



# HHS Public Access

Author manuscript

*Soc Networks*. Author manuscript; available in PMC 2017 August 25.

Published in final edited form as:

*Soc Networks*. 2017 January ; 48: 157–168. doi:10.1016/j.socnet.2016.08.008.

## An exploratory comparison of name generator content: Data from rural India

Holly B. Shakya<sup>a,b,\*</sup>, Nicholas A. Christakis<sup>b,\*</sup>, and James H. Fowler<sup>a,c</sup>

<sup>a</sup>School of Medicine, University of California San Diego, 9500 Gilman Dr., La Jolla, CA 92093, United States

<sup>b</sup>Yale Institute for Network Science, PO Box 208263, New Haven, CT 06520, United States

<sup>c</sup>Political Science Department, University of California San Diego, SSB 301, 9500 Gilman Dr., La Jolla, CA 92093, United States

### Abstract

Since the 1970s sociologists have explored the best means for measuring social networks, although few name generator analyses have used sociocentric data or data from developing countries, partly because sociocentric studies in developing countries have been scant. Here, we analyze 12 different name generators used in a sociocentric network study conducted in 75 villages in rural Karnataka, India. Having unusual sociocentric data from a non-Western context allowed us to extend previous name generator research through the unique analyses of network structural measures, an extensive consideration of homophily, and investigation of status difference between egos and alters. We found that *domestic interaction* questions generated networks that were highly clustered and highly centralized. Similarity between respondents and their nominated contacts was strongest for gender, caste, and religion. We also found that *domestic interaction* name generators yielded the most homogeneous ties, while *advice* questions yielded the most heterogeneous. Participants were generally more likely to nominate those of higher social status, although certain questions, such as who participants *talk to* uncovered more egalitarian relationships, while other name generators elicited the names of social contacts distinctly higher or lower in status than the respondent. Some questions also seemed to uncover networks that were specific to the cultural context, suggesting that network researchers should balance local relevance with global generalizability when choosing name generators.

### Keywords

Name generators; Social norms; India; Homophily; Sociocentric; Social networks; Social hierarchy

---

\*Corresponding author at: University of California San Diego, 9500 Gilman Drive, Mail Code 0507, La Jolla, CA 92093-0041, United States. hshakya@ucsd.edu (H.B. Shakya), nicholas.christakis@yale.edu (N.A. Christakis), jfowler@ucsd.edu (J.H. Fowler).

## 1. Introduction

Since the 1970s sociologists have explored the best means for measuring social networks. While survey questions are most commonly used, many scholars have experimented with different measurement tools, searching for the most valid and reliable methods including the reverse small-world technique and various forms of personal diaries, including a smart phone app that allows participants to enter social interactions in real time (Bernard et al., 1987, 1990; Fu, 2005; Lerner et al., 2014). The appropriate technique, however may depend upon the type of network study being conducted and the research question being asked (Knipscheer and Antonucci, 1990).

Sociocentric studies focus on a small population and attempt to ascertain all of the social relationships within a set of interconnected individuals (Marin and Wellman, 2011). In the case of such studies, it is important to be able to accurately connect nominated individuals in order to analyze the greater structure of the network. For large-scale sociocentric data collection efforts, the most practical means of eliciting the names of social contacts, therefore, may be through the administration of surveys that are administered uniformly to the entire population.

Surveys, of course, are composed of questions. The network data collection procedure most frequently used in the context of a survey is to simply ask participants (*egos*) the name of those people with whom they have social connections (*alters*) (Burt, 1984; Marin, 2004). This kind of question is called a “name generator”. Although this method has its drawbacks, it is generally reliable, and is more efficient than other methods (Bien et al, 1991). The question can be hypothetical (with whom would you do something) or factual (with whom have you done something) (De Lange et al., 2004). While the way in which a question is asked is important in terms of eliciting network ties, the most crucial component is the content of the question itself (Ferligoj and Hlebec, 1999). The ties elicited by the name generators create the structure of the network, and the specific questions asked to elicit those ties provide the context.

While friendship, as a social phenomenon, occurs throughout the world, the actions that define friendship may differ greatly (Hruschka, 2010). Fischer, for instance, determined that friendship in America is vaguely defined, although generally it refers to people to whom no other specific title, such as co-worker or relative, can be given (Fischer, 1982). Name generator questions, therefore, usually focus on the specific context of relationships. The content of the question determines the type of relationship depicted, which is a crucial component of understanding the significance of the network itself. Network contexts can be categorized as exchange (people with whom an ego engages in reciprocal service provision such as borrowing and lending money); role relation (specific relationships such as spouse or mother); interactive (people with whom an ego interacts during the day); and affective (people with whom an ego shares strong emotional bonds) (Knipscheer and Antonucci, 1990; Marin and Hampton, 2007; Van der Poel, 1993). Exchange, interactive, and role relation questions tend to elicit the largest networks, while affective networks the smallest but are comprised of the closest relationships. Obviously, there is likely to be some degree of overlap between any of these categories.

Ideally, researchers would have the time and resources to collect data on every possible kind of relationship for every person within a given network. However, there are resource limitations both in the ability of most researchers to collect such an exhaustive amount of data and in the ability of the respondents to enumerate social contacts to that extent. For example, respondents might experience survey fatigue, and begin to increasingly underreport alters with each additional name generator question (Pustejovsky and Spillane, 2009). A critical decision faced by any network researcher, then, is how many name generators should be used. If there are too many questions, researchers may be collecting redundant information. De Lange and colleagues analyzed the results of name generator questions administered to employees in a small organization in Belgium, and they were able to uncover three primary conceptual factors as well as nomination overlap in a group of questions regarding relations within the organization (De Lange et al., 2004). This information provided crucial clues regarding what questions could potentially be cut from an instrument with multiple questions. The validity of network data that uses too few name generators, on the other hand, may be compromised by the fact that respondents tend to underreport their important ties, and instruments using only one name generator are not sensitive enough to capture the size and complexity of the real network (Burt, 1997; Marin, 2004; Marin and Hampton, 2007). For example, while one study in the US has shown that the sex and race of nominated alters are relatively invariant across name generators, other work has shown significant gender variation (Campbell and Lee, 1991; Ruan, 1998; Straits, 2000). Two of these studies also showed significant differences in tie characteristics, i.e. the strength and type of relationship (Campbell and Lee, 1991; Ruan, 1998). While McCallister and Fischer were the first to use multiple name generators to ascertain a more comprehensive and multi-dimensional network in 1978 (McCallister and Fischer, 1978), many network studies still limit the number of name generators to 1 or 2 general purpose name generators. One of the most commonly used is “with whom do you discuss important matters?”.

An important shortcoming of using one general purpose name generator is the assumption that it can be correlated with diverse outcomes, such as finding a job, as the pathway by which this is supposed to occur is by no means clear (Parigi and Bearman, 2005). The *functional specificity hypothesis* is based upon the belief that individuals rely on different people for different types of support depending upon their need (Perry and Pescosolido, 2010). A person with whom someone discusses politics may not be the person upon whom they rely for assistance with a sick child. Wellman and Wortley, for instance, show that the types of social support given by a person’s “strong ties” differs substantially from that given between parents and children (Wellman and Wortley, 1990). Milardo demonstrates that what he terms “significant other” relationships (which in his operationalization includes both close ties and kin) do not overlap significantly with exchange networks (Milardo, 1989, 1992). He also emphasizes the importance of understanding the difference in the structural elements of these different types of networks if we are to attribute significance to network structure overall. Given this viewpoint, the “important matters” name generator (or any general sort of name generator focused on non-specific social interaction) could be insufficient for eliciting ties that might be influential in a variety of contexts. Perry and colleagues (Perry and Pescosolido, 2010) show there is a significant difference between the

“important matters” discussion network and a “health matters” discussion network both in structure and ability to predict relevant health outcomes (the health matters network could predict health outcomes in this study but not the important matters network). Researchers in Mali collected network data on women from two different ethnic groups using questions to elicit support networks on four different dimensions: emotional, cognitive (meaning information sharing), material, and practical (help with childcare, etc.) (Adams et al., 2002). Their research demonstrated that the correlation between network composition and child mortality differed between network types as well as across ethnic groups. Ruan’s study (Ruan, 1998) showed that there was considerable overlap between the names of people generated by the important matters question and those with whom the respondent socializes, but not those from whom the respondent expects help or goes to regarding family matters.

One of the main purposes of applied network research is to understand how behaviors cluster and shift within communities. By using network methods to implement interventions, researchers and program specialists can increase their reach and impact. Often this approach involves strategies to magnify possible social effects so that a behavior adopted by one person spreads to others in her social network. In order to most successfully exploit these potential social effects, it is important that network researchers use questions that are the most likely to elicit the network in which these effects occur.

In some contexts this may be best achieved by finding individuals with whom participants are similar, as individuals may be more likely to adopt a new behavior if that behavior has already been adopted by someone with whom she is similar (Centola, 2011). In other contexts this may be best achieved by finding individuals with whom participants are different. High levels of *homophily* (similarity between socially connected individuals) can cause norms to become entrenched making social change difficult. For instance, past research has suggested that smoking behavior may become solidified within small pockets of smokers as a result of increasing anti-smoking sentiment (Christakis and Fowler, 2008). In contrast, in other contexts, lower status individuals tend to emulate higher status individuals (DiMaggio and Garip, 2011), suggesting that it may be important to elicit network ties between individuals of different status. Often behavior change interventions will utilize a “peer educator” model, in which highly connected, high status individuals are educated to disseminate new ideas within their focal communities (Valente, 2012). Recent research has also suggested that targeting friends of friends might be beneficial, given that nominated individuals tend to be more central than those that nominated them (Cobb et al., 2010; Kim et al., 2015). These sorts of strategies require an approach that goes beyond considering just network size, and that takes into consideration the function of the name generators themselves. Nevertheless, few studies have considered the degree to which egos and alters are similar in socio-centric contexts, and, relatedly, the degree to which they hold different status within their communities.

The choice of which name generator to use in a study is therefore a matter of critical interest to network scientists. It serves the function of the primary survey instrument, and, as such, must be both valid and reliable in capturing the social structure of the respondent population. Despite the importance of name generators in creating network datasets, studies analyzing the use of different name generators have been sparse, particularly when looking at

populations in non-Western cultural settings. The purpose of this study is an analysis of the characteristics of the networks elicited by 12 different name generators used for the collection of sociocentric network data in rural Karnataka, India. Sociocentric studies in developing countries have been scant (Perkins et al., 2015), so little if any work has considered name generators in a context such as this. Furthermore, lack of resources and strongly held social structures mean that studies that consider network dynamics in countries such as India are particularly important (Apicella et al., 2012; Perkins et al., 2015).

This study allows us to build on previous name generator research in a completely different setting, while exploring questions that would be difficult to answer using more the more limited sources of data normally available in egocentric studies. First, we can compare full network structure across name generators, beyond the limited comparisons of simple density and degree. Second, because alters have been given the full behavioral and demographic survey, we can consider homophily between egos and alters beyond the ego-reported alter characteristics (such as age and gender) available in most egocentric network studies. Third, while a small body of research has suggested that egos are likely to nominate higher status alters, in this data we have both the social network and demographic characteristics for both egos and alters with which to test this. Importantly, this also allows us to not only quantify the extent to which egos may nominate higher-status alters, but also to identify the name generators that are most likely to elicit more egalitarian ties versus those that may be more likely to elicit more hierarchical ties. In the context of a strongly collective hierarchical society, such as rural India, cultural aspects that may affect network responses include sharply defined gender roles, and intra-group and intra-family hierarchies which may result in very discernible differences in gender and status differentials for nominations.

Here, we compared the networks formed by each name generator to test the following hypotheses. (1) Name generators within the same domain will overlap sufficiently to allow us to use only one per domain; (2) name generators will differ in the degree to which they elicit status differential in the ties; for instance *giving advice* will elicit the names of lower status alters, while *getting advice from* will elicit the names of higher status alters; (3) we will find that ties elicited through name generators that focus on social interaction (such as who you talk to, and who you visit or receive in your home) will be the most *homophilic* (same-type) while those focused on exchange will be the most *heterophilic* (opposite-type) with respect to demographic, behavioral, and household characteristics; and (4) The relations network will be the densest and most reciprocated, while those that involve one-way provision of resources (such as giving advice or lending money) will be the least dense and least reciprocated.

## 2. Data

In the Karnataka data, information was collected as part of a study to understand the network diffusion of micro-finance (Banerjee et al, 2013; Jackson et al., 2012). A complete census was taken by interviewing one person within each household in the village regarding household characteristics such as latrine ownership. Individual surveys were then used to collect demographic and network data from over one half of eligible households (those with women between the ages of 18 and 57), which were randomly chosen using stratification by

religion and geographic sub-location (Banerjee et al., 2013; Jackson et al, 2012). Respondents included eligible women within each household and their spouses in 75 villages. The total number of individuals interviewed was 16,984. After removing observations with missing data, our analysis included data on 16,403 individuals coming from 6811 households or approximately 46% of all households per village. Of these, 6543 were household heads, 5919 were spouses of household heads, and 4117 were other individuals in the household. We defined our population to be eligible women and their spouses. Social ties to others outside this population were excluded.

## 2.1. Name generators

In this study, respondents (termed here *egos*) were asked to name up to 8 individuals (termed here *alters*) for each name generator, and there were 12 name generators administered. Name generator questions included interactive, role relation, and exchange ties.

1. *Talk to:* Name the 4 non-relatives whom you speak to the most.
2. *Visit their home:* In your free time, whose house do you visit?
3. *Invite home:* Who visits your house in his or her free time?
4. *Borrow rice from:* If you needed to borrow kerosene or rice, to whom would you go?
5. *Lend rice to:* Who would come to you if he/she needed to borrow kerosene or rice?
6. *Borrow money from:* If you suddenly needed to borrow 50 Rupees for a day, whom would you ask? (This represents roughly one days wages in these villages).
7. *Lend money to:* Who do you trust enough that if he/she needed to borrow 50 Rupees for a day you would lend it the him/her?
8. *Give advice to:* Who comes to you for advice?
9. *Take advice from:* If you had to make a difficult personal decision, whom would you ask for advice?
10. *Help during emergency:* If you had a medical emergency and were alone at home whom would you ask for help in getting to a hospital?
11. *Related to:* Name any close relatives, aside those in this household, who also live in this village?
12. *Go to temple with:* Do you visit temple/mosque/church? Do you go with anyone else? What are the names of these people?

## 2.2. Network ties

To better understand the factors predicting a tie between each ego and alter pair, we used the network with directed ties, meaning that for each nomination we know who is the ego (the nominator) and who is the alter (the nominated). Our final dataset consisted of one

observation for each ego-alter dyad and included covariates for both individuals. The complete network from all 12 name generators yielded 80,838 directed social ties.

### 2.3. Demographic and household level traits

Participants reported their age, gender, religion (Hindu or Muslim), and mother tongue (Kannada, Tamil, Telugu, or Hindi). Participants were also asked to identify to which caste they belong (scheduled caste: SC, scheduled tribe: ST, other backward caste: OBC, or general: GEN). Education was measured using 16 levels ranging from none to higher degree. Household quality variables included number of rooms in the home, number of beds in the home, and household electricity (whether privately supplied, government supplied, or not available). Consistent with prior work in traditional agrarian societies in which data regarding income is unreliable, we used these household quality measures as a proxy for income (Morris et al., 2000). Ration cards are used in India to guarantee government subsidies for food depending upon income. Therefore, we categorized each individual according to the type of ration card held (below poverty level, above poverty level, or not holding a card). A breakdown of the demographic characteristics by name generator is shown in Table 1.

### 2.4. Behavioral traits

We also included several behavioral level measures of the type that might spread from person to person in a network. These included binary measures of participation in micro-finance programs for all individuals within villages in which the micro-finance program was being introduced (43 villages), holding an election card, having a savings account, membership in a savings group, working outside of the village for those who are employed, and owning a latrine. Although latrine ownership is a household level measure, and one that is reflective of economic status, previous work on latrine ownership has shown that it may be susceptible to social effects so conceptually we included it in this category (Shakya et al., 2015).

### 2.5. Network measures

To explore whether the networks generated by each name generator differed according to structural characteristics, we calculated network level measures for each village and then averaged them across all 75 villages for each name generator. Previous research has suggested that network structure can play a significant role in the adoption, maintenance, and shifting of new behaviors (Barrington et al, 2009; Latkin et al., 2004; Tobin and Latkin, 2008; Uzzi et al., 2007). An understanding of the structural differences between networks elicited by different questions may help to inform decisions around which questions to use for behavior change interventions.

*Density* is a measure of network cohesion, and is calculated by dividing the total number of observed ties by the total number of possible ties. Another measure of cohesion, *transitivity*, is the number of connected triads divided by the total number of possible triads (in other words, what is the probability that two of a person's friends are also friends with one another?). Both density and transitivity can constrain adoption of normatively driven behaviors, as individuals with connected friends are less likely to risk social sanctions by

defying a currently held norm (Granovetter, 1983). On the other hand, once a critical mass of people has begun to adopt a new behavior, it can spread most quickly in networks with high density and/or transitivity as the strong level of connection between network members assures that each person within the network is exposed quickly and frequently (Centola and Macy, 2007).

*Centralization* is a measure of the disparity between the most central individuals and those least central (Valente, 2010). Some interventions target the most central individuals for training or behavior change in order to increase the likelihood that a certain outcome will diffuse throughout the population (Valente et al., 2003, 2007; Valente, 2012). Understanding the distribution of centrality within communities can therefore offer important insights into the possible dynamics of diffusion. If there are only a few prominently positioned people within the network then it may be easier to identify them for diffusion training. However if people within the network are similar in their centrality then identifying the most central may be less important. There are several different network level centralization measures that are each based on individual level centrality measures: *degree centralization*, *closeness centralization*, and *betweenness centralization*. Centralization scores are calculated by taking the difference between the score of the person with the highest centrality measure from the individual scores of all the others in the network. These differences are then summed and divided by what this sum would be under the largest possible centralization for a network of that size. Degree centralization then reflects to what degree a few individuals have the most direct connections; closeness centralization reflects the degree to which a few individuals are most closely connected to all others in the network, and betweenness centralization the degree to which a few individuals act as bridges between others in the network.

Finally, we calculated a measure of *network reciprocity*, which is the proportion of reciprocated ties in the network (Valente, 2010).

### 3. Statistical analyses

We used logit regression models to test whether trait similarity between ego and alter significantly predicted the chance that an ego would name an alter as a social contact. To do this, we followed Apicella and colleagues, applying the method they used to calculate trait similarity among the Hadza in Tanzania (Apicella et al, 2012). We quantified the association between social ties and similarity for each trait across the 12 name generators using a dataset composed of all possible nominations within each village ( $n = 4,197,904$ ) (Apicella et al., 2012). Of these, there were a total of 80,838 observed nominations across all of the name generators in the dataset. Statistical analyses were run for each of 16 traits of interest in order to predict ties for each of the 12 separate name generators. The dependent variable for the name generators was 1 if person  $i$  named person  $j$  as a social contact in that name generator (for instance if  $i$  names  $j$  as someone to whom they go for advice), and 0 if otherwise.

The main independent variables in these basic models included a measure for the trait of interest for the person  $i$  (the “ego”), a measure for the trait of interest for person  $j$  (the “alter”), and a measure of the similarity of their traits (this is the homophily measure). For



example, for gender, we ran 12 different regressions: one for each name generator. Each regression included a variable for ego's gender, a variable for alter's gender, and a variable indicating whether or not they had the same gender. These three variables allow us to measure how much that trait predicts the sending of nominations (do women name more social contacts than men?), the receiving of nominations (are women named more often than men?), and the extent to which homophily on the trait plays a role (are same-gender social ties more frequent than opposite-gender?).

To measure similarity for continuous traits we took the absolute value of the difference between ego's and alter's measures and then coded each dyad as 1 if it was below the median difference in values and 0 if it was above. Because each ego was represented in the dataset multiple times, we adjusted the standard errors by using a general estimating equation, clustering on the ego. Finally, all models included a binary control variable for "*same household or not*" for each dyadic observation to adjust for the fact that certain traits such as latrine ownership will be the same for those in the same household, while others traits such as education maybe very different.

## 4. Results

### 4.1. Descriptive statistics

The number of unique egos and alters differed substantially according to the name generator (see Table 1). While the total number of unique individuals in the dataset numbered 16,935, the largest number of egos that named at least one social contact for a given name generator was 14,090 (*visit their home*) and the largest number of alters named was 10,605 (also *visit their home*). This name generator also had the largest number of total observations, (24,674) which differed substantially from that with the least number of observations (*go to temple with*) in which there were only 7529 total observations.

Table 1 shows the average total degree of respondents across each name generator as well as for the overall network (*total degree* is the total number of social contacts including both those the respondent nominated and those who nominated the respondent). While the average total degree for individual name generators ranged from 1.73 to 3.34 the average total degree for the complete network was 9.55. Fig. 1 is a histogram of out-going nominations for the complete network.

Table 1 shows *there were substantial differences in the demographics of nominated individuals depending upon the name generator*. For instance, while women made up only 17% of alters nominated in the *related to* name generator, they represented 63% of nominated alters in the *go to temple with* category. The *related to* category elicited the names of the oldest egos (mean 46) versus the *go to temple with* category in which the alters were the youngest (mean 36). Other differences became more apparent after stratifying the sample. For instance, women were very likely to nominate another woman as someone from whom they would *borrow money* (79% of alters), while men were very unlikely to nominate a woman as someone from whom they would *borrow money* (5% of alters). Conversely, 86% of the alters nominated by men as someone with whom they would *go to the temple* were women.

Caste also played a role in nominations. Although the likelihood of being nominated was roughly the same regardless of caste across the 12 name generators, the source of the nominations reveals sharp differences. High caste individuals almost never nominated a scheduled caste individual as someone with whom they would *go to the temple* (0% of alters). However, they did sometimes nominate a scheduled caste person as someone with whom they *talk* (6% of alters).

#### 4.2. Overlap between name generators

Table 1 shows what fraction of all unique social ties generated by any of the 12 name generators (the “full network”) was captured by each name generator independently. *Visit their home* yielded the most coverage, generating 32% of the full network, while *go to the temple* with yielded the least (9%). The majority of the name generators each captured only between 20% and 30% of the full network with a quarter of them capturing less than 20%.

Tables SA1 and SA2 show the extent to which different name generators tended to produce the same network ties. In many cases there was strong overlap. For instance, the *visit their home* and *invite home* name generators were highly correlated (Pearson’s  $r = 0.73$ ) as were *borrow rice from* and *lend rice to* ( $r = 0.71$ ) and *borrow money from* and *lend money to* ( $r = 0.62$ ), suggesting that for any of these pairs one question could potentially be omitted without biasing the network substantially (Fig. 2).

To illustrate how all 12 name generators were related to one another, Fig. 3 shows a correlation plot that orders the name generators utilizing a hierarchical clustering algorithm that places each question near other name generators that tend to elicit similar networks. These results yielded several conceptual groups that we subsequently used to group the name generators. These included (1) discussion: *talk to*; (2) domestic interactions: *visit their home*, *invite home*; (3) domestic resource exchange: *borrow kerosene/rice from*, *give kerosene/rice to*; (4) instrumental resource exchange: *borrow money from*, *lend money to*, *help during emergency*; (5) personal resource exchange: *give advice to*, *take advice from*; (6) spiritual interactions: *go to temple with*; and (7) *related to*.

We then subsetting the dataset using just one name generator from each conceptual group to generate a combination network comprised of the connections from 7 of the 12 name generators in the original dataset (see Fig. 2). With this subsetting network, we retained 71,795 of the 80,838 connections (89%) in the full network (see Table 1). Average degree in the subsetting network was 8.44 compared to 9.55 in the full network. Likewise, demographic characteristics in the reduced network were very close to those in the complete network with little if any variation across caste, gender, age, and education. These results show that, consistent with hypothesis 1, a great deal of the full network can be recovered with many fewer name generators, possibly saving time in network surveys.

#### 4.3. Ego and alter characteristics

Our regression models provided us with measures of the likelihood that alters with specific characteristic were nominated, controlling for the same characteristic measure for ego, and the similarity between ego and alter on that measure. Fig. SA1 shows the odds that an alter was nominated conditional on selected traits compared to an alter without that trait. Fig.

SA2 also provides comparable plots for ego's characteristics. A notable pattern in the results is that in almost all cases *the average nominated alter was of higher status than the average nominating ego*. We can see from the plot on gender, for instance, that across all name generators *men were more likely to be nominated than women*. For example, men were much more likely to be nominated as a *relative* (OR 7.10 95% CI 7.05–7.15) than were women. Men were also more likely than women to be the person to whom an ego would go *to borrow money* (OR 2.94 95% CI 2.77–3.12), *take advice from* (OR 2.77 95% CI 2.67–2.88) and or for *help during an emergency* (OR 2.80 95% CI 2.64–2.97). The only exceptions were that women were more likely than men to be nominated as someone with whom an ego might *go to temple with* (OR 1.90 95%, CI 1.84–1.96), and as someone to whom an ego might *lend rice* (1.34 95% CI 1.30–1.38).

General caste individuals were also more likely than scheduled caste individuals to be nominated for almost every name generator, with the exception of to whom you *lend rice* (OR 0.85 95% CI 0.79 to 0.92). Again, as with men versus women, the high caste alters were more likely to be nominated in the case of seeking resources than were scheduled caste alters: *borrow money from* (OR 1.82 95% CI 1.73 to 1.92), *take advice from* (OR 1.92 95% CI 1.77–2.07), and *help during an emergency* (OR 1.46 95% CI 1.35–1.58).

#### 4.4. Hierarchical vs peer nominations

To further explore the difference in characteristics between ego and alter, we calculated a network based measure of status difference using *in-degree* (the number of in-coming nominations), a commonly used measure of social status (Bondonio, 1998; Strauss and Pollack, 2003). If higher status individuals receive more nominations, then the difference in in-degree between ego and alter should indicate the difference in their status. Moreover, the mean difference across all ego-alter pairs should indicate which name generators yield peer relationships (low difference in status) and which yield hierarchical relationships (high difference in status).

Fig. 4 shows that for all 12 name generators, egos were likely to nominate those who had more friends than themselves. This is a consequence of the well-known “friendship paradox” (Feld, 1991), which states that the average number of friends of friends in a network is higher than the average number of friends in any network with variable degree. Nevertheless, there was substantial variation in the measure suggesting that, as predicted in hypothesis 2, *some name generators yield peer relations* (for example *talk to* and *visit their home*) *while others yield hierarchical relationships* (for example *take advice from* and *give advice to*). To test the plausibility of this interpretation, we compared the measure to other measures of status differences. Fig. SA3 shows in-degree differences plotted against status differences for three different demographic traits: gender, poverty status, and age. In all cases we can see a strong positive relationship between the difference in network status and the difference in sociodemographic status. Name generators that elicited more nominations from women to men, from the poor to the rich, and from the young to the old also tended to generate more nominations between the poorly and well connected.

#### 4.5. Homophilic and heterophilic ties

Our regression models yielded a measure of homophily on each trait for each name generator (see Figs. SA3–SA5 for plots of all models and Tables SA4–SA19 for regression tables). *Across all traits, nominations were homophilic*—;egos were more likely to nominate similar alters. Fig. 5 shows that the degree to which trait-homophily predicted ties differed substantially depending upon the trait. Overall homophily (measured across all of the 12 name generators) was strongest for sociodemographic traits like gender, caste, and religion. The regression models suggest that for each of these traits the odds of naming a similar-trait social contact was about four to six times higher than naming a different-trait social contact. Behavioral traits also exhibited homophily, but at much lower levels with odds ratios ranging from 1.2 to 2.2.

We can also characterize which *name generators* yield social ties that are more likely to be homophilic. Fig. 6 shows three plots, one that averages homophily measures across all 16 traits, and two that average across behavioral and sociodemographic traits. These results show that the *discussion and domestic interaction* name generators (*talk to*, *invite home*, *visit their home*) and the *financial exchange* name generators (*borrow money from*, *lend money to*) yielded the most homophilic ties across both demographic and behavioral traits. Meanwhile, the *give advice to*, *take advice from*, and *go to temple with* name generators elicited the networks with the most heterophilic ties (those in which the egos and alters are dissimilar). These results are consistent with hypothesis 3.

If we look at traits individually however, we can see that the detailed picture is more nuanced. For instance while *talk to* was one of the name generators that elicited the most *homophilic* ties overall, it elicited the most *heterophilic* ties across some of the most *homophilic traits*: namely caste, religion, and mother tongue, more so than either *visit their home* or *invite home*.

#### 4.6. Network characteristics

Recent work has shown that when we consider the relationship between networks and behavior it is not simply the relationships between pairs of individuals that should be of interest, but we must also consider the structure of the network itself (Shakya et al., 2014, 2015). The ability for behaviors to change, diffuse, and become established within a community can greatly affect how that community is connected. Likewise, if there are many intertwining connections, knowledge and behavior may change quite differently than it will in a community that is highly segmented or sparsely connected. To understand the difference in the structural characteristics of the networks generated by all 12 name generators, we calculated 6 different network measures across all 75 villages for each name generator. Fig. SA6 shows the averages of 6 different network measures across all 75 villages for each name generator. Network transitivity was highest for *related to* (0.19 95% CI 0.17–0.20). Networks with the next highest transitivity measures were those based on *domestic resource exchange* questions (*borrow rice from* 0.14 95% CI 0.13–0.15 and *give rice to* 0.14 95% CI 0.13–0.15), and *domestic interaction* questions (*visit their home* 0.13 95% CI 0.12–0.14 and *invite home* 0.13 95% CI 0.12–0.14), suggesting that these specific name generators also elicited networks with a high amount of clustered relationships. The *domestic interactions*

networks also showed the highest disparity between those with the highest centrality scores (both closeness and betweenness) and those with the lowest. The *advice networks* on the other hand had some of the lowest closeness and betweenness centrality scores. Contrary to hypothesis 4, the *related to* network was neither the densest nor the most highly reciprocated. The *advice networks* on the other hand were highly reciprocated, although the most highly reciprocated ties were those identified through the *go to temple with* question.

## 5. Discussion

This is one of the first studies to analyze network survey methods in the developing world, and it is also among the first to attempt to identify general rules about what we can expect from the kinds of networks that are elicited by different “name generator” questions. Both in the main text and in the supplement we provide numerous details on these name generators and the network and individual characteristics with which they are associated, but here we summarize what we think are the most important insights from the data.

First, *different name generators can yield very different sets of alters with very different distributions of demographic characteristics*. As an example, while the *related to* name generator tended to identify older male alters, the *go to temple with* name generator identified the opposite: younger female alters. These differences become more pronounced as we considered the demographics of the ego. If we had limited our analysis to the *talk to* name generator we would have concluded that men have virtually no interactions with women (only 3% of alters). Comparing across name generators, however, it is apparent that the *talk to* name generator evinced the lowest proportion of male to female nominations of any question. *Go to the temple with*, on the other hand, brought forth the names of many female alters, especially among male egos (86% of alters), as did *give advice to* (34%) and *borrow rice from* (29%). These results suggest that using only one name generator may not capture the many dimensions of relationships inherent in a real world village network and may bias the observed relationships toward individuals of a certain demographic background.

Second, as hypothesized, *it may be possible to ascertain the full range of network interactions with only a few name generators*. While efficiency in data collection and conservative use of valuable research resources are goals that might motivate researchers to use the smallest number of name generators possible, it is not obvious at the onset how that can be achieved without potentially biasing the network in terms of size and composition. The hierarchical clustering algorithm we used on the network provided us with 7 clearly demarcated categories of name generators, 3 of which (*discussion, spiritual interactions, and relations*) stood alone. These categorizations were essential to our exercise of honing down the network to a smaller version while retaining the demographic dimensions of the original. The groupings are consistent with what would intuitively be expected of some of these questions, particularly ones like *domestic resource exchange (borrow rice from and lend rice to)* in which the behaviors are most likely reciprocal. This exercise also provided us strong evidence of functional specificity. In other words, questions uncovered functional relationships that could be grouped together with other questions from which the observed relationships served a similar function.

Third, *a stand-alone discussion question may be insufficient for capturing a general picture of the full network*. It is interesting that the *talk to* question stood alone, as it would seem reasonable to group together with *visit their home* and *invite home*, to which it is adjacent in the correlation matrix. Like *talk to*, *visit their home* and *invite home* also captured interactions that might be primarily social in nature, while excluding same household nominations (the *talk to* name generator specifically instructed participants to exclude household members). However, the homophily figures in the SA (Fig. SA 4) demonstrate that the *talk to* name generator uncovered the names of some of the least homophilic relationships for strongly homophilic characteristics such as caste, religion, and language, in contrast to *visit their home* and *invite home* which tended to uncover more homophilic relationships along those dimensions. People to whom respondents *talk* were also more highly educated, and much more likely to be the same gender as those nominated in the *visit their home* and *invite home* questions. These results suggest that common stand-alone name generators like the “discuss important matters” question may be particularly problematic in places like rural India where social connections depend on a wide range of complex interactions (Roland, 1991).

Fourth, *it is probably necessary to ask multiple name generator questions to capture the full network*. Researchers engaged in work for which an accurate representation of the size, structure, or demographic composition of the network is important will need to carefully consider how many and specifically which name generators they use to collect their data. The highest proportion of the entire network captured by any one name generator was 32%, with the majority of the name generators capturing between 20% and 30%. Furthermore, while the highest average degree of any name generator on its own was 3.2 (*invite home*), the average for the full network was 9.6, a full 3 times greater. Considering this in conjunction with the demographic variation across name generators, it is likely that asking only one name generator risks grossly underestimating the full network both in terms of size and characteristics. Using one question from each of the conceptual clusters that we created, however, we were able to considerably decrease (by 42%) the number of name generators necessary in order to create a reduced network that encompassed 89% of the total network while maintaining similar composition in terms of demographic characteristics.

Fifth, *people tended to nominate people with whom they are similar (homophily) across all name generators, but there are important differences across cultures*. The fact that people tend to nominate those to whom they are similar has been well-established in previous work (McPherson et al., 2001), but this work has primarily focused on the United States. In our rural Indian sample here we found that the strongest predictors of ties overall were individual demographic characteristics such as gender, religion, and caste. While in the US, adult relationships are relatively well integrated between the genders, in this study homophily on gender was the strongest predictor of social ties among all of the individual characteristics, with the highest average homophily score of any characteristic we measured. Although homophily on mother tongue was a strong predictor of social ties, on average it was weaker than homophily on religion or caste. In the US, social status is arguably a byproduct of higher levels of income and education, however in India caste is a characteristic considered intrinsic to an individual, similar to one’s race (Berreman, 1960) which, in the US, is one of the most highly homophilic traits (McPherson et al., 2001). In the

US there is a strong tendency toward homophily on education and income levels. In this study, education and income were the individual traits least likely to predict ties and in fact they had lower average values on our homophily measures than some of the behavioral traits.

Identifying homophilic social ties is important because similarity between people can reinforce norms (Kitts, 2006). It may be much more difficult to change a behavior that is entrenched within a group of very similar people who are socially connected. An intervention strategy that targets socially connected yet demo-graphically different people might be more effective for highly normative behaviors. On the other hand targeting socially connected people as a group, rather than as individuals who happen to share similar demographic characteristics, can potentially increase the efficacy of a behavioral intervention if the group as a whole can come to a consensus about the need to change the norm in question.

Sixth, *name generators tended to either reveal peer relationships or hierarchical relationships between lower and higher status individuals*. Despite the bias toward homophily, network questions tended to generate the names of alters who were relatively higher in status, either economically or socially, compared to the general population. Named alters were more likely to be male than female, high caste than lower caste, non-impo-verished than impoverished, and they were more likely to own an election card and be older than average. Although there was a general tendency to nominate people of higher status no matter what the question, this effect was markedly less pronounced for questions that were eliciting the names of alters to whom the ego would provide resources (such as *lend money to* or *lend rice to*). Moreover, there was a strong relationship between the average social status of alters nominated using a particular name generator and their *network status*, measured by their number of inward nominations. This suggests a rich-gets-richer phenomenon, with higher status individuals tending to occupy highly central locations in the network. It also indicates that researchers should be careful to consider the status -bias their name generators might create in their ascertainment of the network.

Seventh, *different name generators yielded networks with different structure*. For example, questions related to domestic encounters including *domestic resource exchange (borrow rice and lend rice)* and *domestic interaction (visit their home and invite home)* generated the most transitive network. Interestingly, both of these sets of questions were categorized in our status measure as peer nomination questions. These results suggest that such questions are likely to reveal tightly clustered groups of individuals interacting within a context of relative equality. Norms within these groups may be tightly held, but new behaviors, once accepted by a critical mass of the network, may diffuse quickly. We also found that some networks were highly centralized compared to others. The questions that generated networks with the most disparities between a few well-connected individuals and the rest of the community were those focused on social interaction: *talk to*, and again *domestic interactions*.

The fact that the *domestic interaction* questions were both some of the most transitive and the most centralized is interesting because it suggests these questions could be particularly useful for diffusion by combining approaches that utilize the most connected individuals and

exploit the highly clustered nature of the network for reinforcement. On the other hand, these are also questions that tended to uncover highly homophilic relationships. When dealing with norms and behaviors that are deeply embedded in the network, tightly clustered homophilic networks in which a few influential people have access to the most pathways across the network may be hard to shift. In such cases, it may be beneficial to consider network pathways drawn by the *personal resource exchange* questions (*give advice to* and *take advice from*), which in this study were the least transitive, the least centralized, and the most likely to elicit the names of ties across trait and network status lines.

Eighth, *some name generators stood apart*. For example, the *go to temple with* question elicited the smallest number of alters from the smallest number of egos; it was the most likely to generate the names of female alters, even among male egos; and these ties were also the most heterophilic. In fact the network elicited was so different that it did not cluster with any of the other name generators. It is also interesting to note that this question elicited the most highly reciprocated ties, and the least centralized network. But interpretation of this name generator is potentially problematic because of its cultural specificity. In the cultural setting of Hindu South Asia, activities related to daily rituals are often the duty of younger women, particularly new daughter-in-laws (Bennett, 1983). It is unlikely that the specific type of network generated by this question would be duplicated by a question such as “with whom do you go attend church” in rural Central America for instance. In the context of this study, however, the *go to temple with* question was able to elicit the names of alters who, for the most part, were not named in other name generators. So while the total number of alters nominated was the lowest of all name generators, many of these relationships were unique to the network.

In contrast to the *go to temple with* question, the *related to* question yielded relationships that were highly homophilic for demographic traits, but not for behavioral traits. Furthermore, the *related to* question tended to elicit the names of older male relatives and was also at the high end of the social status measure. In other words, nominated alters were receiving many more nominations than were the egos who were nominating them. This is confirmed by the relatively low reciprocity measure for a question that should be measuring an objective tie. The *related to* question is then problematic within the scope of this study as the results are not consistent with what would be the common sense goal of such a question, which would be to uncover the connections between non-household family members. The question, which should be easily generalizable across settings, may have been asked in a way that generated a regionally specific interpretation. This is most likely a cultural phenomenon, specific to the context of rural South Asia, as respondents seem to have interpreted this question as asking for high status family members, (in this context older men), rather than family members in general. Future qualitative research can investigate in what way this question could be reworded in order to capture a more representative group of related alters.

### 5.1. Limitations of this study

Because the data for this study is cross-sectional we were not able to observe the way in which behavioral trends could be predicted by the different name generators. It would be a



very valuable contribution to network literature, for instance, to differentiate the specific questions most likely to elicit the names of people who *share behavioral traits with each other* from those most likely to elicit the names of people *who bring about behavioral change in one another*. Another important limitation is that the study is exclusive to rural South India, and therefore extrapolating these results to other communities may be problematic. Optimally we would like to be able to compare the networks generated by these same questions to those generated in other cultural settings in order to analyze the differences and similarities of the networks generated depending upon the local context. Nevertheless, it is likely that some of the culturally specific trends we observed in this study are applicable to other tightly knit, hierarchical, and male-dominated cultural contexts. Finally, because this data was collected as part of a study hoping to uncover determinants of micro-finance diffusion, there were no affective questions asked of the participants. We cannot therefore use affective relations as a comparison in our analyses.

On that note, it is important to remember that researchers hoping to use these name generators in different contexts will have to consider the wording and translation of the questions. For instance domestic exchange questions may need to ask about exchange of cornmeal in the mountains of Honduras, rather than the exchange of rice. The unique results generated by the *go to temple with* question may be specific to rural Hindu villages in India. A comparable question regarding church going in Guatemalan villages may elicit a completely different sort of network. Questions regarding reciprocal visits to the homes of others may elicit the smallest and most intimate networks in places that are geographically isolated and home visitation is more difficult. This raises the question, then, of how contextually specific a name generator should be. While very contextually specific name generators may capture dimensions of the network that cannot be captured by more general questions (the “discuss important matters” question for instance), questions that are very specific are also not easily comparable across settings.

It is potentially beneficial to network science as a whole to ask questions that are comparable so that network data gathered in unique settings around the world can be compared. As evident in our analysis however, the most general questions (in this case *talk to*) may be insufficient in capturing the unique dimensions of the network across an important range of social dynamics. This is not an issue that has one clear solution and the answer may only become clear as more network studies are conducted within unique settings, with qualitative work contributing to the understanding of how name generators are interpreted within each one. Future research should investigate these differences so that as network science progresses across the world the tools with which to conduct analyses are the most comparable across studies. Will a question on lending cornmeal in Honduras generate a substantially different network than the question on lending rice that was asked in India? And if so, can we attribute that to genuine differences in the actual networks or as simply an artifact of how the question was asked? The challenge will be to develop a toolbox of questions that are comparable across settings but appropriate to local contexts. The present analysis helps to address this larger problem, but significant work remains in understanding the complexity of human interactions in sociocentric settings.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Appendix A. Supplementary data

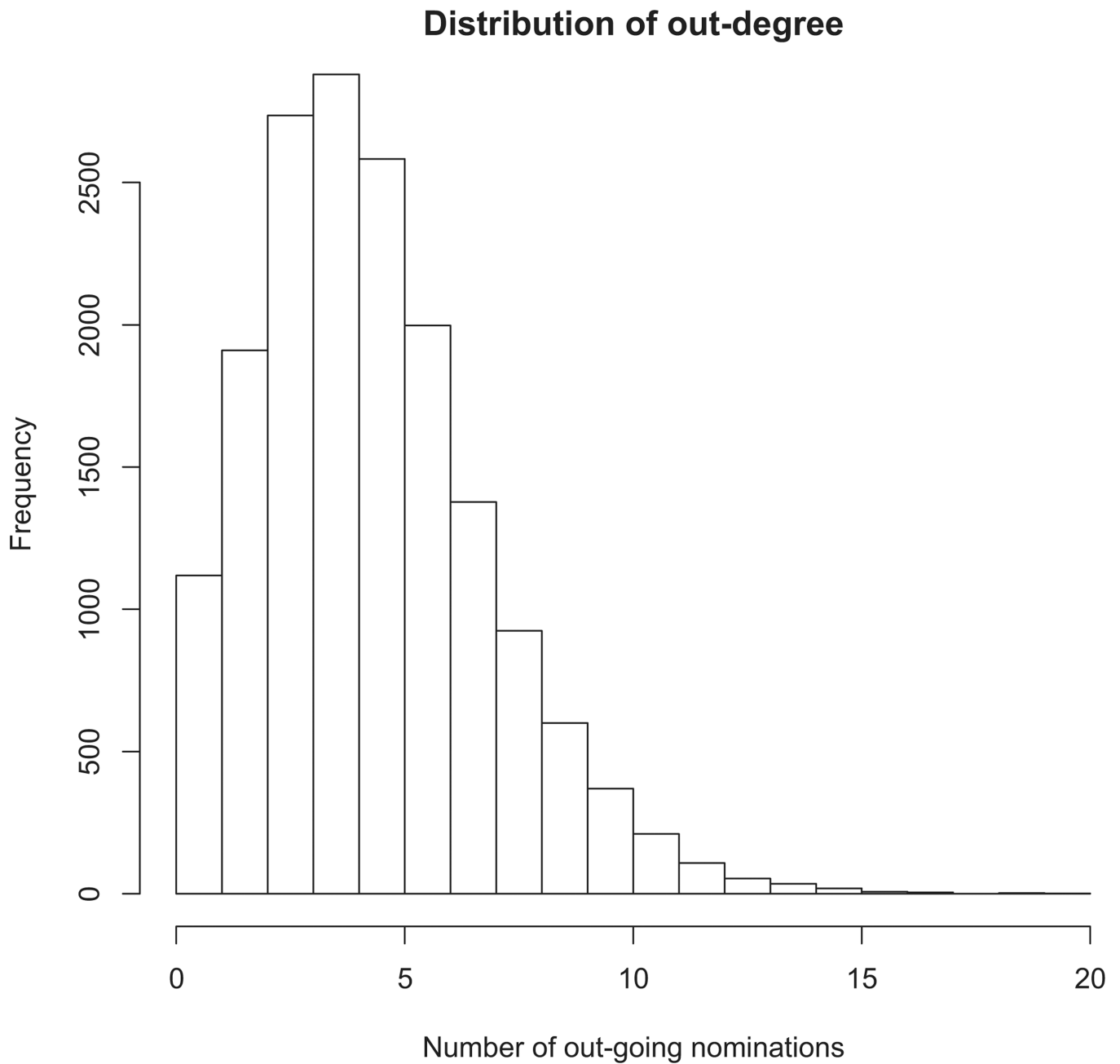
Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.socnet.2016.08.008>.

## References

- Adams AM, Madhavan S, Simon D. Women's social networks and child survival in Mali. *Soc. Sci. Med.* 2002; 54:165–178. [PubMed: 11824923]
- Apicella CL, Marlowe FW, Fowler JH, Christakis NA. Social networks and cooperation in hunter-gatherers. *Nature*. 2012; 481:497–501. [PubMed: 22281599]
- Banerjee A, Chandrasekhar AG, Duflo E, Jackson MO. The diffusion of microfinance. *Science*. 2013:341.
- Barrington C, Latkin C, Sweat MD, Moreno L, Ellen J, Kerrigan D. Talking the talk, walking the walk: social network norms, communication patterns, and condom use among the male partners of female sex workers in La Romana, Dominican Republic. *Soc. Sci. Med.* 2009; 68:2037–2044. [PubMed: 19356834]
- Bennett, L. *Dangerous Wives and Sacred Sisters*. New York: Columbia University Press; 1983.
- Bernard HR, Johnsen EC, Killworth PD, McCarty C, Shelley GA, Robinson S. Comparing four different methods for measuring personal social networks. *Soc. Netw.* 1990; 12:179–215.
- Bernard HR, Shelley GA, Killworth P. How much of a network does the GSS and RSW dredge up? *Soc. Netw.* 1987; 9:49–61.
- Berremán GD. Caste in India and the United States. *Am. J. Sociol.* 1960:120–127.
- Bien W, Marbach J, Neyer F. Using egocentered networks in survey research. A methodological preview on an application of social network analysis in the area of family research. *Soc. Netw.* 1991; 13:75–90.
- Bondonio D. Predictors of accuracy in perceiving informal social networks. *Soc. Netw.* 1998; 20:301–330.
- Burt RS. Network items and the general social survey. *Soc. Netw.* 1984; 6:293–339.
- Burt RS. A note on social capital and network content. *Soc. Netw.* 1997; 19:355–373.
- Campbell KE, Lee BA. Name generators in surveys of personal networks. *Soc. Netw.* 1991; 13:203–221.
- Centola D. An experimental study of homophily in the adoption of health behavior. *Science*. 2011; 334:1269–1272. [PubMed: 22144624]
- Centola D, Macy M. Complex contagions and the weakness of long ties. *Am. J. Sociol.* 2007; 113:702–734.
- Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. *N. Engl. J. Med.* 2008; 358:2249–2258. [PubMed: 18499567]
- Cobb NK, Graham AL, Abrams DB. Social network structure of a large online community for smoking cessation. *Am. J. Public Health*. 2010; 100:1282. [PubMed: 20466971]
- De Lange D, Agneessens F, Waage H. Asking social network questions: a quality assessment of different measures. *Metodoloski zvezki*. 2004; 1:351–378.
- DiMaggio P, Garip F. How network externalities can exacerbate intergroup inequality. *Am. J. Sociol.* 2011; 116:1887–1933.
- Feld SL. Why your friends have more friends than you do. *Am. J. Sociol.* 1991; 96:1464–1477.
- Ferligoj A, Hlebec V. Evaluation of social network measurement instruments. *Soc. Netw.* 1999; 21:111–130.
- Fischer CS. What do we mean by 'friend'? An inductive study. *Soc. Netw.* 1982; 3:287–306.

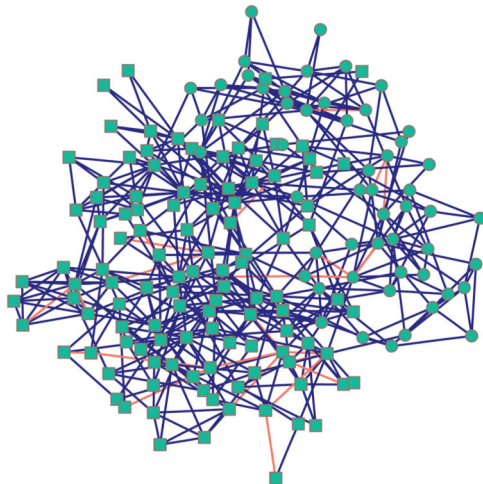
- Fu Y-C. Measuring personal networks with daily contacts: a single-item survey question and the contact diary. *Soc. Netw.* 2005; 27:169–186.
- Granovetter M. The strength of weak ties: a network theory revisited. *Sociol. Theory.* 1983; 1:201–233.
- Hruschka, D. *Friendship: Development, Ecology, and Evolution of a Relationship.* Berkeley, CA: University of California Press; 2010.
- Jackson MO, Rodriguez-Barraquer T, Tan X. Social patterns and social quilts: network patterns of favor exchange. *Am. Econ. Rev.* 2012; 102:1857–1897.
- Kim DA, Hwang AR, Stafford D, Hughes DA, O'Malley AJ, Fowler JH, Christakis NA. Social network targeting to maximise population behaviour change: a cluster randomised controlled trial. *Lancet.* 2015
- Kitts JA. Social influence and the emergence of norms amid ties of amity and enmity. *Simul. Modell. Practice Theory.* 2006; 14:407–422.
- Knipscheer, C., Antonucci, T. *Network Research: Substantive Issues and Methodological Questions.* Amsterdam: Swets and Zeitlinger; 1990.
- Latkin CA, Hua W, Tobin K. Social network correlates of self-reported nonfatal overdose. *Drug Alcohol Depend.* 2004; 73:61–67. [PubMed: 14687960]
- Lerner J, Jessica Lubbers M, Luis Molin J, Brandes U. *Social Capital Companion: Capturing Personal Networks As They ARE Lived.* GRAFO Working Papers. 2014:3.
- Marin A. Are respondents more likely to list alters with certain characteristics? Implications for name generator data. *Soc. Netw.* 2004; 26:289–307.
- Marin A, Hampton KN. Simplifying the personal network name generator. *Field Methods.* 2007; 19:163.
- Marin, A., Wellman, B. Social network analysis: an introduction. In: Scott, J., Carrington, P., editors. *The SAGE Handbook of Social Network Analysis.* Thousand Oaks, CA: Sage; 2011. p. 11
- McCallister L, Fischer CS. A procedure for surveying personal networks. *Sociol. Methods Res.* 1978; 7:131–148.
- McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: homophily in social networks. *Annu. Rev. Sociol.* 2001:415–444.
- Milardo RM. Theoretical and methodological issues in the identification of the social networks of spouses. *J. Marriage Fam.* 1989:165–174.
- Milardo RM. Comparative methods for delineating social networks. *J. Soc. Personal Relatsh.* 1992; 9:447–461.
- Morris SS, Carletto C, Hoddinott J, Christiaensen LJM. Validity of rapid estimates of household wealth and income for health surveys in rural Africa. *J. Epidemiol. Community Health.* 2000; 54:381–387.
- Parigi P, Bearman PS. Cloning headless frogs and other important matters: conversation topics and network structure. *Soc. Forces.* 2005; 83:535–557.
- Perkins JM, Subramanian S, Christakis NA. Social networks and health: a systematic review of sociocentric network studies in low- and middle-income countries. *Soc. Sci. Med.* 2015; 125:60–78. [PubMed: 25442969]
- Perry BL, Pescosolido BA. Functional specificity in discussion networks: the influence of general and problem-specific networks on health outcomes. *Soc. Netw.* 2010; 32:345–357.
- Pustejovsky JE, Spillane JP. Question-order effects in social network name generators. *Soc. Netw.* 2009; 31:221–229.
- Roland, A. Princeton University Press. 1991. *In Search of Self in India and Japan: Toward a Cross-Cultural Psychology.*
- Ruan D. The content of the General Social Survey discussion networks: an exploration of General Social Survey discussion name generator in a Chinese context. *Soc. Netw.* 1998; 20:247–264.
- Shakya HB, Christakis NA, Fowler JH. Association between social network communities and health behavior: an observational sociocentric network study of latrine ownership in rural India. *Am. J. Public Health.* 2014; 104:930–937. [PubMed: 24625175]

- Shakya HB, Christakis NA, Fowler JH. Social network predictors of latrine ownership. *Soc. Sci. Med.* 2015; 125:129–138. [PubMed: 24726688]
- Straits BC. Ego's important discussants or significant people: an experiment in varying the wording of personal network name generators. *Soc. Netw.* 2000; 22:123–140.
- Strauss RS, Pollack HA. Social marginalization of overweight children. *Arch. Pediatr. Adolesc. Med.* 2003; 157:746–752. [PubMed: 12912779]
- Tobin KE, Latkin CA. An examination of social network characteristics of men who have sex with men who use drugs. *Sexually Transm. Infect.* 2008; 84:420–424.
- Uzzi B, Amaral LA, Reed-Tsochas F. Small world networks and management science research: a review. *Eur. Manage. Rev.* 2007; 4:77–91.
- Valente, T. *Social Networks and Health: Models, Methods, and Applications*. New York, NY: Oxford University Press; 2010.
- Valente T, Hoffman BR, Ritt-Olson A, Lichtman K, Johnson CA. Effects of a social-network method for group assignment strategies on peer-led tobacco prevention programs in schools. *Am. J. Public Health.* 2003; 93:1837. [PubMed: 14600050]
- Valente T, Ritt-Olson A, Stacy A, Unger JB, Okamoto J, Sussman S. Peer acceleration: effects of a social network tailored substance abuse prevention program among high risk adolescents. *Addiction.* 2007; 102:1804–1815. [PubMed: 17784893]
- Valente TW. Network interventions. *Science.* 2012; 337:49–53. [PubMed: 22767921]
- Van der Poel MG. Delineating personal support networks. *Soc. Netw.* 1993; 15:49–70.
- Wellman B, Wortley S. Different strokes from different folks: community ties and social support. *Am. J. Sociol.* 1990:558–588.



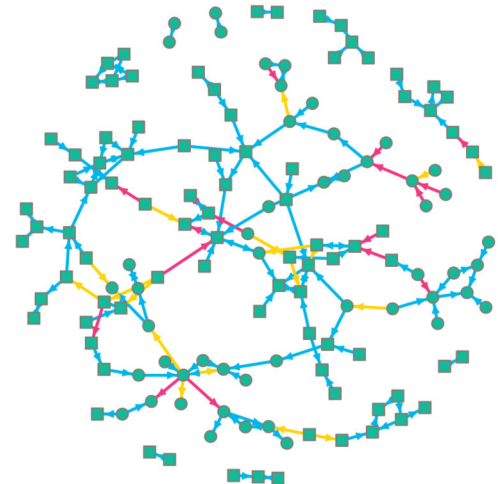
**Fig. 1.** A histogram of out-going nominations for the full network including ties elicited from all 12 name generators.

### Full network vs condensed network



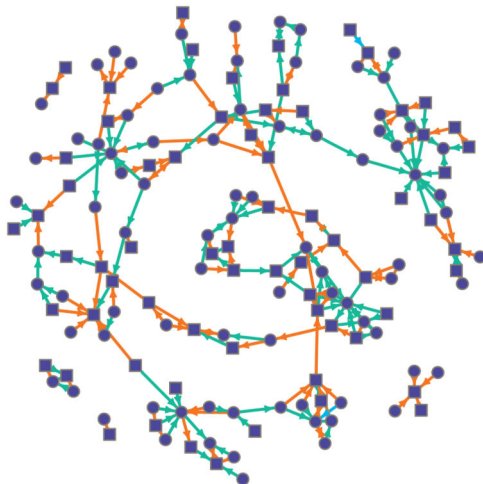
— Full network      ■ High caste  
— Condensed network      ● Low caste

### Invite home vs visit home



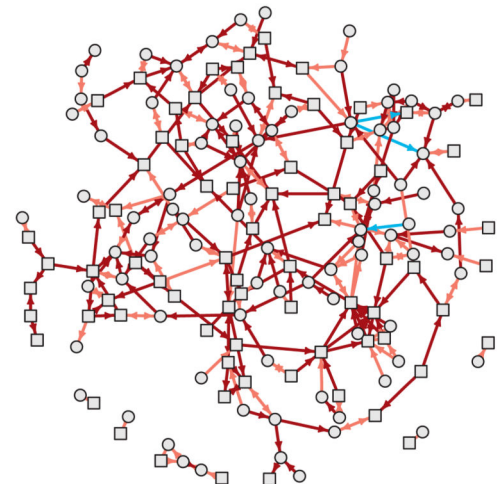
— Invite home      ■ High caste  
— Visit home      ● Low caste  
— Overlap

### Take advice from vs give rice to



— Take advice      ■ Men  
— Give rice to      ● Women  
— Overlap

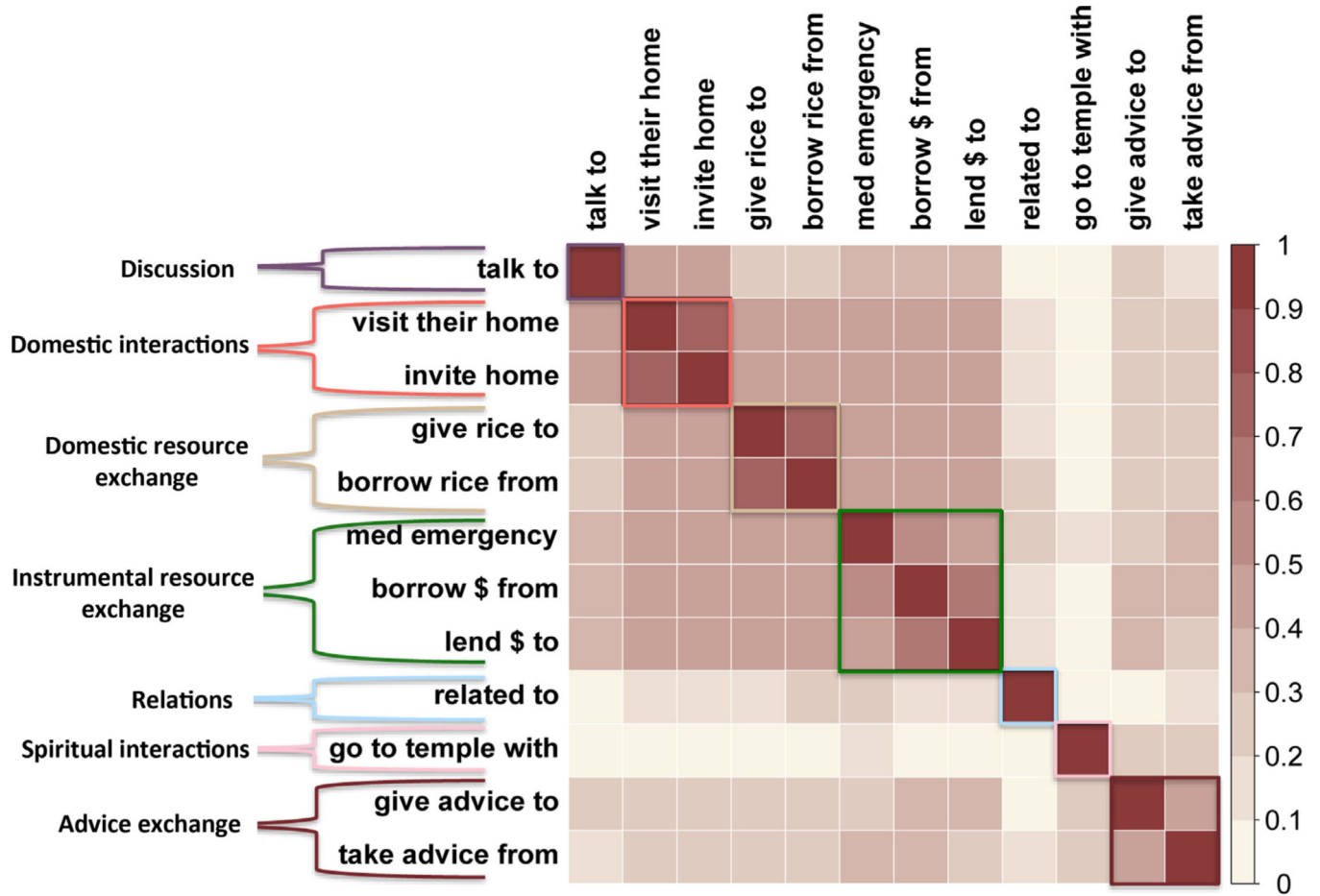
### Go to temple with vs socialize with



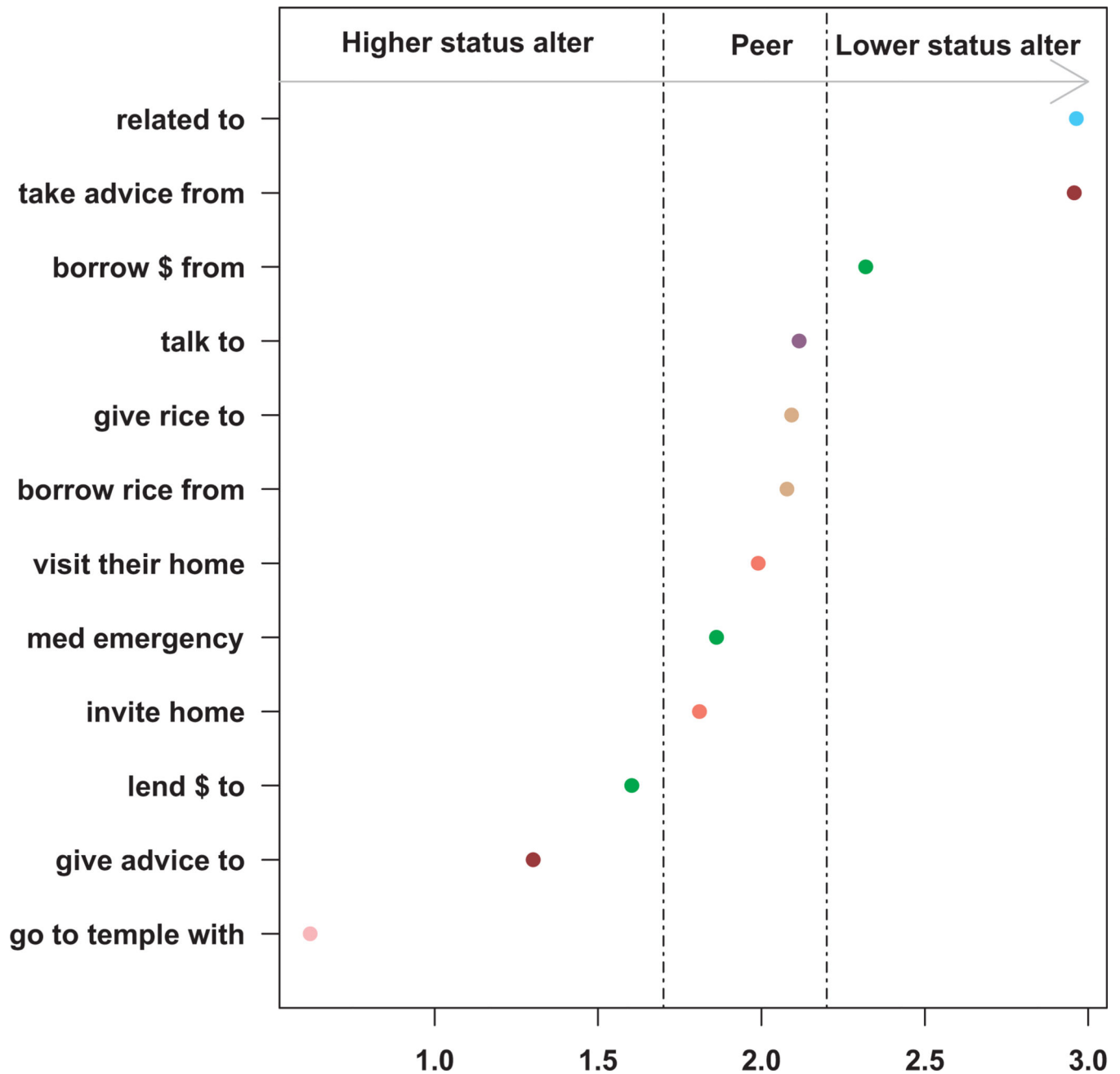
— Socialize with      ■ Men  
— Go to temple with      ● Women  
— Overlap

**Fig. 2.**

Network maps for one sample village, each comparing nominations from two networks in the dataset. It is clear from the first plot that using the condensed network retains much of the full network while substantially reducing the number of questions asked (from 12 down to 7). The subsequent 3 plots show the degree of overlap between nominations in specific name generators. While nominations from *invite home and visit their home* overlapped significantly, the nominations generated by *take advice from* and *give rice to*, for example, were substantially different.

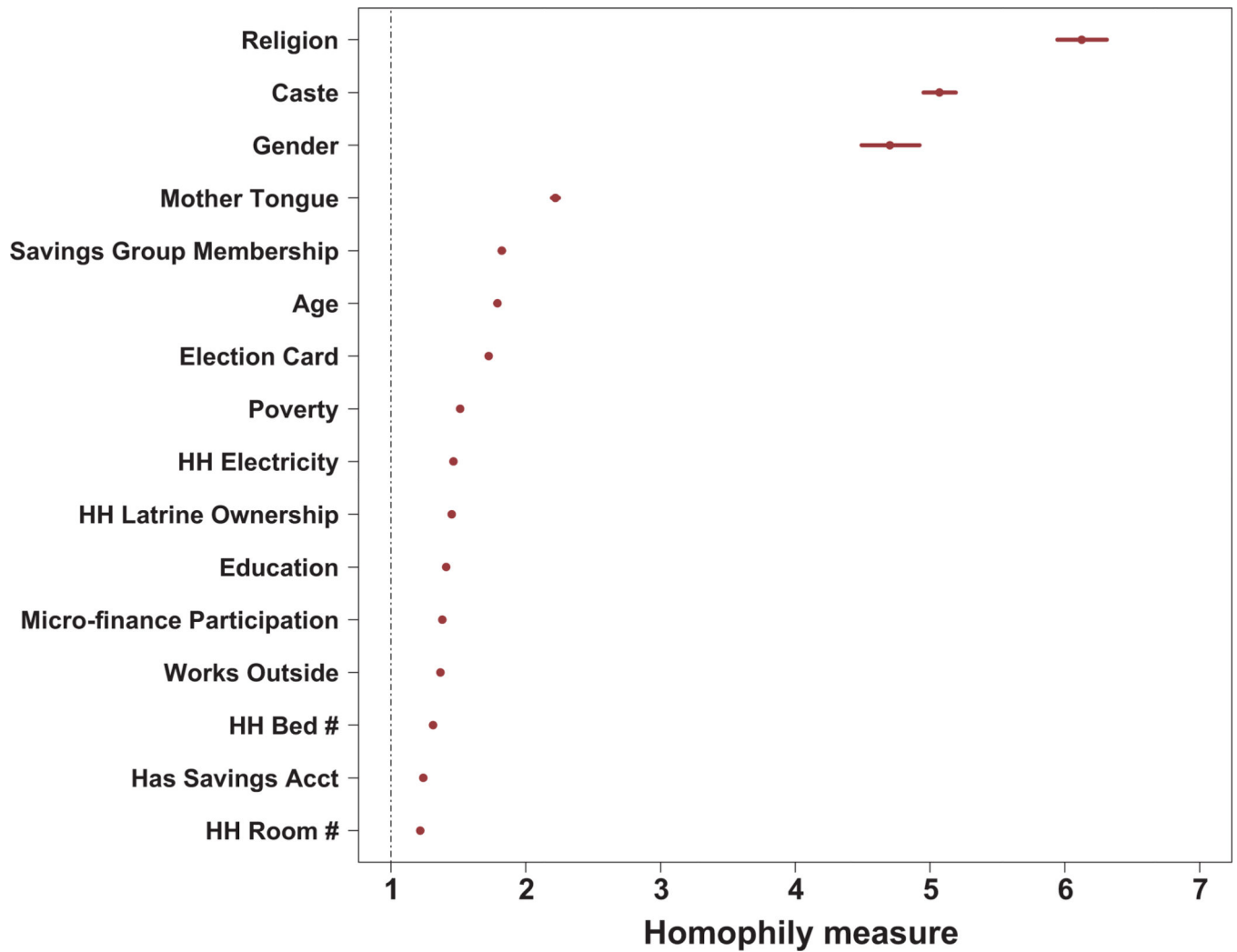


**Fig. 3.** Plot of the correlation matrix of nominations by name generator (see Table SA 2 for specific values). A hierarchical clustering algorithm was used to reveal conceptual groups that are labeled and color-coded.

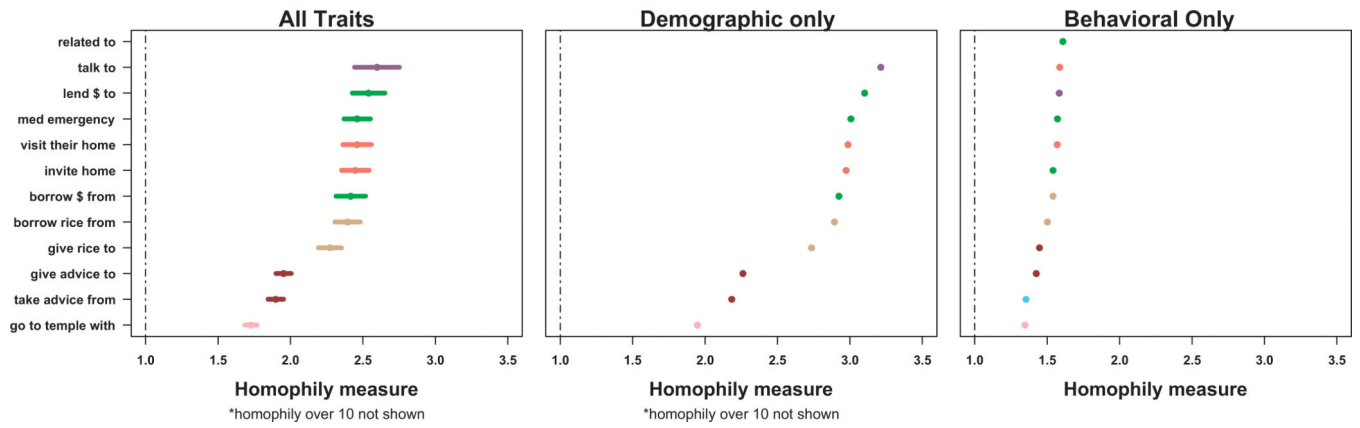


**Fig. 4.** This figure shows the mean difference between alters and egos in-degree measure for each name generator. Across all 12 name generators, egos nominated alters with higher in-degree scores than themselves, as predicted by the “friendship paradox” (Feld, 1991). Name generators with the lowest differences between egos and alters can be categorized as low status questions, those around the median can be categorized as peer questions, and those at the upper end as high status questions. Fig. SA 3 plots these differences against select demographic trait differences.





**Fig. 5.** Using an average of the homophily measures calculated for each trait across all 12 name generators, we compared the degree to which nominations within any one trait were homophilic. Here we see that by far the most homophilic traits were demographic while behavioral traits, although also homophilic, were much less so.



**Fig. 6.** An overview of the degree to which the average nominations among all traits across any one name generator are homophilic. Across all name generators demographic traits showed the most homophily in nominations. See Figs. SA4–SA6 for detailed plots of all traits across all name generators.

**Table 1**  
Demographic characteristics of both egos and alters across 12 different name generators asked of respondents in rural Karnataka India.

	Talk to	Visit their home	Invite home	Borrow rice from	Give rice To	Help during emergency	Borrow \$ from	Lend \$ to	Give advice to	Take advice from	Goto temple w/	Related to	Full network	Reduced network
Egos <i>N</i>	12,875	14,090	13,930	13,243	13,315	12,278	12,424	11,422	11,600	13,177	6736	10,324	16,763	16,699
Alters <i>N</i>	9636	10,605	10,511	9071	9182	8733	8823	8609	8929	9199	6213	5592	15,498	15,156
Total <i>N</i> (NA omitted)	20,927	25,871	24,674	21,109	21,682	17,546	18,677	16,315	14,697	18,228	7529	14,861	80,838	71,795
% of full network	26%	32%	31%	26%	27%	22%	23%	20%	18%	23%	9%	18%	89	89
Total degree (mean)	2.83	3.34	3.20	2.87	2.94	2.47	2.60	2.39	2.12	2.43	1.73	2.54	9.55	8.44
Same household (prop yes)	0	0	0	0	0	0.06	0.02	0.01	0.33	0.36	0.80	0.01	15	15
<i>Age</i>														
Ego age (mean)	39.43	39.16	39.13	38.98	38.92	39.18	39.23	39.44	39.93	38.65	36.6	38.78	38.99	38.83
Alter age (mean)	39.93	40.94	40.49	40.04	40.8	40.76	40.78	39.43	38.97	44.01	36.36	46.43	41.45	41.63
<i>Gender</i>														
Ego % female	0.55	0.56	0.56	0.56	0.56	0.55	0.55	0.55	0.54	0.57	0.62	0.54	0.55	0.55
Alter % female	0.50	0.49	0.52	0.62	0.55	0.43	0.46	0.52	0.51	0.36	0.63	0.17	0.45	0.45
Alter % female (female ego)	0.88	0.83	0.86	0.87	0.84	0.72	0.79	0.89	0.66	0.47	0.49	0.23	0.62	0.62
Alter % female (male ego)	0.03	0.06	0.09	0.29	0.19	0.07	0.05	0.08	0.34	0.23	0.86	0.09	0.24	0.24
<i>Education</i>														
Ego education (mean)	4.94	4.94	4.96	4.99	4.93	4.95	4.87	4.95	4.95	5.04	5.82	5.15	5.14	5.16
Alter education (mean)	5.35	4.99	4.85	4.36	4.68	5.33	5.55	4.93	5.29	5.32	5.77	4.73	5.09	5.04
<i>Caste</i>														
Ego SC %	0.25	0.25	0.25	0.25	0.25	0.25	0.26	0.25	0.25	0.25	0.24	0.24	0.25	0.25
Ego ST %	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.06	0.06	0.06	0.06	0.07	0.07
Ego OBC %	0.55	0.56	0.56	0.56	0.56	0.56	0.56	0.55	0.57	0.57	0.55	0.58	0.55	0.56
Ego GEN %	0.13	0.13	0.13	0.13	0.13	0.13	0.12	0.13	0.13	0.12	0.15	0.12	0.13	0.13
Alter SC%	0.23	0.23	0.24	0.26	0.24	0.23	0.22	0.25	0.24	0.22	0.23	0.25	0.24	0.24
Alter ST %	0.07	0.07	0.07	0.07	0.07	0.07	0.06	0.07	0.06	0.06	0.06	0.06	0.06	0.06
Alter OBC %	0.57	0.56	0.56	0.56	0.56	0.56	0.58	0.56	0.57	0.59	0.55	0.57	0.56	0.56
Alter GEN %	0.14	0.14	0.13	0.12	0.14	0.14	0.15	0.12	0.13	0.14	0.16	0.12	0.14	0.13

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

	Talk to	Visit their home	Invite home	Borrow rice from	Give rice To	Help during emergency	Borrow \$ from	Lend \$ to	Give advice to	Take advice from	Goto temple w/	Related to	Full network	Reduced network
Alter SC % (GEN ego)	0.06	0.04	0.03	0.05	0.02	0.02	0.03	0.06	0.02	0.02	0.00	0.01	0.04	
Alter ST % (GEN ego)	0.05	0.03	0.04	0.04	0.02	0.02	0.03	0.03	0.03	0.01	0.00	0.01	0.03	
Alter OBC % (GEN ego)	0.32	0.25	0.26	0.28	0.22	0.21	0.24	0.29	0.21	0.19	0.12	0.14	0.23	
Alter GEN % (GEN ego)	0.58	0.68	0.66	0.63	0.74	0.74	0.71	0.62	0.74	0.79	0.88	0.85	0.70	
Alter SC% (ST ego)	0.69	0.80	0.82	0.87	0.85	0.83	0.73	0.82	0.87	0.79	0.96	0.94	0.80	
Alter ST % (ST ego)	0.03	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.01	0.01	0.02	
Alter OBC % (ST ego)	0.23	0.15	0.13	0.10	0.11	0.12	0.20	0.13	0.10	0.15	0.03	0.04	0.14	
Alter GEN % (ST ego)	0.06	0.03	0.02	0.01	0.02	0.03	0.06	0.02	0.02	0.05	0.01	0.00	0.03	