community to agree on a ranking together. Although these meetings combat incentives to lie (Rai, 2002), they are sensitive to fatigue. Even in small communities of 54 households on average the ranking procedure took an average of 1.68 hours to complete, and household were ranked more accurately at the beginning than at the end (Alatas, Banerjee, Hanna, Olken, and Tobias, 2012). Friend-based targeting avoids fatigue since individuals do not need to be present at the same meeting. The planner only needs each individual to rank his friends, not the entire community. We focus on only the friend group since it has been shown that information decreases sharply with social distance (Alatas, Banerjee, Chandrasekhar, Hanna, and Olken, 2016). Individuals can compare their friends but are less accurate and less confident about their comparisons at distance two (friends of friends) or greater.

Friend-based targeting is not limited to social relationships. Colleagues at work or neighboring households may have also have good information about private characteristics. Our results apply to any network where each node reports ranking of his neighbors on the network. A recent field experiment Maharashtra, India grouped five geographically close entrepreneurs and asked that they rank each other according to business profitability (Hussam, Rigol, and Roth, 2017). The population of 1380 individuals is connected on a network of 277 separate cliques of around five individuals. Even though this network is close to lower bound of information, it proved to be a more accurate method of targeting than using observable information about the entrepreneurs.

We see friend-based targeting as a complement to current targeting popular methods such as proxy-means tests. Although widely used, the proxy means test has shown to perform only slightly better than universal transfers at reducing poverty (Brown, Ravallion, and Van de Walle, 2016) and to lack adjustment to transitory shocks (Coady, Grosh, and Hoddinott, 2004; Basurto, Dupas, and Robinson, 2017). Given the rise of mobile money, friend-based targeting and the corresponding transfers could be implemented quickly via mobile phones to respond to transitory shocks. The proxy means test could help to validate the accuracy of the comparisons reported via friend-based targeting.

The theoretical analysis of the paper is closely related to the small literature studying peer selection in computer science and social choice theory. Alon, Fischer, Procaccia, and Tennenholtz (2011) analyzes the design of mechanisms to select a group of k individuals among their peers. As in approval voting, each individual provides a list of peers she wishes to nominate. The mechanism then selects a subset of agents. Each agent obtains a positive utility if she is nominated. The mechanism is called *impartial* if it is strategy-proof: no individual has an incentive to lie in order to increase his chances of being selected. Alon, Fischer, Procaccia, and Tennenholtz (2011) prove a strong negative result: there does not exist any deterministic impartial mechanism which guarantees an outcome which reaches a positive fraction of the efficiency frontier. Approximately efficient stochastic impartial mechanisms can be constructed. They are based on the random partition of individuals into clusters of fixed size such that agents inside a cluster rank agents outside the cluster, as in the case of bipartite social graphs we consider here. Holzman and Moulin (2013) analyze impartial voting rules when agents nominate a single individual. They identify a class of desirable voting rules as two-step mechanisms, where voters are first partitioned into districts which elect local winners, who are then competing against one another to select the global winner. However, Holzman and Moulin (2013) also highlight a number of impossibility results, showing that there is no impartial voting procedure which treats voters symmetrically, and that there is no impartial voting procedure which guarantees that an agent whom nobody considers best will be elected and that an agent whom everybody considers best will always be elected. Kurokawa, Lev, Morgenstern, and Procaccia (2015) and Aziz, Lev, Mattei, Rosenschein, and Walsh (2016) improve on the partition algorithm proposed in Alon, Fischer, Procaccia, and Tennenholtz (2011). They consider a more general setting, inspired by the new peer review system experimented by the National Science Foundation to fund the Sensors and Sensing System program. Applicants are reviewers and each reviews and assigns grades to m proposals. The funding agency selects k projects using an impartial mechanism. Kurokawa, Lev, Morgenstern, and Procaccia (2015) propose the "credible subset mechanism", a process which first identifies candidates which are likely to win, and with some probability does not output any result. The credible subset mechanism achieves a high fraction ($\frac{k}{k+m}$) of the total efficiency. Aziz, Lev, Mattei, Rosenschein, and Walsh (2016) propose a mechanism combining the insights of the partition mechanism of Alon, Fischer, Procaccia, and Tennenholtz (2011) with the impartial "divide the dollar" mechanism of De Clippel, Moulin, and Tideman (2008). They show that the mechanism satisfies strategy-proofness and monotonicity.

Our model departs from the models of peer selection in a number of ways. First, we consider *ordinal* rather than cardinal information as inputs of the mechanism. In our model, individuals do not assign grades to other individuals, but have only access to bilateral comparisons. Second, we consider as output a *complete ranking* of individuals rather than a coarse ranking into two sets of acceptable and non-acceptable candidates. (However, we also consider coarser rankings as a possible extension of our model). Third, because dominant-strategy mechanisms do not exist, we weaken the incentive requirement to ex-post implementation, thereby obtaining positive results which differ from the results obtained in the peer selection literature. Fourth and foremost, we do not assume a specific assignment of proposals to reviewers, but consider an arbitrary network of observation captured by a social network. Our main objective is then to characterize those social networks (or structures of observability) for which mechanisms satisfying desirable properties can be constructed.

The paper which is probably the more closely connected to ours is a recent paper by Baumann (Baumann (2017)) which analyzes network structures for which it is possible to identify the agent with the highest characteristic. Baumann (2017) constructs a specific multi-tier mechanism identifying the top agent from the reports of his neighbors, shows that the mechanism is ex-post incentive compatible and identifies network structures for which truth-telling is the only equilibrium of the mechanism. Our paper differs in many dimensions. First, we consider an ordinal rather than a cardinal setting, giving rise to the possibility of incompleteness of the social ranking. Second, we suppose that the objective of the planner is to rank all agents rather than identify the top agent. Third, we do not assume an exogenous bound on the way in which agents can misreport (this exogenous bound plays a crucial role to allow the construction of an ex-post incentive compatible mechanism without transfers in Baumann (2017)). Our results are thus not directly comparable to hers.

1.3 Structure of the paper

The rest of the paper is organized as follows. The next Section presents the model. Section 3 discusses societies in which complete rankings can be constructed. Section 4 focuses on mechanism design when the social ranking is incomplete. Section 5 illustrates our finding in two real world contexts: social networks in India and Indonesia. Section 6 discusses extensions of the model and Section 7 contains our conclusions.

2 Model

2.1 Individuals and communities

We consider a community N of n individuals indexed by $i = 1, 2, ..., n$. Each individual i has a characteristic $\theta_i \in \mathbb{R}$. Examples of θ_i include wealth, aptitude for a job, or quality of a project. Characteristics are privately known and are drawn from a non-atomic continuous distribution F.

Members of the community are linked by a connected, undirected graph g. The social network g is commonly known by the agents and the planner. The characteristic of agent i, θ_i, can be observed by agent i and by all her direct neighbors in the social network g. We suppose that individuals cannot provide an accurate value for the characteristic θ_i. Either the characteristic cannot be measured precisely, or individuals do not have the ability or the language to express a precise measure for θ_i. Instead, we assume that individuals only possess *ordinal* information and are able to compare the characteristics of two individuals. For any agent i and any pair of agents (j, k) that agent i can observe, we let $t^i_{jk} = 1$ if agent i observes that $\theta_j > \theta_k$ and $t^i_{jk} = -1$ if agent i observes that $\theta_j < \theta_k$. The ordinal comparison is assumed to be perfect: agent i always perfectly observes whether agent j's characteristic is higher than agent k's. Given that the characteristics are drawn from a non-atomic continuous distribution, we ignore situations where the two characteristics are equal.

Agent i's information (and type) can thus be summarized by a matrix $T^i = [t^i_{jk}]$ where $t^i_{jk} \in \{-1, 0, 1\}$ and $t^i_{jk} \neq 0$ if and only if i observes the comparison between j and k, namely either $i = j$ or $i = k$ or $g_{ij}g_{ik} = 1$. When $i = j$ or $i = k$, we call the comparison t^i_{jk} a *self-comparison*. When $g_{ij}g_{ik} = 1$, we call the comparison t^i_{jk} a *friend-based comparison*. Notice that the matrix T^i is formally identical to a *tournament matrix*, as defined in social choice theory ((Laslier, 1997)). We let $\mathcal{T}$ denote the set of $n \times n$ tournament matrices.

The vector $\mathbf{T} = (T^1, .., T^n)$ describes the information possessed by the community on the ranking of the characteristics θ_i. Obviously, because agent's observations are perfectly correlated, individual types T^i and T^j will be correlated if there exists a pair of agents (k, l) such that $t^i_{kl} \neq 0$ and $t^j_{kl} \neq 0$. Hence, if the planner could construct a punishment for contradictory reports as in Crémer and McLean (1985), he would be able to induce the agents to report their true type. However, we do not permit punishments in our setting, as we assume that transfers are not possible, and that the only instrument available to the planner is the ranking ρ.

The information contained in the vector $\mathbf{T} = (T^1, .., T^n)$ results in a partial ranking of the characteristics of the agents, that we denote by $\succ$. We let $i \succ_{\mathbf{T}} j$ if the information contained in $\mathbf{T}$ allows to conclude that $\theta_i > \theta_j$.

For a fixed social network g, the information contained in the vector $\mathbf{T} = (T^1, .., T^n)$ may not be the same for different realizations of $(\theta^1, .., \theta^n)$. This is due to the fact that ordinal comparisons are transitive, so that basic bilateral comparisons may induce additional bilateral comparisons by transitivity. However the transitive closure depends on the bilateral comparisons, which are different for different realizations of $(\theta^1, .., \theta^n)$.To illustrate this point, consider four agents $i = 1, 2, 3, 4$ organized in a line as in Figure 1

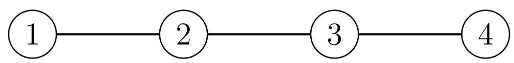


Figure 1: A line of four individuals

If $\theta_1 < \theta_2 < \theta_3 < \theta_4$, then as $t^1_{12} = t^2_{12} = -1, t^2_{23} = t^3_{23} = -1, t^3_{34} = t^4_{34} = -1$, the comparisons result in a complete ranking $1 \prec 2 \prec 3 \prec 4$. However, for other possible realizations of $(\theta_1, \theta_2, \theta_3, \theta_4)$, the ranking generated by the types $\mathbf{T}$ may be incomplete. For example if $\theta_1 < \theta_4 < \theta_2 < \theta_3$, we obtain $1 \prec 2 \prec 3$ and $4 \prec 2 \prec 3$, but 1 and 4 cannot be compared.

A social network g is called *completely informative* if, for any realization of the characteristics $(\theta_1, .., \theta_n)$, the information contained in $\mathbf{T}$ results in a complete ranking of the members of the community. The following Lemma characterizes completely informative social networks.

Lemma 1. *A social network g is completely informative if and only if, for any pair of individuals (i, j) either $g_{ij} = 1$ or there exists an agent k such that $g_{ik}g_{jk} = 1$.*

Proof. The condition is obviously sufficient, as it guarantees that for any pair (i, j) there exists an individual k such that $t^k_{ij} \neq 0$, Hence the matrix generated by $(T^1, .., T^n)$ contains nonzero entries everywhere outside the diagonal. Conversely, suppose that there exists a pair of individuals (i, j) who is observed by no other player and such that $g_{ij} = 0$. Consider a realization of the characteristics such that θ_i and θ_j are two consecutive values. No individual can directly compare i and j. In addition, because there is no k such that $\theta_k \in (\theta_i, \theta_j)$, there is no k such that $\theta_i \prec \theta_k \prec \theta_j$ or $\theta_j \prec \theta_k \prec \theta_i$. Hence the social network g is not completely informative. □

Lemma 1 that a social network is completely informative if and only if *any pair of individuals* can be compared either by self-comparisons or by friend-based comparisons.

2.2 Planner and mechanism design

The objective of the planner is to construct an ordinal ranking of agents according to the value of the characteristic θ_i. For example, an outside agency wants to rank members of the community according to their wealth, an employer wants to rank workers according to their ability, a bank wants to rank projects according to their profitability. We let ρ denote the complete ranking chosen by the planner, a permutation over the set N of agents. The set of all permutations is denoted $\mathcal{P}$. The rank of agent i is denoted ρ_i, with the convention that higher values of ρ_i are associated to higher values of the characteristic θ_i.

The planner wishes to construct a ranking as close as possible to the true ranking of characteristics θ_i. We do not specify explicitly the preferences of the planner. In the ordinal setting we consider, different measures of distances between rankings can be constructed. Instead of describing explicitly the loss function associated to differences in rankings, we will focus attention on efficient mechanisms. Noting that the community may not be able to completely rank individuals, efficiency will only require that the ranking ρ_i coincides with the ranking generated by $\mathbf{T}$ for any pair of agents (i, j) which can be compared under $\mathbf{T}$.

We finally specify the preferences of individuals. Members of the community only care about their rank ρ_i and have strict preferences over ρ_i. By convention, agents prefer higher rankings. Hence ρ_i is preferred to ρ'_i if and only if $\rho_i \geq \rho'_i$. In particular, we assume that there are no externalities in the community: individuals do not care about the ranking of other members of the community. This assumption implies that individuals do not derive any satisfaction from high rankings of friends or low rankings of foes.

A *direct mechanism* associates to any vector of reported tournament matrices $\mathbf{T} \in \mathcal{T}^n$ a complete ranking $\rho \in \mathcal{P}$. We impose two conditions on the mechanism:

Ex-post incentive compatibility For any agent i, for any vector of types $\mathbf{T} = (T^i, T^{-1})$, any type T'^i,

$$\rho_i(\mathbf{T}) \geq \rho_i(T'^i, T^{-i}).$$

Ex-post efficiency For any vector of types $\mathbf{T}$,

$$\text{If } i \succ_{\mathbf{T}} j, \text{ then } \rho_i(\mathbf{T}) > \rho_j(\mathbf{T}).$$

We focus on ex-post implementation for two reasons. First, given that we consider an ordinal setting, we do not wish to specify the distribution of types, and hence select a robust implementation concept which does not depend on the distribution of types. Second, as we show in Section 6, the alternative robust implementation concept, dominant strategy implementation, is too strong for our setting.

Ex-post efficiency requires that the planner's ranking coincide with the true ranking of characteristics in the following weak sense. Whenever two agents i and j can be ranked using the information contained in $\mathbf{T}$, the ranking ρ_i must be consistent with the ranking between i and j. As the order relation induced by $\mathbf{T}$, $\succ_{\mathbf{T}}$, may be very incomplete, the requirement may be very weak. The ranking ρ must be a completion of the ranking $\succ_{\mathbf{T}}$, If $\succ_{\mathbf{T}}$ is a very small subset of N^2, the ranking ρ may end up being very different from the "true" ranking of characteristics θ_i. However as the "true" ranking of characteristics cannot be constructed using the local information of the social network, the difference between ρ and the "true" ranking should not be a matter of concern, as the planner chooses an efficient mechanism given the information available to the community.

3 Completely informative rankings

3.1 Completely informative rankings

We first analyze conditions under which an ex-post incentive compatible and efficient mechanism can be constructed when the information available in the community results in a complete ranking. By Lemma 1, all pairs of agents must be either directly connected or commonly observed by a third agent. The next Theorem shows that for an ex-post incentive compatible and efficient mechanism to exist, all pairs of agents must be observed by a third agent.

Theorem 1. *Suppose that the social network g is completely informative. An ex-post incentive compatible and efficient mechanism exists if and only if, for all pair of agents (i, j) there exists a third agent k who observes both i and j, $g_{ik}g_{jk} = 1$.*

Proof. (Sufficiency) Suppose that for any pair of agents (i, j), there exists a third agent k for whom $g_{ik} = g_{jk} = 1$. We define the mechanism ρ by constructing bilateral comparisons. Let r_{ij} denote the bilateral comparison between i and j chosen by the planner.

First consider a pair of agents (i, j) who observe each other, $g_{ij} = 1$. By assumption, there are at least three reports on the ranking of i and j. If all agents transmit the same report on (i, j), let $r_{ij} = t_{ij}$. If all agents but one transmit the same report t_{ij} and one agent reports $t'_{ij} = -t_{ij}$, ignore the ranking t'_{ij} and let $r_{ij} = t_{ij}$. In all other cases, let $r_{ij} = 1$ if and only if $i > j$.

Second consider a pair of agents (i, j) who do not observe each other, $g_{ij} = 0$. By assumption, there exists at least one agent k who observes them both. If there are at least three agents who observe i and j, use as above a mechanism such that $r_{ij} = t_{ij}$ if all agents agree on t_{ij} or only one agent chooses $t'_{ij} = -t_{ij}$, and let $r_{ij} = 1$ if $i > j$ otherwise.

If one or two agents observe i, j, pick the agent k with the highest index. Consider the vector of announcements $\tilde{T}^{-k}$ where one disregards the announcements of agent k. Let $\succ_{\tilde{T}^{-k}}$ be the binary creation created by letting $t_{ij} = 1$ if and only if $t^l_{ij} = 1$ for all $l \neq k$. If there exists a directed path of length greater or equal to 2 between i and j in $\succ_{\tilde{T}^{-k}}$, and for all directed paths between i and j in $\succ_{\tilde{T}^{-k}}$, $i^0, .., i^L$ we have $t_{i^l i^{l+1}} = 1$, then $r_{ij} = 1$. If on the other hand for all directed paths between i and j in $\succ_{\tilde{T}^{-k}}$, $t_{i^l i^{l+1}} = -1$, then $r_{ij} = -1$. In all other cases, let agent k dictate the bilateral comparison between i and j, $r_{ij} = t^k_{ij}$.

Now consider all bilateral comparisons r_{ij}. If they induce a transitive binary relation on N, let ρ be the complete order generated by the bilateral comparisons. Otherwise, consider all shortest cycles generated by the binary relation r_{ij}. If there exists a single agent i who dictates at least two bilateral comparisons in all shortest cycles, agent i is punished by setting $\rho_i = 1$ and $\rho_j > \rho_k$ if and only if $j > k$ for all $j, k \neq i$. If this is not the case, pick the arbitrary ranking where $\rho_i > \rho_j$ if and only if $i > j$.

We now show that the mechanism ρ is ex-post incentive compatible and ex-post efficient.

Suppose that all agents except k report their true type, and consider agent k's incentive to report $T'^k \neq T^k$. On any link (i, j) such that $g_{ij} = 1$, as all other agents make the same announcement, agent k cannot change the bilateral comparison r_{ij} by misreporting.

Consider a link (i, j) such that $g_{ij} = 0$ and $g_{ik}g_{jk} = 1$. If there are at least three agents who observe i and j, agent k cannot affect the outcome. Otherwise, if there is a directed path of length greater than equal to 2 in $\succ_{\tilde{T}^{-k}}$, agent k's report cannot change the ranking. If agent k is not the highest ranked agent who observes i and j, then she cannot change the bilateral comparison r_{ij} by misreporting. Hence we only need to focus attention on pairs (i, j) such that k is the highest index agent who observes i and j and there is no directed path between i and j in $\succ_{\tilde{T}^{-k}}$.

Suppose that all agents $l \neq k$ announce the truth, so that $\succ_{\tilde{T}^{-k}} = \succ_{T^{-k}}$. We first show that individual k cannot gain by making an announcement which generates cycles in the ranking r_{ij}. Suppose that the bilateral relation r_{ij} exhibits cycles. We first claim that the shortest cycles must be of length 3.

Suppose that there exists a cycle of length L, $i^0, i^1, .., i^L$. Because the society is completely informative, the binary comparisons generated by the announcements are complete, so that for any l, m, either $r_{i^l i^m} = 1$ or $r_{i^l i^m} = -1$. Now consider i^0, i^1, i^2. If $r_{i^0 i^2} = -1$, $i^0, i^1 i^2 i^0$ forms a shortest cycle of length 3. If not, $r_{i^0 i^2} = 1$ and we can construct a cycle of length $L - 1$, $i^0 i^2, .., i^L$. By repeating this argument, we either find shortest cycles of length 3 or end up reducing the initial cycle to a cycle of length 3.

Consider next a shortest cycle of $ijli$. We claim that individual k must dictate at least two bilateral comparisons in the cycle.

First note that if k does not dictate the comparison between i and j, there must be a directed path between i and j in $\succ_{T^{-k}}$. To see this, notice that either $g_{ij} = 1$ and then $i \succ_{T^{-k}} j$ or $g_{ij} = 0$ but i and j are not observed by k or are observed by k and another agent with a higher index than k, in which case $i \succ_{T^{-k}} j$. Finally, it could be that $g_{ij} = 0$, i and j are observed by k, k is the highest index agent observing i and j, but then as k

does not dictate the comparison (i, j), there must exist a directed path of length 2 between i and j in $\succ_{T^{-k}}$.

Now suppose first that k does not dictate any bilateral comparison in the cycle. There must exist a directed path between i and j, j and l and l and i in $\succ_{T^{-k}}$, a contradiction since, as all agents tell the truth, the binary relation generated by $\succ_{T^{-k}}$ is transitive.

Next suppose that k dictates a single bilateral comparison (i, j) in the cycle but not the bilateral comparisons (j, l) and (l, i). Then there exists a directed path between j and l and a directed path between l and i in $\succ_{T^{-k}}$. Hence there exists a directed path of length greater than or equal to 2 between j and i in $\succ_{T^{-k}}$. Furthermore, as all agents tell the truth, for all directed paths between j and i, $r_{ji} = 1$. Hence the mechanism cannot let agent k dictate the choice between i and j, yielding a contradiction.

We conclude that all shortest cycles are of length 3, and that in any cycle of length 3, agent k must dictate at least two of the three bilateral comparisons. Hence the mechanism assigns $\rho(k) = 1$ and agent k cannot benefit from inducing a cycle.

Finally suppose that all bilateral comparisons r_{ij} result in a transitive relation so that ρ can be constructed as the complete order generated by these bilateral comparisons. We claim that the comparisons generated by T^{-k} are sufficient to compute the rank of k. In fact, for any $i \neq k$, either $g_{ik} = 1$ and as all other agent tell the truth, r_{ik} is independent of the report t^k_{ik}, or $g_{ik} = 0$ and the report on (i, k) is truthfully made by another agent l. In both cases, the information contained in T^{-k} is sufficient to construct the bilateral comparison r_{ik}. Hence ρ_k is independent of the announcement T^k, concluding the proof that the mechanism is ex-post incentive compatible.

To show that the mechanism is ex-post efficient notice that, when all agents truthfully report their types, the bilateral rankings r_{ij} induce a transitive relation, and yield the complete ranking generated by $\succ_{\mathbf{T}}$.

(Necessity) Suppose that the social network g satisfies the conditions of Lemma 1 but that there exists a pair of agents (i, j) who observe each other but are not observed by any third agent k. Consider a realization of the characteristics such that θ_i and θ_j are the two lowest characteristics. Let $\mathbf{T}_1$ be the type profile if $\theta_i < \theta_j$ and $\mathbf{T}_2$ the type profile if $\theta_j < \theta_i$.

By ex-post efficiency, because the rankings generated by $\mathbf{T}_1$ and $\mathbf{T}_2$ are complete,

$$\begin{aligned} \rho_i(\mathbf{T}_1) &= \rho_j(\mathbf{T}_2) &= 1, \\ \rho_j(\mathbf{T}_2) &= \rho_i(\mathbf{T}_1) &= 2. \end{aligned}$$

Because there are only two announcements t^i_{ij} and t^j_{ij} on the link (i, j), ex-post incentive compatibility requires that agents i and j cannot improve their ranking by changing their reports on the link (i, j). Let T_{-ij} denote the announcements on all links but link ij. We must have

$$\begin{aligned} \rho_i(T_{-ij}, t_{ij}^i = 1, t_{ij}^j = -1) &= \rho_i(\mathbf{T}_1) = 1, \\ \rho_j(T_{-ij}, t_{ij}^i = 1, t_{ij}^j = -1) &= \rho_j(\mathbf{T}_2) = 1. \end{aligned}$$

resulting in a contradiction as i and j cannot both be ranked at position 1. □

Theorem 1 shows that an ex-post incentive compatible and efficient mechanism exists in completely informative societies if and only if, for any pair of individuals (i, j), there exists a third individual k who is a neighbor of both individuals in the graph g. Self-comparisons cannot be used and any bilateral comparison requires the presence of a third agent ("a friend") who observes both agents. If the two agents i and j are connected, this implies that there must exist a triangle (i, j, k) in the network g. In the terminology of Jackson, Rodriguez-Barraquer, and Tan (2012), the link (ij) is "supported". Hence the social network must be a "social quilt" as in Jackson, Rodriguez-Barraquer, and Tan (2012), because all links need to be supported. However, the social network need not be complete: a bilateral comparison between i and j can be obtained by the report of a third agent k even when i and j are not connected.

The intuition underlying Theorem 1 is easy to grasp. If the comparison between θ_i and θ_j can only be reported by the two agents, in an ex-post efficient mechanism, one of them has an incentive to lie. Consider a ranking which places i and j as the two agents with the lowest characteristics in the community. If both announce that θ_i is smaller than θ_j, then $\rho_i = 1, \rho_j = 2$. Similarly, if both announce that θ_j is smaller than θ_i, then $\rho_j = 1, \rho_i = 2$. But by incentive compatibility, neither of the agents can improve his rank by changing her report on t_{ij}. Hence i must still be ranked at position 1 when she announces $\theta_i > \theta_j$ and j announces $\theta_i < \theta_j$, and similarly agent j must still be ranked at position 1 when she announces $\theta_j > \theta_i$ and agent i announces $\theta_i > \theta_j$. As two agents cannot occupy the same position in the ranking, this contradiction shows that there is no ex-post incentive compatible and efficient mechanism relying on self-reports. Notice that this impossibility result stems from the fact that the planner is restricted by the set of possible outcomes and can only select a ranking of the members of the community. If he could impose any other punishment (for example by excluding agents who provide inconsistent reports, he could implement an ex-post efficient mechanism in dominant strategies as in (Crémer and McLean, 1985) for any network architecture.

The construction of an ex-post incentive compatible and efficient mechanism when all links are supported is also very intuitive. First consider a comparison between i and j which is observed by at least three agents. The mechanism disregards the report of any agent who deviates from the reports of all other agents. Hence no agent can unilaterally change the outcome of the mechanism when all other agents report the truth. Next suppose that the comparison between i and j is dictated by a third party, a common friend k of i and j. There is no change in reports which could improve the rank of k given that all other individuals tell the truth and that the social network is completely informative. The reason is the following: if the change in report creates an inconsistency in the ranking, the planner can detect if a single individual has cheated and punish him by ranking him at the worst position in the ranking. If the change in report does not create a violation in transitivity, because the social network is completely informative, individual k can be ranked with any other individual in the community. Hence, the rank of individual k is fixed, and no change in report can improve the position of individual k in the ranking. This "friend-based" targeting mechanism is ex-post incentive compatible and efficient.

Theorem 1 characterizes societies for which friend-based targeting mechanisms can be constructed. The social network need not be complete, but must allow all links to be supported. Our next result characterizes the *sparsest* networks for which the condition of Theorem 1 holds. This characterization is based on the "friendship theorem" due to Erdős, Rényi, and Sós (1966).

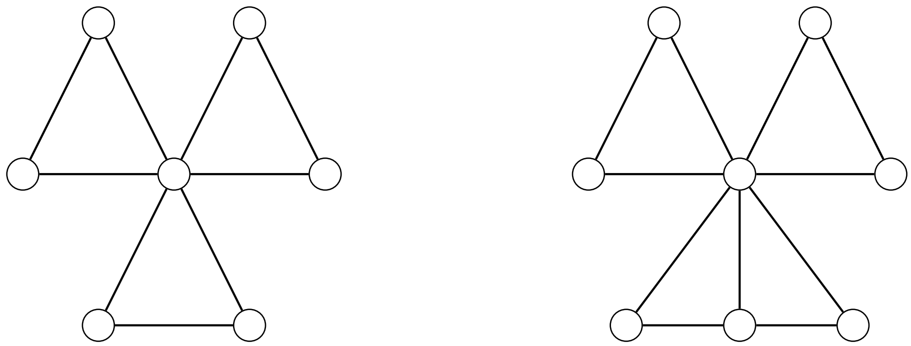
Theorem 2. *(The "friendship theorem") If G is a graph of order n in which any two vertices i and j have one neighbor in common, then $n = 2m+1$ and G contains m triangles which are connected at a common vertex.*

The "friendship theorem", initially stated and proved in (Erdős, Rényi, and Sós, 1966), asserts that in any society where every pair of agents has exactly one friend in common, this friend must be the same for all agents. Different proofs of the friendship theorem have been proposed, often using complex combinatorial arguments (see Wilf (1971), Longyear and Parsons (1972), and Huneke (2002)). The "friendship graph" is illustrated in Figure 2 for $n = 7$. For obvious reasons, it is also termed the "windmill graph". There are exactly $3m$ edges in the friendship graph. Our next theorem shows that this is actually the smallest number of edges for which a completely informative mechanism can be constructed when n is odd. When n is even, the graph which minimizes the number of edges is a variation of the friendship graph, where one of the sails of the windmill contains three vertices, as illustrated in Figure 2 for $n = 8$.

Theorem 3. *Suppose that $n \geq 3$. Let g be a social network for which friend-based targeting generates a complete ranking. Then g must contain at least $\frac{3n}{2} - 1$ links if n is even and $\frac{3(n-1)}{2}$ links is n is odd. If n is odd, the unique sparsest network architecture is the friendship network. If n is even, the unique sparsest network architecture is a modified windmill network where one of the wings contains three nodes i, j, k such that i, j and k are connected to the hub, i is connected to j and j is connected to k.*

Proof. We establish the Theorem through a sequence of claims. Let $\ell(g)$ be the number of links in the social network g.

Claim 1. *If the social network is completely informative, then $d_i \geq 2$ for all $i \in N$.*



(i) $n = 7$. For odd number of nodes, the windmill is also called a friendship graph.

(ii) $n = 8$. For even number of nodes the windmill is modified and one sail has three nodes.

Figure 2: Windmill graphs

Proof of the Claim: As g is connected, $d_i \geq 1$ for all $i \in N$. Suppose that $d_i = 1$, and consider the unique neighbor j of i. As $d_i = 1$, there is no $k \neq j$ which is connected to i and can draw a comparison between i and j. Hence the network g is not completely informative, establishing a contradiction.

Claim 2. *If for any (i, j) there exists k such that $g_{ik}g_{jk} = 1$, then $\ell(g) \geq \frac{3(n-1)}{2}$ if n is odd and $\ell(g) \geq \frac{3n}{2} - 1$ if n is even.*

Proof of the Claim: Consider the following problem: For a fixed number of links L, compute the maximal number of pairwise comparisons of neighbors that can be generated by a social network g when all nodes have degree contained in $[2, n-1]$. More precisely, let $(d_1, .., d_n)$ denote the degree sequence of g with the understanding that $d_{i-1} \geq d_i$ for all $i = 1, .., n$. Then consider the problem:

$$\max_{(d_1,\ldots,d_n)} \frac{d_1(d_1-1)}{2} + \frac{d_2(d_2-1)}{2} + ... + \frac{d_n(d_n-1)}{2}$$
$$\text{subject to } 2 \leq d_i \leq n-1 \;\; \forall i \;,$$
$$d_1 + d_2 + ... + d_n = 2L \;.$$

Notice that the objective function $V(d_1, .., d_n) = \frac{d_1(d_1-1)}{2} + \frac{d_2(d_2-1)}{2} + ... + \frac{d_n(d_n-1)}{2}$ is strictly increasing and convex in $(d_1, ..., d_n)$.

Assume first that n is odd. Then pick $L = \frac{3(n-1)}{2}$ and $d_1 = n-1, d_2 = .. = d_n = 2$.

Because V is strictly convex,

$$\begin{aligned} V(n-1,2,...,2) &= \frac{(n-1)(n-2)}{2} + n - 1 \\ &= \frac{n(n-1)}{2} \\ &> V(d_1,...,d_n) \end{aligned}$$

for any $(d_1,..,d_n) \neq (n-1,2,...2)$ such that $d_1+..+d_n = 3(n-1)$ and $d_i \geq 2$ for all i. Now $\frac{n(n-1)}{2}$ is the total number of bilateral comparisons. So, as $V(d_1,..,d_n)$ is strictly increasing in n, the social network g must contain at least $\frac{3(n-1)}{2}$ links for all pairwise comparisons to be constructed.

Assume next that n is even. Pick $L = \frac{3n}{2} - 1$ and $d_1 = n-1, d_2 = 3, d_3 = .. = d_n = 2$. Because V is strictly convex,

$$\begin{aligned} V(n-1,3,2,...,2) &= \frac{(n-1)(n-2)}{2} + 3 + n - 2 \\ &= \frac{n(n-1)}{2} + 1 \\ &> V(d_1,...,d_n) \end{aligned}$$

for any $(d_1,..,d_n) \neq (n-1,3,2,...2)$ such that $d_1+..+d_n = 3n-2$ and $d_i \geq 2$ for all i.

In addition notice that for $L' = \frac{3n}{2} - 2$,

$$\begin{aligned} V(n-2,2,2,...,2) &= \frac{(n-2)(n-3)}{2} + n - 1 \\ &= \frac{(n-2)^2 - n}{2} \\ &> V(d_1,...,d_n) \end{aligned}$$

for any $(d_1,..,d_n) \neq (n-2,2,2,...2)$ such that $d_1+..+d_n = 3n-4$ and $d_i \geq 2$ for all i.

Hence, the maximum of V_i is smaller than $\frac{n(n-1)}{2}$ when $\ell(g) = \frac{3n}{2} - 2$ and greater than $\frac{n(n-1)}{2}$ when $\ell(g) = \frac{3n}{2} - 1$, establishing that the social network g must contain at least $\frac{3n}{2} - 1$ links for all pairwise comparisons to be constructed.

Next we observe that the friendship network and the modified windmill network generate all pairwise comparisons.

Claim 3. *If n is odd, the friendship network containing exactly $\frac{3(n-1)}{2}$ links, generates all pairwise comparisons. If n is even,The windmill with sails of size 2 and one sail of size 3 with an additional link, containing exactly $\frac{3n}{2} - 1$ links, generates all pairwise comparisons.*

Proof of the Claim: The hub of the network, node n_h, provides the bilateral comparisons between all other $(n-1)$ nodes. If n is odd, in any petal (i,j), i provides the bilateral comparison between j and n_h and j provides the bilateral comparison between i and n_h. If n is even, in any sail of size 2, (i,j),i provides the bilateral comparison between j and n_h and j provides the bilateral comparison between i and n_h. In the unique sail of size 3, (i,j,k), i provides the bilateral comparison between j and n_h, j provides the bilateral comparisons between i and n_h and k and n_h and k provides a (redundant) bilateral comparison between j and n_h.

Finally we establish that the friendship network and the modified windmill network are the only network architectures generating all pairwise comparisons with the minimal number of edges.

Claim 4. *If n is odd, the friendship network is the only network with degree sequence $(n-1,2,...,2)$. If n is even, the modified windmill network with $\frac{n}{2}-2$ sails of size 2 and one sail of size 3 with an additional link is the only network with degree sequence $(n-1,3,2,..2)$.*

Proof of the Claim: Let n be odd. Because one node has degree $n-1$, the network is connected and this node is a hub. All other nodes must be connected to the hub, and if they have degree 2, they must be mutually connected to one other node. Let n be even. The same argument shows that all nodes with degree 2 must be connected to the hub and one other node. These nodes are mutually connected except for the petal of size 3, where one node is connected to the two other nodes in the sail.

□

Theorem 3 establishes a lower bound on the number of edges needed to obtain a complete ranking of the community. It also identifies the unique network architecture which reaches this lower bound: a windmill network where one of the nodes, the hub, connects all other nodes which form triangles.[1] This network architecture implies a very unequal distribution of degrees: one node is connected to all other nodes, whereas all other nodes have degree two or three.

3.2 Regular networks

Theorem 3 shows that as soon as the density of the social network is above $\frac{3}{n}$, the windmill provides a completely informative, incentive compatible, and efficient network. Notice that the hub of the windmill is friends with everyone in the community. Although this

[1]We note in passing that the proof of the Theorem, based on characterization of degree sequences, is very different from the known proofs of the Friendship Theorem. Remark however that we do the statement of Theorem 3 is related but different from the statement of the Friendship Theorem.

is possible in rural villages (Basurto, Dupas, and Robinson, 2017), it is unlikely in larger communities. Even if there does exist some individual who is friends with everyone, friend-based targeting would place a large cognitive cost on this individual since he must provide $\frac{(n-1)(n-2)}{2}$ comparisons.

Suppose the planner sets a maximum degree d in the social network—she caps the number of comparisons each individual provides. What is the minimal value of d for which one can construct a social network satisfying the conditions of Theorem 1? To explore this question, we construct a family of graphs called the rook graphs. This family is derived from the Cartesian product of two graphs, which is defined as follows.

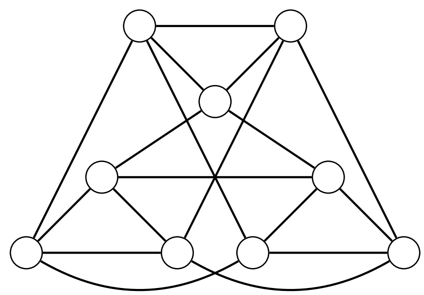


Figure 3: Rook graph for $n = 9$

Cartesian product $g \Box g'$. Given two undirected graphs g and g', the Cartesian product, denoted by symbol $\Box$, is a graph such that the node set is the Cartesian product of the node sets $N(g) \times N(g')$, and any two nodes (i, i'), (j, j') are connected in $g \Box g'$ if and only if $i = j$ and i' is connected to j' in g', or $i' = j'$ and i is connected to j in g.

The Cartesian product of two complete graphs of m nodes $K_m \Box K_m$ is a regular graph on m^2 nodes of degree $2(m-1)$ where every pair of connected nodes has $m-2$ common neighbors and every pair of unconnected nodes has 2 common neighbors (Brouwer and Haemers, 2011). These graphs are called *rook graphs* since the graphs describe the connectivity of the rook chess piece on an $m \times m$ chess board.[2] Figure 3 shows a rook graph for $n = 9$.

For $m \geq 3$ both connected and unconnected pairs of nodes on the rook graph have more than one common neighbor. Thus, by Lemma 1 and Theorem 1, the rook graph is completely informative, incentive compatible, and efficient. For order n, the maximal degree of the rook graph is $d = 2(\sqrt{n} - 1)$. It is an open question whether there exists a family of graphs which improve upon the rook graph.

[2]See Shrikhande (1959) for an early reference.

4 Incomplete rankings and friend-based comparisons

4.1 Comparison networks

We now consider societies where the condition of theorem 1 fails. In these societies, the planner cannot construct an ex-post incentive compatible and efficient mechanism to generate a complete ranking of the agents. Instead, the objective of the planner will be to construct a mechanism which guarantees that a subset of all potential bilateral comparisons will be obtained. More precisely, if the society is not completely informative, there exist some type profiles $\mathbf{T}$ for which a complete ranking of the individuals cannot be constructed. In this situation, we let $i \bowtie_{\mathbf{T}} j$ denote the fact that i and j cannot be compared using the information contained in $\mathbf{T}$. As the mechanism ρ defines a complete ranking, it must then choose an arbitrary ranking between i and j at $\mathbf{T}$. We first define a condition on the mechanism ρ guaranteeing that the arbitrary ranking chosen between i and j remains the same independently of the type profile $\mathbf{T}$.

Independence: The ranking ρ satisfies independence if for any two type profiles $\mathbf{T}$ and $\mathbf{T}'$ such that $i \bowtie_{\mathbf{T}} j$ and $i \bowtie_{\mathbf{T}'} j$, then $\rho_i(\mathbf{T}) > \rho_j(\mathbf{T}) \Leftrightarrow \rho_i(\mathbf{T}') > \rho_j(\mathbf{T}')$.

When the independence condition is satisfied, if the comparison between some pair of agents i and j relies on self-reports, the mechanism ρ cannot simultaneously satisfy ex-post incentive compatibility and efficiency.

Proposition 1. *Suppose that there exists a pair of agents (i, j) such that $g_{ij} = 1$ but there is no agent k such that $g_{ik}g_{jk} = 1$. Then there exists no mechanism satisfying independence, ex-post incentive compatibility and efficiency.*

Proof. As in the necessity part of the proof of Proposition 1, consider a realization of the characteristics such that θ_i and θ_j are the two lowest characteristics. Fix two type profiles $\mathbf{T}$ and $\mathbf{T}'$ which agree on all comparisons except that $\theta_i < \theta_j$ in $\mathbf{T}$ and $\theta_j < \theta_i$ in $\mathbf{T}'$. Clearly, for any $k \neq i, j$, if $k \succ_{\mathbf{T}} i$ then $k \succ_{\mathbf{T}} j$, as i and j can be compared under $\mathbf{T}$. Similarly, if $k \succ_{\mathbf{T}'} i$ then $k \succ_{\mathbf{T}'} j$. Furthermore, as θ_i and θ_j are the two smallest characteristics, all agents k which can be compared to i and j have higher rank than i and j. Next consider $k \neq i, j$ such that $k \bowtie_{\mathbf{T}} i$. Then we must also have $k \bowtie_{\mathbf{T}} j$, as otherwise $k \succ_{\mathbf{T}} j$ which implies $k \succ_{\mathbf{T}} i$. Similarly, if $k \bowtie_{\mathbf{T}'} j$ then $k \bowtie_{\mathbf{T}'} i$. Hence if an individual k cannot be compared to i under $\mathbf{T}$, it cannot be compared to j under $\mathbf{T}$, nor to j under $\mathbf{T}'$ nor to i under $\mathbf{T}'$.

By ex-post efficiency, for all k which can be compared to i, j, $\rho_k(\mathbf{T}) > \rho_j(\mathbf{T}) > \rho_i(\mathbf{T})$. Similarly, by ex-post efficiency, for all k which can be compared to i, j, $\rho_k(\mathbf{T}') > \rho_i(\mathbf{T}) > \rho_j(\mathbf{T})$. By independence, for all k which cannot be compared to i, j, $\rho_k(\mathbf{T}) > \rho_i(\mathbf{T})$ if and only if $\rho_k(\mathbf{T}') > \rho_i(\mathbf{T}')$ and $\rho_k(\mathbf{T}) > \rho_j(\mathbf{T})$ if and only if $\rho_k(\mathbf{T}') > \rho_j(\mathbf{T}')$. Hence the set of agents who are incomparable to i, j and are ranked below i and j under $\mathbf{T}$ and $\mathbf{T}'$ are identical. But this implies that

$$\begin{aligned}\rho_i(\mathbf{T}) &= \rho_i(\mathbf{T}') - 1,\\ \rho_j(\mathbf{T}') &= \rho_j(\mathbf{T}) - 1\end{aligned}$$

As we also have $\rho_i(\mathbf{T}) < \rho_j(\mathbf{T})$ and $\rho_j(\mathbf{T}') < \rho_i(\mathbf{T}')$, we must have

$$\rho_i(\mathbf{T} + 2 > \rho_j(\mathbf{T}) > \rho_i(\mathbf{T}),$$

so that $\rho_j(\mathbf{T}) = \rho_i(\mathbf{T}) + 1$. Hence

$$\begin{aligned}\rho_i(\mathbf{T}) &= \rho_j(\mathbf{T}'),\\ \rho_j(\mathbf{T}) &= \rho_i(\mathbf{T}').\end{aligned}$$

Because there are only two announcements t^i_{ij} and t^j_{ij} on the link (i, j), ex-post incentive compatibility requires that agents i and j cannot improve their ranking by changing their reports on the link (i, j). Let T_{-ij} denote the announcements on all links but link ij. We must have

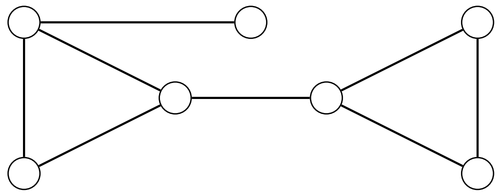
$$\begin{aligned}\rho_i(T_{-ij}, t^i_{ij} = 1, t^j_{ij} = -1) &= \rho_i(\mathbf{T})\\ \rho_j(T_{-ij}, t^i_{ij} = 1, t^j_{ij} = -1) &= \rho_j(\mathbf{T}'),\end{aligned}$$

resulting in a contradiction as i and j cannot both be ranked at the same position. □

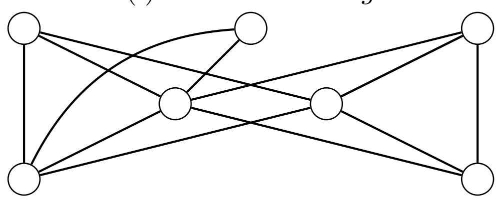
Proposition 1 extends the necessity argument of theorem 1 to show that the planner cannot construct an ex-post incentive compatible and efficient mechanism when two agents provide self-reports. Hence comparisons based on self-reports cannot be elicited by the planner. This implies that, when some links (i, j) are not supported, the planner cannot expect to construct a meaningful comparison between i and j. The planner will only be able to construct a subset of all possible bilateral comparisons. By Proposition 1, in characterizing the set of comparisons that the planner can obtain, we will ignore comparisons based on self-reports, and focus instead on friend-based comparisons.

To this end, we first modify the type of agent i, T_i by removing any comparison t^i_{ij} which is not supported by a third agent, i.e. we let $t^i_{ij} = 0$ if there exists no $k \neq i, j$ such that $t^k_{ij} \neq 0$. We now search for mechanisms satisfying independence, ex-post incentive compatibility and efficiency in the modified society where self-comparisons are ignored. Second, we construct a *comparison network* h which captures all comparisons that can be obtained using friend-based comparisons. Formally, we let $h_{ij} = 1$ if and only if there exists

k such that $g_{ik}g_{jk} = 1$. The network h collects all pairs of agents which can be compared by a third agent. It differs from the social network g in two ways: (i) pairs of agents which are linked in g but do not have a common friend appear in g but not in h, (ii) pairs of agents which are not directly linked in g but have a common friend appear in h but not in g. Hence the comparison network h may be denser or sparser than the social network g. Figure 4 illustrates a social network g and the corresponding comparison network h.



(i) Social network g



(ii) Comparison network h

Figure 4: Social and comparison networks

The comparison network h is the minimal set of comparisons that the planner can guarantee for any possible realization of the characteristics. For some realizations, the planner may complement the basic comparisons contained in h by taking the transitive closure of these basic bilateral comparisons. In the end, the planner will be able to construct a partial order on the individuals, which will need to be completed by an arbitrary ranking of incomparable agents. If the conditions of Theorem 1 hold, the comparison network h is the complete network. When the conditions fail, our objective is to characterize all comparison networks which can be supported by a mechanism satisfying independence, ex-post incentive compatibility and efficiency.

4.2 Connected comparison networks and bipartite social networks

We first provide a characterization of social networks which generate connected comparison networks.

Proposition 2. *Suppose that $n \geq 3$. The comparison network h is connected if and only if g is not bipartite.*

Proof. We first prove the following Claim.

Claim 5. *The comparison network is connected if and only if for all $i, j \in N$, there exists an even walk between i and j.*

Proof of the Claim Suppose first that h is connected. Pick any two nodes $i, j \in N$ and a walk $i = i^0, ..., i^m = j$ in h. By definition, for any (i^k, i^{k+1}) in the walk, there exists $j^k \in N$ such that $i^k, i^{k+1} \in N_{j^k}$. But this implies that there exists a walk in g connecting i to j given by $i^0, j^0, i^1, j^1, ..., i^{m-1}, j^{m-1}, i^m$.[3] This walk contains an even number of edges, proving necessity of the claim.

Next suppose that h is not connected and let i and j be two nodes in different components of h. We want to show that all walks between i and j in g are odd. Consider first a path between i and j. If the path is even, there exists a sequence of nodes $i = i^0, i^1, .., i^m = j$ where $m = 2l$ is even such that $g_{i^k, i^{k+1}} = 1$ for all k. But then, for any $l = 0, \frac{m}{2} - 1$, $h_{i^{2l}, i^{2l+1}} = 1$, and hence there exists a path $i = i^0, i^2, ..., i^m = j \in f$, contradicting the fact that i and j belong to two different components in h. Hence all paths between i and j are odd. If there exists an even walk between i and j in g, it must thus involve an odd cycle starting at i or starting at j. Without loss of generality, suppose that there exists an odd cycle starting at i, $i = i^0, ..., i^m = i$, where $m = 2l + 1$ is odd. Consider any even path between i and j, where we index $i = i^m, ...i^r = j$ and $r = 2p$ is even. We construct a path in h between i and j as follows. Because m is odd, we first construct the sequence of connected nodes in h, $i = i^{2l+1}, i^{2l-1}, i^{2l-3}, .., i^1$. Because i is connected in g both to i^1 and i^{m+1}, we then link i^1 to i^{m+1} in h. Now $m + 1$ is even, so we can use the path between i and j to construct a sequence $i^{m+1}, .., i^r = j$ in h. Concatenating the two sequences, we construct a sequence $i, i^{m-2}, ..., i^1, i^{m+1}, .., i^r = j$ in h, contradicting the fact that i and j belong to two different components in h. Hence if h is not connected, there exists a pair of nodes i, j such that all walks between i and j are odd, proving the necessity of the claim.

We now prove the second claim

Claim 6. *For all $i, j \in N$ there exists an even walk between i and j if and only if g is not bipartite.*

Proof of the Claim: Suppose that g is bipartite with sets A and B. As $N \geq 3$, at least one of the two sets has more than one element. Pick i, j such that $i \in A$ and $j \in B$, then we claim that all walks between i and j must be odd. Any walk between i and j must contain an even number of edges alternating between nodes in A and B and a single edge between a node in A and a node in B. Hence the total number of edges must be odd, proving the necessity of the claim.

[3]Note that this walk is not necessarily a path even if the initial walk in h is a path, as the same node j^k can be used several times in the walk.

Conversely, suppose that there exists a pair of nodes i, j such that all walks between i and j are odd. Consider the sets of nodes $A = \{k|\delta(i,k) \text{ is even}\}$ and $B = \{k|\delta(i,k) \text{ is odd }\}$, where $\delta(i,k)$ denotes the geodesic distance between i and k in the graph. We first claim that if $k \in A$, *all walks between i and k must be even*. Suppose not, then there exist two different walks between i and k, one w_1 which is even (the shortest path between i and k) and one w_2 which is odd. Pick one particular path p between k and j. If this path is odd, then the walk between i and j containing w_2 followed by p is even, contradicting the assumption. If the path is even, then the walk between i and j containing w_1 followed by p is even, contradicting the assumption again. Hence all walks between i and nodes in A are even and all walks between i and nodes in B are odd. Next notice that there cannot be any edge between nodes in A. Suppose by contradiction that there exists an edge between k and l in A, and consider a walk between i and k, w_1 followed by the edge kl. This forms an odd walk between i and l, contradicting the fact that all walks between i and l must be even. Hence, there is no edge between nodes in A and similarly no edge between nodes in B, showing that the graph g is bipartite.

□

Proposition 2 establishes that the network h is connected if and only if the social network g is not bipartite. If the network g is bipartite, and the nodes partitioned into the two sets A and B, the comparison network h is disconnected into two components: individuals in A rank individuals in B and individuals in B rank individuals in A. In that situation, there is no realization of the characteristics for which the society can reach a complete ranking. Individuals can be ranked inside the two sets A and B but rankings of individuals across the two sets must be arbitrary. Notice however that an agent in A cannot improve his ranking by lying about the ranking of agents in B. Hence, when the social network g is bipartite (and the comparison network h disconnected), it is easy to construct a mechanism satisfying independence, ex-post incentive compatibility and efficiency.

Proposition 3. *Suppose that the social network g is bipartite with two sets of nodes A and B. Then there exists a mechanism satisfying independence, ex-post incentive compatibility and efficiency, which generates a ranking which coincides with the comparison network h on its two components A and B.*

Proof. Consider a mechanism where all agents in A are ranked above agents in B. For any two agents i and j in A, let $\rho(i) > \rho(j)$ if $i \succ_{\mathbf{T}} j$. If $i \bowtie_{\mathbf{T}} j$ or if the reports on i and j are incompatible, construct an arbitrary ranking by letting $\rho(i) > \rho(j)$ if and only if $i > j$. Similarly, for any two agents i and j in B, let $\rho(i) > \rho(j)$ if $i \succ_{\mathbf{T}} j$. If $i \bowtie_{\mathbf{T}} j$ or if the reports on i and j are incompatible, let $\rho(i) > \rho(j)$ if and only if $i > j$.

We will show that the mechanism satisfies the three properties. Clearly if i and j are incomparable under two profile types $\mathbf{T}$ and $\mathbf{T}'$, either one belongs to A and the other to B (in which case the mechanism ranks them in the same way under $\mathbf{T}$ and $\mathbf{T}'$), or they belong to the same set, and the mechanism ranks them identically under $\mathbf{T}$ and $\mathbf{T}'$ as it only uses the index to rank them. Hence independence is satisfied.

The mechanism satisfies strategy-proofness, a stronger incentive compatibility notion than ex-post incentive compatibility. Consider an agent i in A. Then we claim that if $t^i_{jk} \neq 0$ it must be that both j and k are in B. To see this notice that as g is bipartite it does not contain any triangle. Hence no self-report can be supported by a third agent, and hence $t^i_{ij} = 0$ for all $j \neq i$. The only case where $t^i_{jk} \neq 0$ is thus when $g_{ij}g_{ik} = 1$ and $j, k \in B$. Hence, by changing his report t^i_{jk}, agent i can only affect the ranking of agents in B. As all agents in B are ranked below agents in A, this does not affect the rank of agent i, and hence agent i's ranking is independent of his announcement, proving that the mechanism is strategy-proof.

Finally, notice that by construction, the mechanism ρ achieves an ex-post efficient ranking separately on each of the two components A and B. by Proposition 2, the comparison network h is disconnected into two components A and B. Hence the mechanism ρ is also ex-post efficient. □

Proposition 3 characterizes one situation where the planner can elicit information about bilateral comparisons: when the set of individuals in the community can be partitioned into two subsets where members of one subset observe members of the other subset. This suggests the following design to extract information about individual characteristics in villages in developing economies. One could survey separately men and women and ask men about the characteristics of women and women about the characteristics of men. This would provide a complete ranking among all men and among all women. However, this design would not allow the planner to obtain information about the ranking of agents across the two sets. The mechanism completes the partial ranking by an arbitrary ranking across individuals in the two sets, possibly resulting in a final ranking which is very different from the true ranking.

4.3 Social quilts and mechanism design in incomplete societies

We now consider societies for which the comparison network h is connected. Our first result provides a sufficient condition under which the planner can construct a mechanism satisfying independence, ex-post incentive compatibility and efficiency. The mechanism is an extension of the mechanism constructed in the proof of Proposition 1 for completely informative societies.

Proposition 4. *Suppose that all links in g are supported (for all i, j such that $g_{ij} = 1$, there exists a k such that $g_{ik}g_{jk} = 1$). Then there exists a mechanism satisfying independence, ex-post incentive compatibility and efficiency.*

Proof. We consider the same mechanism as in the proof of Theorem 1: We define the mechanism ρ by constructing bilateral comparisons. Let r_{ij} denote the bilateral comparison between i and j chosen by the planner.

First consider a pair of agents (i,j) who observe each other, $g_{ij} = 1$. By assumption, there are at least three reports on the ranking of i and j. If all agents transmit the same report on (i,j), let $r_{ij} = t_{ij}$. If all agents but one transmit the same report t_{ij} and one agent reports $t'^{ij} = -t_{ij}$, ignore the ranking t'_{ij} and let $r_{ij} = t_{ij}$. In all other cases, let $r_{ij} = 1$ if and only if $i > j$.

Second consider a pair of agents (i,j) who do not observe each other, $g_{ij} = 0$. By assumption, there exists at least one agent k who observes them both. If there are at least three agents who observe i and j, use as above a mechanism such that $r_{ij} = t_{ij}$ if all agents agree on t_{ij} or only one agent chooses $t'_{ij} = -t_{ij}$, and let $r_{ij} = 1$ if $i > j$ otherwise.

If one or two agents observe i, j, pick the agent k with the highest index. Consider the vector of announcements $\tilde{T}^{-k}$ where one disregards the announcements of agent k. Let $\succ_{\tilde{T}^{-k}}$ be the binary creation created by letting $t_{ij} = 1$ if and only if $t^l_{ij} = 1$ for all $l \neq k$. If there exists a directed path of length greater or equal to 2 between i and j in $\succ_{\tilde{T}^{-k}}$, and for all directed paths between i and j in $\succ_{\tilde{T}^{-k}}$, $i^0, .., i^L$ we have $t_{i^l i^{l+1}} = 1$, then $r_{ij} = 1$. If on the other hand for all directed paths between i and j in $\succ_{\tilde{T}^{-k}}$, $t_{i^l i^{l+1}} = -1$, then $r_{ij} = -1$. In all other cases, let agent k dictate the bilateral comparison between i and j, $r_{ij} = t^k_{ij}$.

Now consider all bilateral comparisons r_{ij}. If they induce a transitive binary relation on N, let ρ be the complete order generated by the bilateral comparisons. Otherwise, consider all shortest cycles generated by the binary relation r_{ij}. If there exists a single agent i who dictates at least two bilateral comparisons in all shortest cycles, agent i is punished by setting $\rho_i = 1$ and $\rho_j > \rho_k$ if and only if $j > k$ for all $j, k \neq i$. If this is not the case, pick the arbitrary ranking where $\rho_i > \rho_j$ if and only if $i > j$.

We now prove that this mechanism satisfies all three conditions. Consider two type profiles $\mathbf{T}$ and $\mathbf{T}'$, and two agents i and j such that $i \bowtie_{\mathbf{T}} j$ and $i \bowtie_{\mathbf{T}'} j$. Because $\mathbf{T}$ and $\mathbf{T}'$ generate identical truthful reports and result in a transitive partial order, i and j must be ranked at the final completion phase of the mechanism, using the same ranking $\rho(i) > \rho(j)$ if and only if $i > j$. Hence independence holds.

Next consider any pair (i,j) such that $i \succ_{\mathbf{T}} j$. There must exist a sequence of bilateral comparisons $(i, i^1, .., i^t, .., i^T, j)$ such that $h_{i^{t-1} i^t} = 1$ and $i^{t-1} \succ_{\mathbf{T}} i^t$. For any of these pairs, we must have $r_{i^{t-1} i^t} = 1$ and hence, because the announcement $\mathbf{T}$ generates a transitive partial order, $\rho(i^{t-1}) > \rho(i^t)$. But this implies that $\rho(i) > \rho(j)$, establishing that the mechanism satisfies ex-post efficiency.

Finally, we show that the mechanism is ex-post incentive compatible. Consider agent k's incentive to change his announcement on a link ij when all other agents tell the truth. If the link ij is supported, this change does not affect the outcome of the mechanism. So consider an unsupported link ij and let agent k be the highest index agent observing i and

j. Suppose that all agents $l \neq k$ announce the truth, so that $\succ_{\tilde{T}^{-k}} = \succ_{T^{-k}}$. We first show that individual k cannot gain by making an announcement which generates cycles in the ranking r_{ij}.

Suppose that the binary relation r_{ij} exhibits a cycle $i^0 i^1 ... i^L$

By the same argument as in the proof of Theorem 1, agent k must dictate at least two bilateral comparisons in the cycle. We will show that the initial cycle must contain a cycle of length 3. Suppose that the initial cycle has length greater than or equal to 4. Let ij and lm be two bilateral comparisons dictated by agent k. Suppose first that the rank of l is strictly higher than the rank of j. As individual k observes both (i, j) and (l, m), he also observes both i and l. Hence i and l must be compared under r and either $r_{il} = 1$ or $r_{il} = -1$. Now if $r_{il} = -1$, one can construct a shorter cycle by replacing the path $lm..i$ by the path li. If $r_{il} = +1$, one can construct a shorter cycle by replacing the path $ij, , l$ by the path il. Next suppose that $j = l$ so that the two comparisons (i, j) and (l, m) are adjacent in the cycle. Again because individual k observes both i and m, then i and m must be compared under r and either $r_{im} = +1$ or $r_{im} = -1$. If $r_{im} = +1$, one can construct a shorter cycle by replacing ijm with im. If $r_{im} = -1$, one can construct a cycle of length 3 $ijmi$.

We conclude that if the binary relation r exhibits a cycle, there must exist a subcycle of length 3, so that all shortest cycles are of length 3. Furthermore, individual k must dictate at least two of the bilateral comparisons in all cycles. Hence, agent individual k has no incentive to make an announcement generating a cycle in r, as he will be punished and obtain the lowest rank

We finally assume that the ranking generated by r is acyclic and show that the rank of agent k must remain the same if he changes his report on any pair (i, j). To this end, let $J = (i_1, j_1), ..., (i_l, j_l), ..(i_L, j_L)$ be the pairs on which k is a dictator and let J^+ denote the set of pairs (i_l, j_l) such that $k succ_{T^{-k}} i_l, j_l$ and J^- the set of pairs such that $k \prec_{T^{-k}} i_l, j_l$. Let $\mathbf{T}' = (T^{-k}, T'^k)$ the announcement obtained when i changes his report on some of the pairs in J while keeping a transitive partial order. For any m such that $k \succ_{T-k} m$, $k \succ_{\mathbf{T}} m$ and $k \succ_{\mathbf{T}'} m$. Hence $\rho_k(\mathbf{T}) > \rho_m(\mathbf{T})$ and $\rho_k(\mathbf{T}') > \rho_m(\mathbf{T}')$. Similarly, for any m such that $k \prec_{T-k} m$, $k \prec_{\mathbf{T}} m$ and $k \prec_{\mathbf{T}'} m$. Hence $\rho_k(\mathbf{T}) < \rho_m(\mathbf{T})$ and $\rho_k(\mathbf{T}') < \rho_m(\mathbf{T}')$. We also have, for any $m \in J^+$, $\rho_k(\mathbf{T}) > \rho_m(\mathbf{T})$ and $\rho_k(\mathbf{T}') > \rho_m(\mathbf{T}')$. For any $m \in J^-$, $\rho_k(\mathbf{T}) < \rho_m(\mathbf{T})$ and $\rho_k(\mathbf{T}') < \rho_m(\mathbf{T}')$. Next consider m such that $k \bowtie_{T^{-k}} m$ and $m \notin J$. If $m \prec_{T^{-k}} i_l$ for some $i_l \in J^+$, then $k \succ_{\mathbf{T}} m$ and $k \succ_{\mathbf{T}'} m$ so that $\rho_k(\mathbf{T}) > \rho_m(\mathbf{T})$ and $\rho_k(\mathbf{T}') > \rho_m(\mathbf{T}')$. Similarly, if $m \succ_{T^{-k}} i_l$ for some $i_l \in J^-$, then $k \prec_{\mathbf{T}} m$ and $k \prec_{\mathbf{T}'} m$ so that $\rho_k(\mathbf{T}) < \rho_m(\mathbf{T})$ and $\rho_k(\mathbf{T}') < \rho_m(\mathbf{T}')$. Finally, if $m \succ_{T^{-i}} i_l \forall i_l \in J^+, m \prec_{T^{-k}} i_l \forall i_l \in J^-$ and $k \bowtie_{T^{-k}} m$, then $k \bowtie_{\mathbf{T}} m$ and $k \bowtie_{\mathbf{T}'} m$. Whenever $k \bowtie_{\mathbf{T}} m$ and $k \bowtie_{\mathbf{T}'} m$, then the ranking between k and m is independent of the type profile. Hence, in all cases the ranking between k and m is identical under $\mathbf{T}$ and $\mathbf{T}'$. This argument completes the proof that the mechanism satisfies ex-post incentive compatibility. □

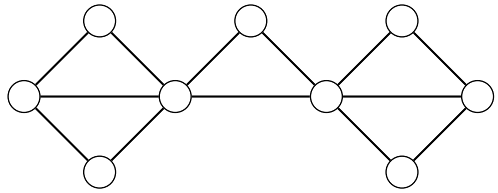


Figure 5: A supported social network g

Proposition 4 identifies social networks which allow the planner to construct an incomplete ranking of the agents: all links must be supported and the social network is thus formed of a collection of triangles. Following Jackson, Rodriguez-Barraquer, and Tan (2012), we call these societies "social quilts". Figure 5 illustrates one of these networks. Notice that some comparisons are supported as links in the triangles, and other comparisons are supported as links across triangles. Whether there exist other social networks g generating connected comparison networks h for which the planner can construct a mechanism satisfying independence, ex-post incentive compatibility and ex-post efficiency remains an open question. However, there are clearly social networks for which the planner will not be able to construct a mechanism satisfying these three properties using friend-based comparisons, as shown in the following example.

Example 1. Let $n = 4$. Agents i, j, k are connected in a triangle and agent l is connected to i. △

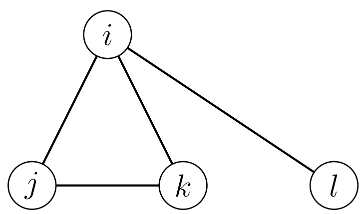


Figure 6: A social network g where a mechanism does not exist

In this example, the links $(i, j), (i, k), (j, k)$ are supported, but the friend-based links (j, l) and (k, l) are not supported. Consider a realization of the characteristics such that $\theta_l > \theta_j > \theta_i > \theta_k$. If agent i announces $\theta_l > \theta_j > \theta_k$, by ex-post efficiency, the planner constructs the rankings k, i, j, l and the rank of agent i must be equal to 2. If on the other hand agent i announces $\theta_j > \theta_k > \theta_l$, the planner constructs the ranking l, k, i, j and the ranking of agent i is now equal to 3. Hence agent i has an incentive to lie and announce $\theta_j > \theta_k > \theta_l$.

In Example 1, the planner's ranking of i depends on his announcement on the rankings (j, l) and (k, l). Given that $\theta_j > \theta_i > \theta_k$, and that (i, j, k) form a triangle, the planner must rank i between j and k. Hence she cannot rank all three agents j, k and l on the same side of agent i, as in the mechanism constructed in the proof of Proposition 4. But then, the announcement of agent i on (j, k, l), by changing the rank of l with respect to j and k, will also affect the ranking of agent i. Because agent i can manipulate his rank by his announcements on the unsupported links (j, k) and (j, l), there is no mechanism satisfying ex-post incentive compatibility and efficiency in this society.

5 Real-life social networks

We use social network data from India (Banerjee et al., 2013) and Indonesia (Alatas et al., 2016) to highlight three implications of our analytical results. Firstly, the success of friend-based targeting depends on the structure of the social network and not simply on the number of links. We show that for a given density there is large variation in information. Secondly, triangles are important for ensuring incentive compatibility. Networks with many triangles have a high share of incentive compatible comparisons. Thirdly, capping the number of comparisons per individual reduces information considerably. The capped information is bounded to a range that tightens as the community size increases.

The data from India and Indonesia is particularly useful because it contains multiple independent networks: 75 villages from Karnataka, India and 622 neighborhoods from three provinces in Indonesia. We focus on the giant component of each network. Table 1 provides summary statistics of the networks. Indonesian networks are smaller and denser than the Indian networks. We report the mean, minimum, and maximum for each measure. The combined sample of networks provides a large range in network size and structure.

Table 1: Summary statistics of social networks

	India		**Indonesia**	
Networks	75		622	
Number of households	198.72	[77, 356]	52.85	[11, 263]
Share in giant component	.95	[.85, .99]	.65	[.22, 1.00]
Average degree	9.34	[6.82, 13.83]	17.96	[2.00, 218.00]
Density	.05	[.02, .12]	.53	[.10, 1.00]
Average clustering	.26	[.16, .45]	.82	[.48, 1.00]
Average distance	2.75	[2.30, 3.32]	1.77	[1.00, 4.32]
Information	.37	[.18, .62]	.78	[.25, 1.00]

Notes: Means are reported with minimum and maximum in brackets. Information is measured by the density of the comparison network. All statistics (except the number of households) are calculated on the giant component. Data is sourced from Banerjee et al. (2013) for India and Alatas et al. (2016) for Indonesia.

We measure information using the density of the comparison network. In appendix A, we show that this measure is almost perfectly correlated with counting spanning trees—a measure with strong theoretical foundations that corresponds to the D-criterion from the optimal design of experiments Pukelsheim (1993). We prefer the density of the comparison network as it is easier to interpret. The density of the comparison network is simply the count of unique comparisons as a share of the $\frac{n(n-1)}{2}$ possible comparisons.

5.1 Large variation of information for a given density

Dense social networks provide many comparisons, but density is not necessary for information. The windmill, from Theorem 3, is completely informative and the density is just $\frac{3}{n}$. This insight is also present in the data. In the bottom left panel of Figure 7, we plot the density against information for social networks of more than 50 households. We highlight two networks and plot their corresponding network diagrams. The orange network, from India, has 75 nodes and a density of 0.12. The green network, from Indonesia, has 69 nodes and the same density of 0.12. Despite having equal density, the orange network provides information of 0.62—nearly double the green network with 0.33.

Two factors contribute to the superior information in the orange network. Firstly, as shown in the top right panel of Figure 7 the degree distribution of the orange is spread more widely than the green. The number of comparisons provided by a single node is a convex function of degree. For a given density, the greater the spread in the degree distribution, the more comparisons the social network provides.

Secondly, the green network is a combination of cliques that are weakly connected to each other. Cliques repeat comparisons. Take a clique of 7 nodes as an example. These seven nodes provide 105 comparisons yet 84 of these comparisons are repeated. Since the green and orange networks have a similar number of nodes and equal density, they produce a similar total number of comparisons. The difference is that a greater share of the orange network's comparisons are repeated. This example shows that the success friend-based targeting depends not only on the number of links, but, more importantly, how those links are structured.

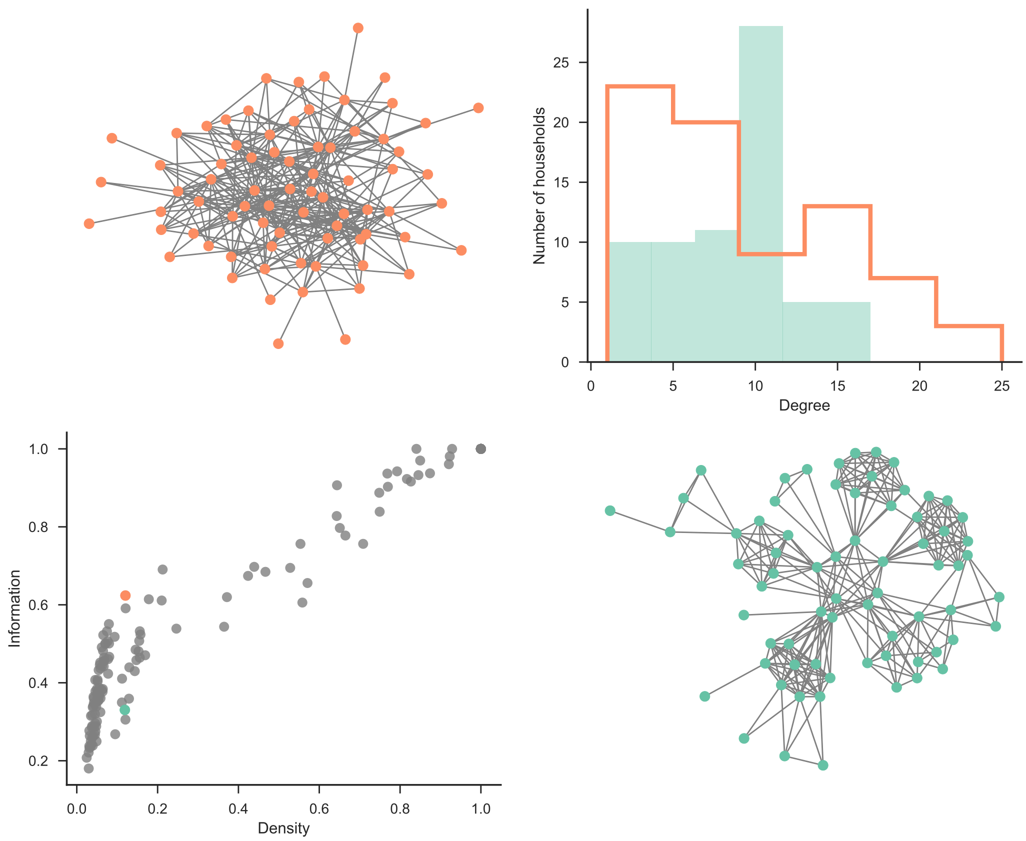


Figure 7: Large variation of information for a given density

Note: The bottom left panel shows a scatter-plot of information and density for networks of more than 50 households. Two networks of similar density are highlighted by orange and green points on the scatter plots. The network diagrams corresponding to those two points are plotted in the top left and bottom right panel. The degree distribution of the highlighted networks is shown in the top right panel.

5.2 Triangles and incentive compatibility

Even if the social network provides many unique comparisons, the social planner needs to ensure that the comparisons are truthful. Proposition 4 showed that triangles ensure incentive compatibility. An individual i may have an incentive to lie about a friend j if the friendship is not supported—there is no k who is friends with both i and j.

Proposition 4 shows the triangles provide incentive compatibility in two ways. Both the within triangle and across triangle comparisons are incentive compatible. Consider a simple example of a windmill on 5 nodes, shown in Figure 8. There are several within triangle comparisons, such as (m, n), but i also provides across triangle comparisons, such as (m, x). For a given network we can decompose information into comparisons provided within and across triangles.

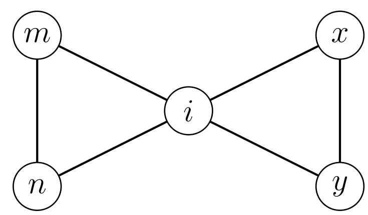


Figure 8: Dutch windmill graph for $n = 5$

We approach the decomposition by removing all unsupported links from the network and recalculating information. The resulting supported network is incentive compatible but information is reduced—from 0.37 to 0.27 on average for India and from 0.78 to 0.75 on average for Indonesia. A greater share of the Indonesian links are supported, which is a product of denser and more clustered networks.

Then we decompose the supported comparisons into within and across triangles. If a comparison is repeated and provided both within and across a triangle, we categorize the comparison as within. In Indonesia, where the networks are denser, on average 70 percent of the of the comparisons are within triangles. In the sparser Indian networks on average only 19 percent of the comparisons are within triangles.

The importance of support for incentive compatibility of friend-based targeting echoes the results of Jackson, Rodriguez-Barraquer, and Tan (2012) who study favor exchange.

5.3 Capping comparisons

In Section 3.2 we analyzed networks with a cap on degree. We focused on completely informative networks and noted that rook graphs are completely informative for degree $d = 2(\sqrt{n} - 1)$. Below the degree of the rook graphs the network may simply lack the necessary comparisons to provide complete information.

Suppose no comparison is repeated and every individual has at most d friends. This social network provides $n\frac{d(d-1)}{2}$ comparisons. Therefore, for maximum degree d a simple upper bound on information is $\frac{d(d-1)}{n-1}$.

We may also construct a lower bound by maximizing repetitions. First, construct a size $d+1$ clique. The clique provides $\frac{(d+1)d}{2}$ unique comparisons. Have the remaining $n-d$ nodes provide comparisons only about this clique. The lower bound on information is $\frac{(d+1)d}{n(n-1)}$. The likelihood this lower bound is met exactly is very small since it requires that at least $d+1$ nodes are full connected in the original social network and everyone happens to choose the same friends to compare. The bound could be tightened by making assumptions about the degree distribution of the original social network.

We simulate the process of capping degree by having each individual provide comparisons between at most d friends. The resulting network is directed. Suppose j picks i and i has more than d friends. There is no guarantee i will also pick j.

Unsurprisingly information is reduced by capping degree. It is less intuitive that there is little variation in this reduction. For 100 iterations, the standard deviation of the change in information is less than .01 for any given network in our sample. Each individual samples their d friends from the original social network so the space of resulting networks is restricted.

In Figure 9 we plot the information from the original social network in color (India in orange and Indonesia in green) and then network capped at $d = 5$ in grey. Since network size, n, is on the horizontal axis the capped network is directly below the original network. The large reduction in information is evident.

The lines on the figure plot the upper bound $\overline{I}(d, n) = \frac{d(d-1)}{n-1}$ and lower bound $\underline{I}(d, n) = \frac{d(d+1)}{n(n-1)}$ for $d = 5$. As the number of nodes increasing the space between the bounds tightens and we may use the bounds to predict information in the capped network.

Policy makers may be concerned that information is too low in the capped network to be of any use. In a recent field experiment, peer comparisons with information of just 0.004 were more accurate in predicting income and assets than a collection of proxies from survey data (Hussam, Rigol, and Roth, 2017). Friend-based targeting may be helpful even when the number of unique comparisons is small relative to the community size.

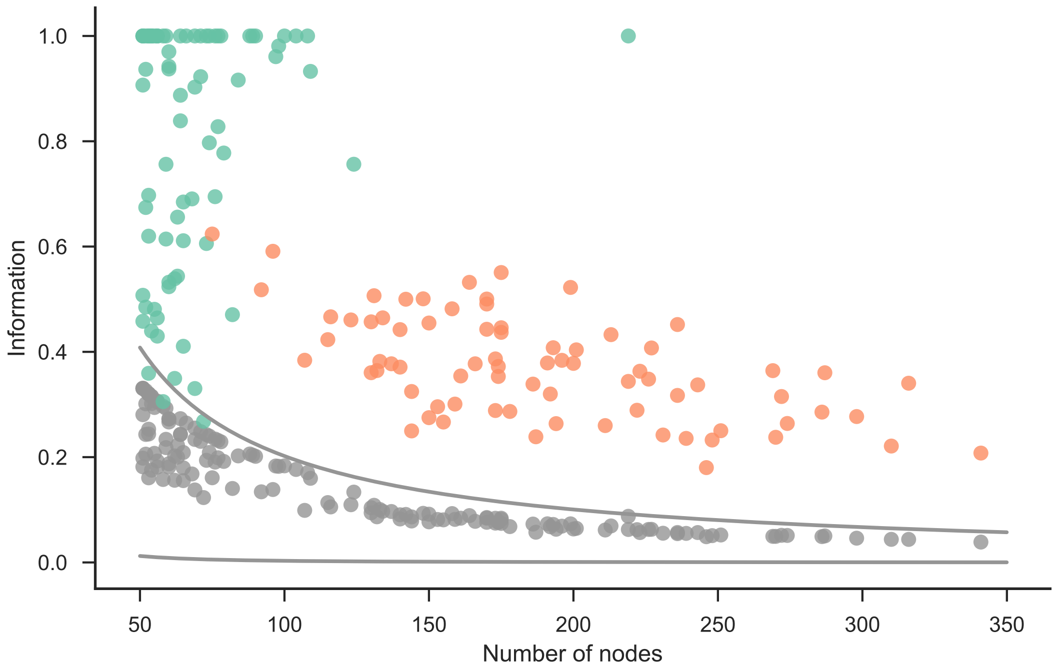


Figure 9: Capping comparisons at 5 friends

Note: The original social network is shown in color (green for Indonesia and orange for India) and the same network capped to a maximum degree of 5 is shown in grey. The number of nodes is constant but information decreases to the region between the upper bound of $\frac{d(d-1)}{n-1}$ and the lower bound of $\frac{(d+1)d}{n(n-1)}$.

6 Robustness and extensions

6.1 Dominant strategy implementation

The following Proposition shows why dominant strategy implementation is too strong in our setting. We first define strategy-proofness for a mechanism:

Strategy-proofness For any agent i, for any vector of announcements $(\hat{T}^{-i})$ and any types T^i, T'^i,

$$\rho_i(T^i, \hat{T}^{-i}) \geq \rho_i(T'^i, \hat{T}^{-i}).$$

Proposition 5. *Let g be a triangle (a three-agent complete network). Then there exists no mechanism satisfying strategy-proofness and ex-post efficiency.*

Proof. We first establish the following simple general claim:

Claim 7. *If ρ is strategy-proof, $\rho_i(T^i, \hat{T}^{-i}) = \rho^i(T'^i, \hat{T}^{-i})$ for all $i, T^i, T'^i, \hat{T}^{-i}$.*

Proof. Suppose by contradiction that there exists $i, T^i, T'^i, \hat{T}^{-i}$ such that $\rho_i(T^i, \hat{T}^{-i}) > \mu^i(T'^i, \hat{T}^{-i})$. Let T'^i be the true type of individual i. Then, individual i has an incentive to announce T^i, contradicting the fact that ρ is strategy-proof. □

Consider next two vectors of types:

- $\mathbf{T}_1$: $t_{ij} = t_{jk} = t_{ik} = 1$
- $\mathbf{T}_2$: $t_{ij} = -1, t_{jk} = 1, t_{ik} = -1$

As the mechanism is ex-post efficient, it must assign ranks $\rho_i(\mathbf{T}_1) = 3, \rho_j(\mathbf{T}_1) = 2, \rho_k(\mathbf{T}_1) = 1, \rho_i(\mathbf{T}_2) = 1, \rho_j(\mathbf{T}_2) = 3, \rho_k(\mathbf{T}_2) = 2$.

Now let t^i denote the announcement $t^i_{ij} = t^i_{jk} = t^i_{ik} = 1$ and t'^i the announcement $t^i_{ij} = -1, t^i_{jk} = 1, t^i_{ik} = -1$. By Claim 7,

$$\begin{aligned} \rho_i(t'^i, t^j, t^k) &= \rho_i(t^i, t^j, t^k) &= 3, \\ \rho_j(t^i, t'^j, t^k) &= \rho_j(t^i, t^j, t^k) &= 2. \\ \rho_k(t'^i, t'^j, t^k) &= \rho_k(t'^i, t'^j, t'^k) &= 2. \end{aligned}$$

Hence we conclude that, at (t'^i, t'^j, t^k) either $\rho_i = 3, \rho_j = 1$ or $\rho_i = 1, \rho_j = 3$. But $\rho_j = 3$ is impossible, as, by claim 7, $\rho_j(t'^i, t'^j, t^k) = \rho_j(t'^i, t^j, t^k)$ and $\rho_j(t'^i, t^j, t^k) \neq \rho_i(t'^i, t^j, t^k) = 3$. Hence we conclude that

$$\rho_i(t'^i, t'^j, t^k) = 3, \rho_j(t'^i, t'^j, t^k) = 1, \rho_k(t'^i, t'^j, t^k) = 2. \tag{1}$$

A similar reasoning shows that

$$\rho_j(t'^i, t^j, t'^k) = 3,$$

and hence either $\rho_i = 2, \rho_k = 1$ or $\rho_i = 1, \rho_k = 3$ at (t'^i, t^j, t'^k). But $\rho_i(t'^i, t^j, t'^k) = \rho_i(t^i, t^j, t'^k) \neq \rho_k(t^i, t^j, t'^k) = \rho_k(t^i, t^j, t^k) = 1$. So we conclude that

$$\rho_i(t'^i, t^j, t'^k) = 2, \rho_j(t'^i, t^j, t'^k) = 3, \rho_k(t'^i, t^j, t'^k) = 1. \tag{2}$$

Now, $\rho_j(t'^i, t^j, t^k) = \rho_j(t'^i, t'^j, t^k)$. By equation 1,

$$\rho_j(t'^i, t'^j, t^k) = 1$$

,

so that $\rho_j(t'^i, t^j, t^k) = 1$. As $\rho_i(t'^i, t^j, t^k) = \rho_i(t^i, t^j, t^k) = 1$,

$$\rho_k(t'^i, t^j, t^k) = 2.$$

Similarly, $\rho_k(t'^i, t^j, t^k) = \rho_k(t'^i, t^j, t'^k)$ and by equation 2,

$$\rho_k(t'^i, t^j, t'^k) = 1$$

so that

$$\rho_k(t'^i, t^j, t^k) = 1.$$

establishing a contradiction. □

6.2 Coarse rankings

We assume in the main analysis that the planner constructs a complete ranking, and that agents have strict preferences over all ranks. What happens if the planner chooses a coarser ranking, leaving some agents in the same indifference class? For example, the planner could select a dichotomous ranking, categorizing agents as "poor" or "rich" in order to implement poverty-targeted programs and subsidies.

If the planner only chooses broad categories, she might be able to construct ex-post incentive compatible and efficient mechanisms even if agents' reports are not supported by a third report. The intuition is immediate: if there exist two "worst spots" in the ranking, the planner may punish individuals who send conflicting self-reports by placing them both on the worst spot. Formally

Proposition 6. *There exists an ex-post incentive compatible and efficient mechanism in any completely informative society if and only if the planner can place two individuals in the worst spot, i.e. if and only if any individual i is indifferent between $\rho(i) = 1$ and $\rho(i) = 2$.*

Proof. If agents strictly prefer being ranked at $\rho(i) = 2$ to being ranked at $\rho(i)$, the necessity part of the proof of Theorem 1 shows that whenever there exists a pair of agents who are not observed by a third agent, there cannot exist an ex-post incentive compatible and efficient mechanism.

Conversely, if agents are indifferent between being ranked at $\rho(i) = 1$ and $\rho(i) = 2$, let $\rho(i) = 1$ and $\rho(j) = 2$ whenever i and j are the only two agents observing the ranking between i and j and $t^i_{ij} \neq t^j_{ij}$. This guarantees that agents have no incentive to send conflicting reports, and hence that this mechanism, completed by the mechanism constructed in the sufficiency part of Theorem 1, satisfies ex-post incentive compatibility and efficiency. □

Proposition 6 thus shows that it is easier to construct ex-post incentive compatible and efficient mechanisms when the planner does not construct a complete ranking of the agents. This possibility raises new interesting questions: For example, when is it possible to construct strategy-proof efficient mechanisms if the planner only selects coarse categories? We plan to address this question in future work.

6.3 Group incentive compatibility

So far we have only considered agents' individual incentives to report truthfully their types. We now briefly analyze what happens when agents can coordinate their reports, and jointly misreport their types. Consider a triangle with three agents. Each agent reports on one of the two links. The mechanism we construct in Theorem 1 assigns a ranking $\rho(i) > \rho(j)$ when at least two of the individuals report that i is higher than j. This clearly creates an incentive for any pair of agents to misrepresent their information. For example, if the true tanking is $\theta_3 > \theta_2 > \theta_1$, agents 1 and 2 have an incentive to misreport and announce that 2 is higher than 1, higher than 3, so that in the end, $\rho(2) = 3 > 2$ and $\rho(1) = 2 > 1$.

This intuition can actually be exploited to show that there does not exist any mechanism satisfying ec-post group incentive compatibility and efficiency when $n = 3$. We first provide a formal definition of ex-post group incentive compatibility:

Ex-post group incentive compatibility For any vector of types $\mathbf{T}$, there does not exist a coalition S and a vector of types $\mathbf{T}'^{\mathbf{S}}$ such that for all agents i in S,

$$\rho_i(\mathbf{T}'^{\mathbf{S}}, \mathbf{T}^{\mathbf{S}}) \geq \rho_i(\mathbf{T})$$

and

$$\rho_i(\mathbf{T}'^{\mathbf{S}}, \mathbf{T}^{\mathbf{S}}) > \rho_i(\mathbf{T}).$$

for some $i \in S$.

Proposition 7. *Let g be a triangle (a three-agent complete network). There does not exist a mechanism satisfying ex-post group incentive compatibility and efficiency.*

Proof. Consider a vector of announcements where all three agents agree on $t_{13} = -1, t_{23} = -1, t_{12} = 1$. By ex-post efficiency, $\rho(1) = 2, \rho(2) = 1, \rho(3) = 3$. We claim that ex-post group incentive compatibility implies that, whenever individual 3 announces $t^3_{13} = -1, t^3_{23} = -1, t^3_{12} = 1$, the rank of agent 1 must be different from 3. If that were not the case, there would exist an announcement (t'^1, t'^2) for agents 1 and 2 resulting in a rank $\rho(1) = 3 > 2, \rho(2) \geq 1$, contradicting ex-post group incentive compatibility. By a similar reasoning, whenever individual 1 announces $t^1_{12} = 1, t^1_{13} = 1, t^1_{23} = 1$, the rank of agent 2 must be different from 3. Finally, when individual 2 announces $t^2_{12} = -1, t^2_{23} = 1, t^2_{13} = -1$, the rank of agent 3 must be different from 3.

So consider the announcement $t^1 = (t^1_{12} = 1, t^1_{13} = 1, t^1_{23} = 1), t^2 = (t^2_{12} = -1, t^2_{23} = 1, t^2_{13} = -1), t^3 = (t^3_{13} = -1, t^3_{23} = -1, t^3_{12} = 1)$. For this announcement, neither of the three agents can be in position 3, a contradiction which completes the proof of the Proposition. □

6.4 Alternative measures of informativeness

Consider two comparison networks, h and h'. Clearly, if the comparisons provided by h are a subset of those provided by h', then h' provides more information. If this is not the case, it is not obvious which comparison network is more informative.

We choose to measure information by counting the number of edges on h. Recall that an edge on h is a comparison so we simply count comparisons. Although this measure is extremely simple, it has a theoretical foundation as the T-criterion from the field of optimal design of experiments (Pukelsheim, 1993).

An arguably better measure is to count spanning trees (which is known as the D-criterion). However, as we show in Appendix A, counting edges is almost perfect correlated with counting spanning trees for a dataset of comparison networks derived from real-life social networks. The simplicity of counting edges on the comparison network is a great advantage for studying the relationship between the social network and comparison network. We are confident that any results we derive with this simple measure of information will also be valid for more complex measures.

We study communities of different sizes so our measure must be normalized. We divide the number of edges by the number of edges on the complete network $\frac{n(n-1)}{2}$ so that our normalized measure of information is the density of the comparison network.

6.5 Homophily

In many settings, individuals with similar characteristics are more likely to form friendships. Wealthy individuals socialize together, smart students take the same classes, and high ability workers may join the same company. How does such homophily impact friend-based targeting? Using Golub and Jackson's (2012) *islands model* of network formation, we show that the probability of finding the full ranking of characteristics initially increases and then decreases in the homophily parameter.

In a community N of n individuals there are $n-1$ comparisons that are necessary and sufficient to determine the full ranking of characteristics. The lowest is compared to the second lowest, the second to the third, and so on. Call this set of comparisons A.

Suppose individuals form friendships with a given probability p (which is the Erdős-Rényi random graph model) and through friend-based targeting the realized network provides a set of comparisons C. What is the probability $\Pr[A \in C]$ that the set of comparisons A necessary and sufficient to find the full ranking is included in the realized comparisons C?

For convenience, number the individuals in the community $\{1, 2, ..., n-1, n\}$ so that the private characteristic $\theta_i > \theta_j$ if and only if $i > j$. The set $A = \{(1,2), (2,3), ..., (n-2, n-1), (n-1, n)\}$. Consider a pair (i, j). $\Pr[(i,j) \in C] = \Pr[\exists\ k \neq i, j : g_{ik} = g_{jk} = 1] = 1 - (1-p^2)^{n-2}$. Since there are $n-1$ pairs in A, $\Pr[A \in C] = \big(1 - (1-p^2)^{n-2}\big)^{n-1}$.

Now divide the community into two groups of equal size, N^L for low and N^H for high such that $\theta_l < \theta_h\ \forall\ l \in N^L, h \in N^H$. Individuals form friendships within their group with probability p_w and outside of their group with probability p_o (Golub and Jackson's (2012) *islands model* with 2 groups). For $p_w \geq p_o$, the gap between p_w and p_o is a measure of homophily.

Keeping p_w constant, if we increase p_o, $\Pr[A \in C]$ will increase since we have raised the expected density of the social network. More links provides more comparisons. This would not be an interesting test of the impact of homophily on friend-based targeting. We need to keep the expected density of the network constant while changing homophily.

Since an individual cannot form a friendship with himself, there are more outside group links than within group links in the ratio $\frac{n-5}{n-1}$. Starting from a zero homophily base of $p = p_w = p_o$, we can analyze the impact of homophily by increasing p_w and decreasing p_o to keep the expected number of links constant. Let $p_w = p + \eta$, where η is the homophily parameter. To keep the number of edges constant, $p_o = p - \eta\frac{n-1}{n-5}$.

From a base of $\eta = 0$ so $p = p_w = p_o$ we can increase η and observe how $\Pr[A \in C]$ responds. First, we must define $\Pr[A \in C]$ in terms of p_w and p_o. Except for the comparison of the highest ranked in N^L to the lowest ranked in N^H, all other comparisons in A are within group comparisons. $\Pr[A \in C]$ is the product of two terms, the $n-2$ within group comparisons and the single cross group comparison of the highest ranked in N^L to the lowest ranked in N^H.

$$\Pr[A \in C] = \Big(1 - \big(1 - p_w^2\big)^{\frac{n}{2}-2}\big(1 - p_o^2\big)^{\frac{n}{2}}\Big)^{n-2}\ \Big(1 - \big(1 - p_w p_o\big)^{n-2}\Big)$$

Plugging in $p_w = p + \eta$ and $p_o = p - \eta\frac{n-1}{n-5}$ gives a complicated expression for $\Pr[A \in C]$ that cannot be easily simplified. Instead, we show the relationship between the homophily parameter η and $\Pr[A \in C]$ graphically in Figure 10 for community size $n = 250$ and base probability of friendship $p = 0.15$. Along the horizontal axis, as η increases from 0 to around

0.1 the probability that the realized comparisons contain the comparisons needed to derive the full ranking increases. The intuition is simple. As homophily increases the probability that some individual k is friends with two individuals in the same group increases while the probability that k is friends with two individuals in different groups decreases. Since nearly all of the comparisons in A are pairs within the same group, $\Pr[A \in C]$ rises with homophily. However, when the probability of outside group friendships p_o approaches zero, $\Pr[A \in C]$ approaches zero since there is little chance that the comparison of the highest in N^L to the lowest in N^H is in C. In Figure 10 $\Pr[A \in C]$ drops sharply as η is above 0.12 and $p_o = p - \eta$ approaches zero. Low levels of homophily improves friend-based targeting whereas extreme homophily is damaging.

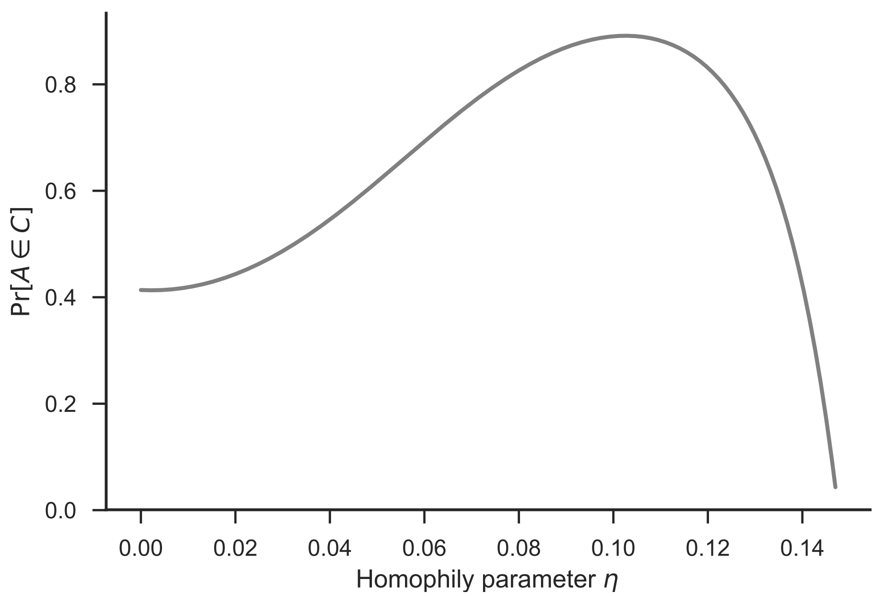


Figure 10: Impact of homophily ($p = 0.15$, $n = 250$)

Note:

Figure 10 was constructed to highlight a strong impact of homophily. If the base probability of friendship p is very high, $\Pr[A \in C]$ is close to 1 and homophily has little impact. If p is very low, $\Pr[A \in C]$ remains close to zero. In addition to changing p, the model may also be enriched.

In the current model, the private characteristic θ is perfectly correlated with the groups N^L and N^H. We can relax this condition by changing the exponents on the formula of $\Pr[A \in C]$. If θ and membership in N^L and N^H is not perfectly correlated, some of the comparisons in A will change from within group to outside group comparisons. Let t be the number of outside group comparisons. The adjusted equation for $\Pr[A \in C]$ is shown below. In the perfect correlation model $t = 1$.

$$\Pr[A \in C] = \Big(1 - \big(1 - p_w^2\big)^{\frac{n}{2}-2}\big(1 - p_o^2\big)^{\frac{n}{2}}\Big)^{n-1-t} \, \Big(1 - \big(1 - p_w p_o\big)^{n-2}\Big)^{t}$$

As t and therefore the number of outside group comparisons in A increases, q_o has a stronger bite. All else constant, $\Pr[A \in C]$ begins decreasing for smaller values of η when t is larger.

7 Conclusion

This paper analyzes the design of mechanisms to rank individuals in communities where individuals only have local, ordinal information on the characteristics of their neighbors. In these communities, pooling the information of all individuals may not be sufficient to obtain a complete social ranking and we distinguish between completely informative societies and societies where only incomplete social rankings can be obtained. In completely informative societies, we show that the planner can construct an ex-post incentive compatible and ex-post efficient mechanism if and only if all pair of agents are observed by a third agent, i.e. any pair of agents in the social network has a common friend. We use this insight to characterize the sparsest social network for which a complete ranking exists as the "friendship network" (or Dutch windmill network) of (Erdős, Rényi, and Sós, 1966). When the society cannot achieve a complete ranking, we show that any self-report which is not supported by a third party must be discarded. We provide two sufficient conditions on the social network under which an ex-post incentive compatible and ex-post efficient mechanism can be constructed. First, in bipartite networks, agents on one side of the graph can be used to rank agents on the other side, resulting in an ex-post efficient but very incomplete social ranking. Second, in "social quilts", where all links are supported in triangles, the planner can use the congruence of reports to construct truthful rankings over any pair of agents. We use data on social networks in two real world settings in India and Indonesia to illustrate the results of the theoretical analysis. In both cases, we compute the fraction of bilateral comparisons which can be obtained by the planner, and analyze how this index correlates with other characteristics of the social network, like density, average distance or clustering. Finally, we discuss robustness and extensions of the model, focusing on strategy-proofness, group incentive compatibility, coarse rankings, homophily and alternative definitions of incompleteness of the social ranking.

To the best of our knowledge, the theoretical problem analyzed in this paper – the design of a mechanism constructing a complete ranking when agents have local, ordinal information – is new. We would like to further our understanding of the problem in future work, by considering in more detail the difference between ordinal and cardinal information, between complete and coarse rankings, and between different concepts of implementation. We also plan on extending the empirical and policy implications of the theoretical model by analyzing in more detail specific institutional settings.

References

Alatas, Vivi, Abhijit Banerjee, Arun G Chandrasekhar, Rema Hanna, and Benjamin A Olken. 2016. "Network structure and the aggregation of information: Theory and evidence from Indonesia." *The American Economic Review* 106 (7):1663–1704.

Alatas, Vivi, Abhijit Banerjee, Rema Hanna, Benjamin A Olken, and Julia Tobias. 2012. "Targeting the poor: evidence from a field experiment in Indonesia." *American Economic Review* 102 (4):1206–40.

Alon, Noga, Felix Fischer, Ariel Procaccia, and Moshe Tennenholtz. 2011. "Sum of us: Strategyproof selection from the selectors." In *Proceedings of the 13th Conference on Theoretical Aspects of Rationality and Knowledge*. ACM, 101–110.

Aziz, Haris, Omer Lev, Nicholas Mattei, Jeffrey S Rosenschein, and Toby Walsh. 2016. "Strategyproof Peer Selection: Mechanisms, Analyses, and Experiments." In *AAAI*. 397–403.

Banerjee, Abhijit, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson. 2013. "The diffusion of microfinance." *Science* 341 (6144):1236498.

Barrow, Daniel, Ian Drayer, Peter Elliott, Garren Gaut, and Braxton Osting. 2013. "Ranking rankings: an empirical comparison of the predictive power of sports ranking methods." *Journal of Quantitative Analysis in Sports* 9 (2):187–202.

Basurto, Pia M, Pascaline Dupas, and Jonathan Robinson. 2017. "Decentralization and Efficiency of Subsidy Targeting: Evidence from Chiefs in Rural Malawi." Tech. rep., National Bureau of Economic Research.

Baumann, Leonie. 2017. "Identifying the Best Agent in a Network." Tech. rep., Working paper.

Brouwer, Andries E and Willem H Haemers. 2011. *Spectra of graphs*. Springer Science & Business Media.

Brown, Caitlin, Martin Ravallion, and Dominique Van de Walle. 2016. "A poor means test? Econometric targeting in Africa." Tech. rep., National Bureau of Economic Research.

Coady, David, Margaret Grosh, and John Hoddinott. 2004. "Targeting outcomes redux." *The World Bank Research Observer* 19 (1):61–85.

Crémer, Jacques and Richard McLean. 1985. "Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist When Demands Are Interdependent." *Econometrica* 53 (2):345–61.

De Clippel, Geoffroy, Herve Moulin, and Nicolaus Tideman. 2008. "Impartial division of a dollar." *Journal of Economic Theory* 139 (1):176–191.

Erdős, Paul, Alfréd Rényi, and Vera Sós. 1966. "On a Problem of Graph Theory." *Studia Scientiarium Mathematicarum Hungarica* 1 (1):215–235.

Golub, Benjamin and Matthew O Jackson. 2012. "How homophily affects the speed of learning and best-response dynamics." *The Quarterly Journal of Economics* 127 (3):1287–1338.

Holzman, Ron and Hervé Moulin. 2013. "Impartial nominations for a prize." *Econometrica* 81 (1):173–196.

Huneke, Craig. 2002. "The friendship theorem." *The American mathematical monthly* 109 (2):192–194.

Hussam, Reshmaan, Natalia Rigol, and B Roth. 2017. "Targeting high ability entrepreneurs using community information: Mechanism design in the field." Tech. rep., Working paper.

Jackson, Matthew O, Tomas Rodriguez-Barraquer, and Xu Tan. 2012. "Social capital and social quilts: Network patterns of favor exchange." *American Economic Review* 102 (5):1857–97.

Jiang, Xiaoye, Lek-Heng Lim, Yuan Yao, and Yinyu Ye. 2011. "Statistical ranking and combinatorial Hodge theory." *Mathematical Programming* 127 (1):203–244.

Kurokawa, David, Omer Lev, Jamie Morgenstern, and Ariel D Procaccia. 2015. "Impartial Peer Review." In *IJCAI*. 582–588.

Laslier, Jean-Francois. 1997. *Tournament solutions and majority voting*. 7. Springer Verlag.

Longyear, Judith Q and Torrence D Parsons. 1972. "The friendship theorem." In *Indagationes Mathematicae (Proceedings)*, vol. 75. Elsevier, 257–262.

Osting, Braxton, Christoph Brune, and Stanley J Osher. 2014. "Optimal data collection for informative rankings expose well-connected graphs." *Journal of machine learning research* 15:2981–3012.

Pukelsheim, Friedrich. 1993. *Optimal design of experiments*, vol. 50. SIAM.

Rai, Ashok S. 2002. "Targeting the poor using community information." *Journal of Development Economics* 69 (1):71–83.

Shrikhande, SS. 1959. "The Uniqueness of the L_2 Association Scheme." *The annals of mathematical statistics* :781–798.

Wilf, Herbert S. 1971. "The friendship theorem." In *Combinatorial Mathematics and its Applications (Proc. Conf., Oxford, 1969)*. 307–309.

A Further details on measures incomplete comparison networks

This appendix expands on Section 6.4. We chose to count edges on the comparison network, h, to measure the information h provides about the unknown ranking, ρ. An alternate measure is to count spanning trees. We use an example to show that counting spanning trees is an intuitive approach. Fortunately, when the comparison network is derived from a social network, counting spanning trees is almost perfectly correlated with counting edges. Our results will be the same with either measure.

Researchers have developed many algorithms to infer a ranking from a matrix of pairwise comparisons. Least-squares ranking performs well (Barrow et al., 2013), can handle both ordinal and cardinal comparisons, and allows for measurement of the quality of the ranking ex-post (Jiang et al., 2011). We minimize the squared distance between the observed pairwise comparisons and the unknown private characteristic ω.

$$\min_{\Omega \in M} \sum_{i,j} H_{ij}(\Omega_{ij} - T_{ij})^2, \tag{3}$$

where M is the set of rank-2 skew-symmetric matrices.

$$M = \{\Omega \in \mathbb{R}^{nxn} \mid \Omega_{ij} = \text{sign}(\omega_i - \omega_j),\ \omega : N \to \mathbb{R}\}. \tag{4}$$

The sign function represents ordinal comparisons. Negative numbers are converted to -1 and non-negative to $+1$. The pairwise comparison data in T is used to estimate the unknown private characteristics ω. The ranking is derived by the rule that $i > j$ if $\omega_i > \omega_j$.

The estimate of ω depends on the observed pairwise comparisons, H. More comparisons provides more information about ω. Osting, Brune, and Osher (2014) show that the information may be quantified in a simple way.

Assume pairwise comparison data on T is generated as follows:

$$T_{ij} = \text{sign}((\omega_i - \omega_j) + \epsilon_{ij}) \tag{5}$$

where $\mathbb{E}[\epsilon_{ij}] = 0$ and $\text{Var}[\epsilon_{ij}] = \sigma^2/H_{ij}$ for some constant σ. The variance of ϵ_{ij} decreases as the planner receives more comparisons between i and j. Under the data generating process in Equation A, Osting, Brune, and Osher (2014) show that the information provided by the matrix of comparisons, H, does not depend on the private characteristics, ω. We may measure the information provided by H without having to make further assumptions about ω. Osting, Brune, and Osher (2014) then show that the Laplacian of H is equal to the Fisher Information.

Definition 1. Laplacian. The Laplacian $L(H)$ equals $D(H) - H$ for the adjacency matrix H and the degree matrix $D(H)$. The degree matrix $D(H)$ is an nxn matrix with the degree of i on the diagonal of row i and zeros elsewhere.

The Fisher Information is the variance of the partial derivative of the log-likelihood function with respect to a parameter to be estimated. It may be understood as the curvature of the log-likelihood function. If the log-likelihood function is steep around the estimate of the parameter, the estimate is highly informative. For friend-based targeting using least-squares the Fisher Information is an nxn co-variance matrix since there are n private characteristics to estimate. Greater Fisher Information shows greater sensitivity of the estimate to the realized comparison data, T, and a more informative estimate. The research field of Optimal Design of Experiments summarize this matrix into a single measure to compare the informativeness of different experimental designs. There are several popular measures including the D-criterion, T-criterion, and E-criterion (Pukelsheim, 1993).

The D-criterion measures the number of spanning trees on h, calculated using the determinant of any co-factor of $L(H)$; the T-criterion measures the trace of $L(H)$, which is simply the number of edges on h; and E-criterion measures *algebraic connectivity*—the second smallest eigenvalue of $L(H)$. We prefer the D-criterion. The measure is used frequently by researchers in the field of Optimal Design and it provides a simple intuition, which we explain in the following example.

Consider the social network of four individuals shown in Figure A1. Individual j is friends with all other individuals in the population, k and l are friends, and i is only friends with j. This social network, with the adjacency matrix g', defines the matrix of comparisons, H', shown below. For instance, since k provides the comparison between j and l, $C'_{jl} = 1$.

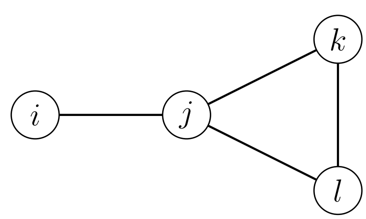


Figure A1: Example social network of four individuals.

$$g' = \begin{array}{c} \\ i \\ j \\ k \\ l \end{array} \begin{array}{c} \begin{matrix} i & j & k & l \end{matrix} \\ \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \end{array} \quad H' = \begin{array}{c} \\ i \\ j \\ k \\ l \end{array} \begin{array}{c} \begin{matrix} i & j & k & l \end{matrix} \\ \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \end{array} \quad L(H') = \begin{array}{c} \\ i \\ j \\ k \\ l \end{array} \begin{array}{c} \begin{matrix} i & j & k & l \end{matrix} \\ \begin{pmatrix} 2 & 0 & -1 & -1 \\ 0 & 2 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{pmatrix} \end{array}$$

The D-criterion (and the number of spanning trees on H') is the determinant of any co-factor of $L(H')$, which is also equal to $\prod_{x\geq 2} \lambda_x$, where λ_x is the eigenvalues of $L(H')$ ordered from smallest to largest. For H' we find the number of spanning trees is eight. In Figure A2, H' is shown in graph form and we enumerate the eight spanning trees in Figure A3.

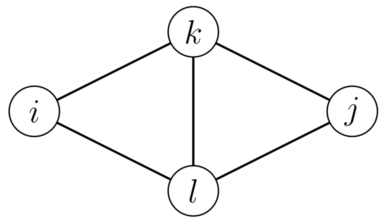


Figure A2: Matrix of comparisons H' as a graph.

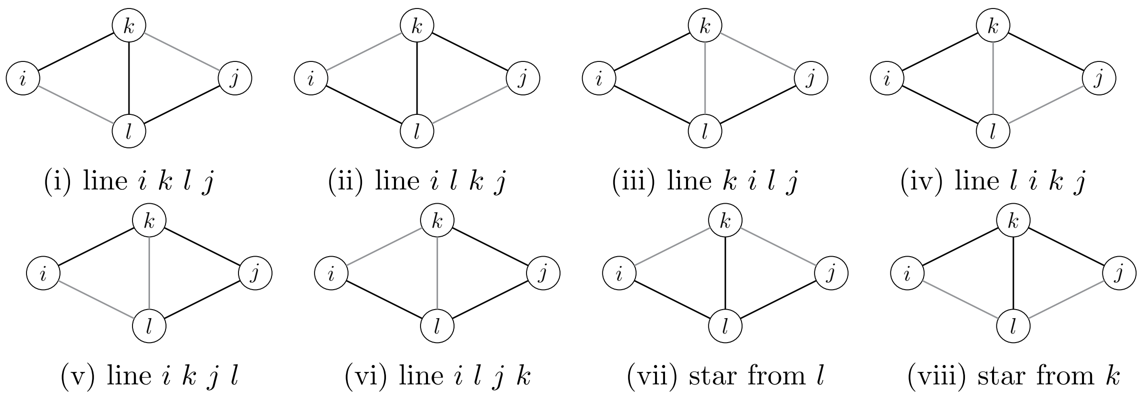


Figure A3: Spanning trees from the comparison graph h'.

The relationship between the number of spanning trees and information from the network of comparisons is intuitive. The true ranking lies along one of the spanning trees. For example, if $\omega_i > \omega_k > \omega_l > \omega_j$, friend-based targeting would reveal the ranking $i > k > l > j$, and the necessary comparisons to identify this ranking are shown by the first spanning tree enumerated in Figure A2. However, if $\omega_i > \omega_j > \omega_k > \omega_l$, friend-based targeting would reveal the ranking $i \lesseqgtr j > k > l$. Since the comparison (i, j) is not reported to the planner, she cannot infer if $i > j$ or $j > i$.[4] The spanning tree i, j, k, l that represents the necessary comparisons to find a complete ranking given $\omega_i > \omega_k > \omega_l > \omega_j$ is not present on C'. Interestingly, the comparisons on partial ranking $i \lesseqgtr j > k > l$ is represented by the spanning tree in the last panel of Figure A2. Since ω is unknown, the more spanning trees on H, the more likely a complete ranking is revealed.

[4] If the pairwise comparisons were cardinal rather than ordinal, there is additional information for the planner to infer if $i > j$ or $j < i$.

As the example illustrated, the number of spanning trees provides an intuitive measure of information for ranking from pairwise comparisons. Unfortunately, this intuitive measure is difficult to study analytically. It is much easier to count edges (T-criterion) than to count spanning trees (D-criterion).

The number of spanning trees has a very large support that depends on the number of nodes. A complete network has n^{n-2} spanning trees. We can normalize by dividing by n^{n-2}, which gives us the percentage of pairwise comparisons we would need to draw uniformly at random without replacement to find the same number of spanning trees. We also normalize the edge count by dividing by $\frac{n(n-1)}{2}$ so the normalized T-criterion is the density of the comparison network.

We compare the T- and D-criteria in Figure A4 using network data from villages in India and neighborhoods in Indonesia. You will notice that these two measures are highly correlated, which is not surprising. On a given comparison network, h, adding edges adds spanning trees. Only in cases when the number of edges on two comparison network h and h' is very close will the T- and D-criteria provide opposing measures of information.

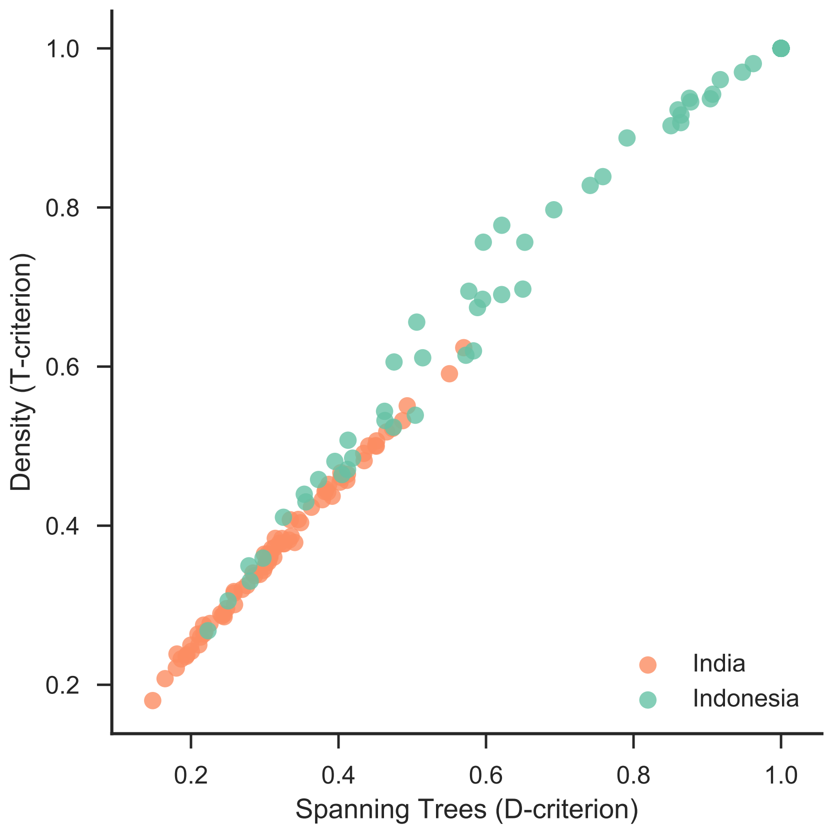


Figure A4: Information measures: edges versus spanning trees

Note: Information can be measured by counting edges (T-criterion) or counting spanning trees (D-criterion). This figure shows a close correlation between these measures. Both measures are calculated on the giant component of the social network to increase the number of non-zero data points. (The number of spanning of trees is zero when the comparison network is not connected.) Only networks with at least 50 individuals in the giant component are plotted.