Appendix
Refugee Networks, Cooperation, and Resource Access *

Daniel Masterson

November 1, 2023

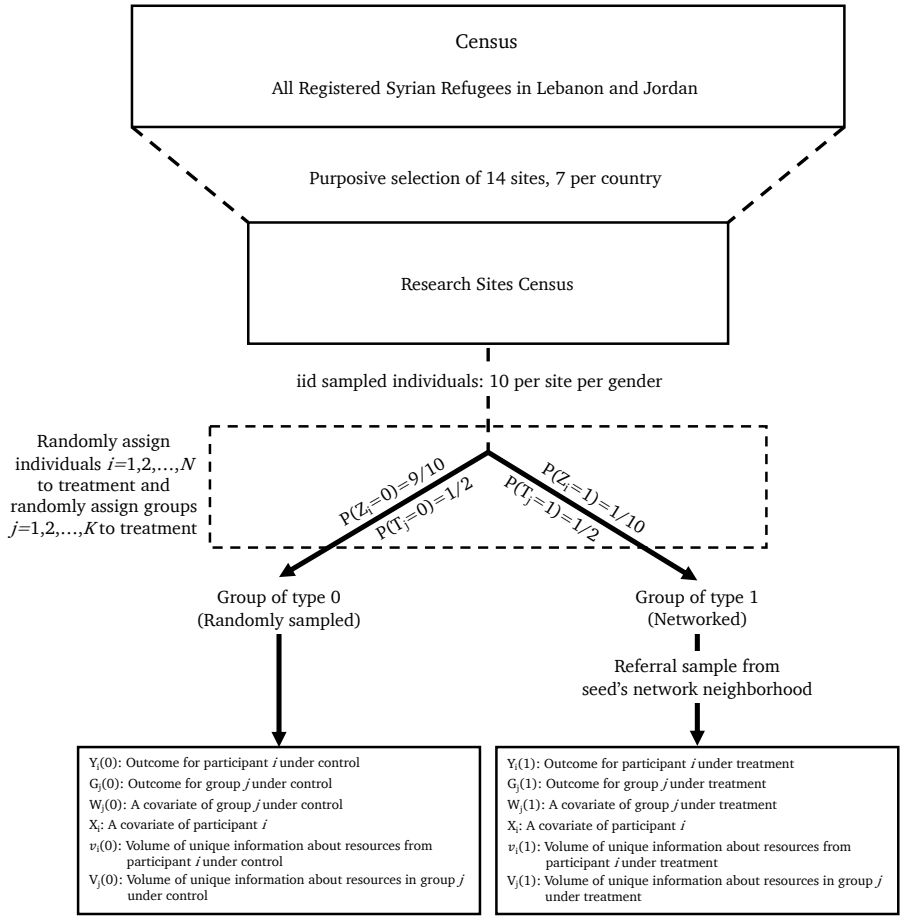
Contents

*Daniel Masterson, Assistant Professor, University of California, Santa Barbara, Department of Political Science. masterson@ucsb.edu. This research is approved by the Yale Human Subjects Committee under protocols #1603017430 and #1508016386.

1 Research Design

Figure 1A presents a schematic of the research design.

Figure 1A: Research Design



1.1 Research Project Background

The community meetings were conducted as a part of the World Bank's Syrian Refugee and Host Community research project, run with the Bank's Middle East and North Africa poverty team in 2015-16. The motivation for the broader multi-PI multi-method research project was threefold. First, it sought to leverage both qualitative and quantitative data for a comprehensive description of Syrian refugees' living conditions, with group discussion in community meetings supplementing a large-n quantitative survey. Second, it sought to identify key needs for Lebanese and Jordanian policy reform, eliciting Syrians' perspectives

on their circumstances and how they were adapting to them. Finally, it sought to support the design of better policy and interventions. Other research output from the broader project includes an internal World Bank report shared with host governments (The World Bank 2017), as well as a number of World Bank working papers (Aguilera et al. 2020; Krishnan et al. 2020*a*,*b*).

1.2 Site Selection

In Lebanon and Jordan, I selected the three governorates in each country with the largest Syrian populations according to UNHCR records. Within each of the governorates, I selected a high-population site and an average-population site. To select an average-sized site I applied both quantitate and qualitative selection criteria. There was generally one clear high-population site per governorate. To select the average-population site, 'average' is defined as the 10% trimmed mean of Syrian town-level populations within that governorate—roughly that is the mean of the middle 80% of the data. I used this statistic because, in this context, both the mean and median have their respective problems. The mean might be too high an estimate of a 'typical' town for each region due to a small number of towns with very large Syrian populations in each region. The median might be too low as an estimate of a 'typical' town in each region due to the large number of towns with a single-digit registered Syrian population size. In short, within each governorate I sampled two sites: a major Syrian population center and a typically sized Syrian community for the region.

Even after applying this quantitative criterion, multiple sites in each governorate were feasible options for the typical town. To further narrow the list of options I attempted to exclude areas with exceptional characteristics for that region (e.g., Christian town in Muslim region, rich town in a poor region). This process was based on my knowledge of the research sites and interviews with key informants who could comment on the characteristics of the potential sites.

In addition to the six sites across three governorates, in each country I included camp settings. In Jordan I included the largest official UNHCR-run camp, the Zaatari camp. In Lebanon, because rural Bekaa has many Syrian camps, in place of an average-sized town and a camp site, for the Bekaa region I included two towns with informal camps in the Bekaa: one with peri-urban, high-density camps, and a second with rural, low-density camps.

In Jordan, the selected sites were Amman and Shafa Badran in the greater Amman area, Mafraq and Irhab in the greater Mafraq greater, Irbid and Huwwarah in the greater Irbid area, and Zaatari camp. In Lebanon, the selected sites were Burj-el-Barajneh in

urban Beirut and Chhim outside Beirut; Baalbek, Khiara (rural camps site), and El-Marj (peri-urban camps site) in the Bekaa; and Tripoli (Tebbaneh) and Kouachra in the North.

1.3 Community Meeting Recruitment

Recruitment for the community meetings was based on a random sample of UNHCR registration records, drawn from all registered refugees living in the research sites between the ages of 20–50. In defining the age range I sought to achieve a balance between breadth to capture variation and narrowness to avoid deferential youth feeling inhibited from speaking around much-older participants. In addition to the age range, other inclusion criteria required that participants in both experimental arms not be members of another participant's nuclear family or household.

Community meetings were either all-male or all-female, with no mixed-gender meetings. This represents a trade-off in the research design between attrition and the realism of the community meeting setting. Some members of the Syrian population would not be willing to sit in community meetings with members of the other sex, although in most Syrian communities the sexes frequently intermix in social settings outside the family.

Recruiters were provided with four lists of names from UNHCR data for each site, one per gender per experimental group. Within each gender, selection into one group or another was randomized and the order of names in each contact list was randomized. Recruiters were instructed to contact people by phone in the order listed until recruiting a sufficient number of participants.

Recruiters read all people contacted a consent script approved by the Yale Human Subjects Committee under protocols #1603017430 and #1508016386. Another consent script was read to participants before community meetings were held. Participants were paid $20 for their participation, in consideration of the fact that the two-hour-long community meeting and travel to and from the meeting site might keep someone from a normal workday. $20 was a typical daily wage for day labor for Syrians in Lebanon and Jordan when research was conducted in 2016.

1.4 Treatment Randomization

I used block randomization at the design stage, creating treatment and controls groups that are balanced with respect to country, site, and gender, based on the expectation that these variables are likely highly predictive of outcomes. Block randomization can ensure that the treatment and control groups have equal proportions of participants in each of the notional cells in the $2 \times 2 \times 7$ research design (that is, two genders, two countries, and seven sites per country). When the blocking variables are predictive of outcomes, blocking

improves precision by preventing chance correlations between treatment assignment and baseline covariates (Miratrix, Sekhon and Yu 2013).

1.5 Experimental Recruitment Methods

In gathering referrals for the networked community meetings, people were requested to give the names of the three Syrians in their community who they had interacted with most over the past two weeks, where interaction was explained to include face-to-face communication and electronic communication via telephone, texting, WhatsApp, Facebook, etc. These three referrals needed to be between the ages of 20 and 50, live in the same town as the referrer, be the same gender as the referrer, and not be members of the referrer's nuclear family or household. Upon contacting these people, these new contacts were only included if they stated that they would be willing and able to refer three people themselves who they had interacted with frequently over the past two weeks. This inclusion criterion was applied to potential participants in the unnetworked group as well, establishing that both potential outcomes are defined for all units in the study.

The referral sampling method I designed and used is distinct from both standard respondent-driven sampling (RDS) and snowball sampling. Magnani et al. (2005) and Heckathorn (2011) describe in detail the specific meanings of snowball sampling and RDS. Snowball sampling is a nonprobability approach to sampling when the researcher does not have a list of population members (that is, a 'sampling frame'), and hence the seeds for the snowball sample are drawn from a convenience sample. In RDS, researchers can estimate selection probabilities by basic mapping of people's network (for example, asking them how many potential recruits they know). Potential recruits are commonly given a coupon by the referrer, and the referred respondent must present themselves at the study site. Researchers never need the names or contact information of potential participants. The referral sampling method in this study is not a convenience sample, as seeds were drawn at random from the UNHCR census, but unlike RDS research staff contacted referred potential participants rather than using a self-referral method.

1.6 Community Meeting and Participant Descriptives

Based on participant self-reports, the average participant in the unnetworked groups had pre-existing weak ties with 11% of the other community meeting participants, compared with 66% in the networked groups. The average participant in the unnetworked groups had pre-existing strong ties with 5% of other community meeting participants, compared with 42% in the networked groups. Weak ties were defined as people in the group who the participant knew by name before the day of the community meeting. Strong ties were

defined as people with whom the participant regularly exchanged visits.

The average community meeting had 8.7 participants and ranged in size from 6 to 10 participants. The population is UNHCR-registered Syrians refugees in Lebanon and Jordan, which includes a vast majority of Syrians living in the two countries. Although precise numbers are lacking, roughly 75% of Syrians living in Lebanon are registered, and Jordan more than 90% in Jordan are registered. UNHCR declared a (*de facto*) blanket refugee-status determination for all Syrians, rather than proceeding on a case-by-case basis as is done in many refugee registration processes, so any Syrian who seeks to register qualifies for refugee status. Based on my fieldwork, unregistered Syrians tend to be either upper- and middle-class Syrians who do not view themselves as 'refugees' in need of assistance from the UN or others who view the risks of legibility to the Jordanian and Lebanese governments as outweighing the benefits of aid provision. This latter class of unregistered often suspect, correctly in fact, that UNHCR shares its registration records with the host governments.

Attendance rates were balanced across treatment and control arms. Although attendance rates across treatment and control arms are balanced, due to an error in recruitment, unnetworked groups were one person smaller on average than networked groups. The difference in the number of people contacted was a flaw in the execution of the research design. Recruiters in Lebanon contacted only 11 people to fill 10 spots in the unnetworked groups, and recruiters in Jordan contacted 12 people to fill 10 spots in the unnetworked groups. This should have been held constant at 13 to match the number of people contacted for the networked group in each country. The attendance rates across treatment and control are balanced, but the number of participants is imbalanced due to differential recruiting targets. In both countries the networked groups were larger by about one person on average. 8.2 vs. 9.3 overall, 8.7 vs. 9.6 in Lebanon, and 7.6 vs. 8.9 in Jordan. In both countries the minimum and maximum community meeting sizes were 6 and 10, with a standard deviation of the number of participants was 1.09 in Lebanon and 1.38 in Jordan.

Existing theory suggests that the difference in group size would create a bias against both central findings. If Olson (1965) is correct that larger groups are more likely to atomize because cooperation is harder, we would be less likely to see the larger networked group engaging in dialogue with each other—the opposite of what we observe. Second, a larger group should have more information about resources for responding to public-goods problems, and we would be less likely to see the smaller unnetworked group turning to outside solutions—again, the opposite of what we observe.

1.7 Data Collection and Community Meeting Conduct

The community meetings were run in Lebanon in May and June 2016, and in Jordan in June and July 2016. Due to delays with obtaining permits for research in Zaʿatari camp, the community meetings there were run in September 2016. Pilot community meetings were conducted in Lebanon in May 2016 to improve the public-goods vignettes, discussion guide, and framing of the study. Community meetings scheduled to be conducted in the Kurdish Region of Iraq (KRI), set to begin around October 20, 2016, were cancelled because of security concerns surrounding the battle of Mosul, which began on October 16, 2016.

All aspects of the study were conducted in Levantine Arabic, including recruitment and data collection, and all documents read to or distributed to participants were in straightforward formal Arabic. I speak the dialect fluently and all recruiters and community meeting moderators were native speakers. I monitored all aspects of the study including recruitment, data collection, and community meeting transcription. Two experienced female Lebanese community meeting moderators conducted the community meetings.

During the community meetings, two recorders were used, one at each side of the group, to increase the transcribers' ability to hear all community meeting content.

1.8 Steps to Preserve Excludability

Research design and conduct were kept as similar as possible across experimental arms. Moderators were not told about the intention of the construction of networked and unnetworked groups, and were not told about the hypotheses under investigation. Although, in order to coordinate recruitment (using different strategies) and the logistics of getting the participants into the right room, moderators did know that the groups were either mostly strangers or mostly friends. Moderators were trained that they should do very little to guide discussion after the audio files were played. At most, if participants asked what they were supposed to do, the moderators were trained to say something minimal like, "What are you going to do?" or "Can you do anything in this situation?" but explain no more and never express expectations that people work together.

I manually assigned moderators to community meetings, with an aim to achieve balance between the moderators on group gender and treatment status. Random assignment of moderators to community meetings was infeasible—scheduling dozens of community meetings across fourteen sites in two countries for two moderators was already a formidable logistical effort. This included scheduling and arranging ground transport for all research participants to and from the group discussion sites, scheduling transport for moderators to sites within each country, and arranging air travel for moderators between Lebanon and Jordan.

1.9 Content of the Community Meetings and Audio Vignettes

The paper examines outcomes drawn from the community meeting discussion of four public goods vignettes, which lasted roughly 20 minutes in each community meeting. The full community meetings lasted about one hour and forty-five minutes and covered location choice, living conditions, pre-migration ties, income generation, local law and order, and intra-household issues including intimate-partner violence.

The order of the audio vignettes was randomized to eliminate order effects.

The audio vignettes were intended to study how Syrians confront community problems, and the impact of social network structure on the responses. Moderators played community meeting participants audio vignettes of two Syrian men discussing problems that Syrian communities commonly face. The vignettes discussed issues common in Syrian refugee communities including resource redistribution, public safety, access to labor and income, property rights, and the ability to run a shop. After playing each vignette the discussion was opened up for the participants. Moderators did little to shape participants' responses to the audio vignettes.

Although the community problems have possible collective solutions, the vignettes did not impose collective responses on participants. Indeed, we see variation in responses in the community meetings from disinterest to heated discussion, and from atomistic responses to communal responses.

One vignette describes a situation where an NGO delivers resources to the community without specific allocation criteria, and participants must distribute the resources. In the second vignette, a Syrian shopkeeper whose shop benefits local Syrians is being challenged and threatened by a local shopkeeper and participants are asked what they might do to respond. The third vignette presents a problem where checkpoints are preventing local Syrians from getting to work, leaving participants to discuss whether there is any response to increase access to work. In the fourth vignette, two young men get in a fight, which spills over into family conflict, and the participants are asked how they might respond. See Appendix Section 12 for the full scripts of the audio vignettes.

I developed the vignettes based on ethnographic research in Syrian communities in Lebanon for approximately one year. I developed these scenarios based on my research experience, and in conjunction with NGO colleagues, some of whom are Syrian, who working with Syrians.

1.10 Participant Protection in Humanitarian Settings

Recruiting and conducting research with participants from a vulnerable population requires great care to minimize potential for harm. To increase the anonymity of recruitment identi-

fiers I utilized the fact that a majority of Syrian adults have a nickname. During recruitment and discussion conduct the research team referred to participants by their nickname whenever possible. The widespread nickname system is based on the name of someone's eldest son or eldest daughter (if they have no son). For example, Um Ali (meaning Ali's mother) is the nickname for a woman whose eldest son is named Ali. Abu Muhammed (meaning Muhammed's father) is the nickname for a man whose eldest son is named Muhammed. Sometimes a man without children uses a similar nickname, but replaces the name of a child with the name of his father, implying that when he has a son, he will name the son after his father, although it is also widespread for young and adult men to use their father's name even if they do not plan to name their child after their father. Women who do not have children less frequently adopt such nicknames, although some adopted them during the early days of the Syrian uprising to protect their identities.

Files including recruitment information and transcripts were password-protected and encrypted. People's responses were further protected by the fact that their statements and real names never appear in the same document. In the audio files of the discussions, moderators referred to people by their nickname and the transcripts identify participants by their nickname. A post-discussion questionnaire recorded each participant's nickname. With participants' nicknames indicated in audio recordings, transcripts, and post-discussion questionnaires, I can link transcript data to questionnaire data while maintaining participant anonymity.

1.11 Coding Guide and Tagging

I tagged transcripts of group discussions according to a coding guide that I developed in partnership with three researchers who were not otherwise involved in the project. Using an iterative process, we each read a random sample of vignettes to define codes that capture salient dynamics in discussions. In the first stage, one outside researcher and I each read a random sample of transcripts, and documented the salient themes that we each found in the discussions. We met to consolidate our respective themes and collaboratively define coding rules for each. Next, a second outside researcher read a random sample of transcripts, and then read the draft of the coding guide, offering comments on existing themes and coding rules, and suggesting revisions and additions. Then a third outside researcher conducted the same procedure as the second. After this feedback, I finalized the coding guide. The researchers who developed the coding guide and I were blind to treatment status while developing the guide.

In my coding of transcripts, I randomized the order in which I read the groups and vignettes within each group. I coded transcripts according to the guide and made no

modifications to the guide after I began coding. Coding was not automated or predictive; I read and hand-coded all transcripts using the qualitative data analysis software Dedoose, which then output the results in a spreadsheet for statistical analysis. I was blind to treatment status and while coding transcripts.

1.12 Considerations for a Two-country Study

I aimed to preserve as much similarity as possible in research conduct across the two countries. Most importantly, I used the same two community meeting moderators in each country, which involved additional costs for moderator travel from Lebanon to Jordan.

Some dialectical modifications were necessary to ensure that the vignettes resonated with participants and therefore prompted substantive discussion of the public goods problems. These changes were necessary because some technical terms vary between Lebanon and Jordan, and not because Syrians in Lebanon and Jordan speak different dialects. All changes necessary were spliced into the same audio files, rather than recording the vignette from scratch. For example, any reference to a host community member in Lebanon used the word "Lebanese," while the word "Jordanian" was used in Jordan. Voice actors recorded both words and I edited audio files to produce otherwise identical audio recordings, but for those key words.

Some technical terminology also required modifications. The commonly used terminology for police checkpoints varies between the two countries. In Lebanon, police checkpoints are referred to as roadblocks/barriers (*hawajiz*) whereas in Jordan they are generally called police patrols (*dawriat shourta*). The relevant legal residency document for Syrians in Lebanon is called a residency (*iqama*) and in Jordan it is a security card (*bataqa amnia*). In Lebanon is the *iqama* is sometimes also referred to as papers (*awraq*) and in Jordan alternatively as Services card (*bataqat khadimat*). Whereas Syrians in Lebanon lack legal residency because they never obtained the document or it expired, the residency document for Syrians in Jordan does not expire, but instead it was replaced by a new document that Syrians needed to go obtain. Therefore, whereas Syrians in Lebanon say their residency is expired or documents are expired, Syrians in Jordan do not say their "security cards are *expired*" (*bataqat amnia khalaseen*). Instead a Syrian in Jordan would be likely to say "I do not have a new security card" (*ma m3na bataqat amnia jadideh*), which is the phrasing I used in Jordan in the relevant vignette.

2 Empirical Strategy

There are N randomly sampled participants, indexed $i = 1, 2, \ldots, N$. There are K community meetings, indexed $j = 1, 2, \ldots, K$. Y_i denotes a metric of randomly sampled participant i's behavior in the group discussion. G_j denotes a metric of group behavior in community meeting j. I consider a binary treatment, where $Z_i = 1$ denotes that randomly sampled participant i is assigned to treatment (placement into group with their personal network, recruited through referral). In the group-level experiment, $T_j = 1$ denotes that community meeting j is assigned to treatment (recruitment through referral). I define a covariate of individual i as X_i. I define a covariate of community meeting j as W_j and the population mean for covariates X and W as $\overline{X}$ and $\overline{W}$.

Random assignment to place at two levels: First, randomly sampled individual i is randomly assigned to treatment status. Sitting with a group of randomly sampled individuals or with a group formed through referral recruitment is denoted respectively $Z_i \in \{0, 1\}$. Randomly assigning an individual to be the first member of a group that is then recruited through referral sampling or random sampling also implicitly randomly assigns the group to its treatment status. Groups are indexed in j and random assignment to recruitment through random sampling or referral recruitment is denoted respectively $T_j \in \{0, 1\}$.[1]

I assume SUTVA at the levels of the randomly sampled individual and the community meeting, such that:

$$Y_i = Y_i(Z_i) \tag{1}$$

$$G_j = G_j(T_j) \tag{2}$$

$$W_j = W_j(T_j) \tag{3}$$

Equations 1, 2, and 3 state that the observed values of Y_i, G_j, and W_j are only a function of the treatment assigned to randomly sampled individual i and to community meeting j.

Random assignment implies the following independence relationships:

$$Y_i(z) \perp\!\!\!\perp Z_i, \text{ for } z \in \{0, 1\} \tag{4}$$

$$G_j(t) \perp\!\!\!\perp T_j, \text{ for } t \in \{0, 1\} \tag{5}$$

$$W_j(t) \perp\!\!\!\perp T_j, \text{ for } t \in \{0, 1\} \tag{6}$$

Equations 4, 5, and 6 imply Equations 7, 8, and 9, respectively. I use the plug-in principle to estimate the inferential targets, the left-hand side of Equations 7, 8, and 9, with the sample analogues of the right-hand side of the equations.

[1] To establish that both potential outcomes are defined for all units in the study, all participants in the study—both referral recruited and randomly recruited—needed to refer three close ties, although referrals were only contacted for people in the networked groups.

$$\mathrm{E}[Y_i(z)] = \mathrm{E}[Y_i|Z_i = z], \text{ for } z \in \{0,1\} \tag{7}$$
$$\mathrm{E}[G_j(t)] = \mathrm{E}[G_j|T_j = t], \text{ for } t \in \{0,1\} \tag{8}$$
$$\mathrm{E}[W_j(t)] = \mathrm{E}[W_j|T_j = t], \text{ for } t \in \{0,1\} \tag{9}$$

Although potential outcomes are independent of treatment assignment, referral recruited participants have different probabilities of recruitment than randomly sampled participants. In traditional experiments, the research design implies both Equation 9 and the observable similarity of the pre-treatment covariate W across treatment arms, that is, $\mathrm{E}[W_j|T_j = 1] = \mathrm{E}[W_j|T_j = 0]$. My research design implies Equation 9 but does not imply that the observed covariate looks similar across treatment arms. That is, in this research design $\mathrm{E}[W_j|T_j = 1]$ does not necessarily equal $\mathrm{E}[W_j|T_j = 0]$.

The research design implies that $\mathrm{E}[X_i] = \mathrm{E}[X_i|Z_i = 0] = \mathrm{E}[X_i|Z_i = 1]$. This states that the characteristics of all randomly sampled individuals are equivalent in expectation. Therefore, in the special case where W denotes the group level mean of a covariate and X denotes the individual values of the same covariate, $\mathrm{E}[W_j(0)] = \mathrm{E}[X_i|Z_i = 0] = \mathrm{E}[X_i|Z_i = 1] = \overline{X} = \overline{W}$. The equality is a result of random sampling from the population. The expectation for the covariates of randomly sampled groups and randomly sampled individuals are equal to the population mean of the covariate.

3 Randomization Check

I check randomization by testing for the observable similarity of randomly sampled participants across treatment arms, which includes all participants in control groups (randomly sampled groups) and seeds in treatment groups (referral recruited groups), and excludes treatment-group referral recruits. I use the plug-in principle to test Equation 10 with my sample.

$$\mathrm{E}[X_i|Z_i = 0] = \mathrm{E}[X_i|Z_i = 1] \tag{10}$$

As we would expect under random assignment, data from a participant questionnaire shows that measured pre-treatment covariates of randomly sampled units are balanced across the two experimental conditions. I run a test of joint balance (aka, joint orthogonality), testing the joint hypothesis: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$, by running an F-test on a linear regression of treatment assignment on measured covariates, subsetted to randomly sampled participants. The test of joint orthogonality fails to reject the null hypothesis of equality (randomization-inference p value: 0.19).

4 Manipulation Check

As a basic test of design, I demonstrate that treatment (referral recruited) groups had higher density and lower diversity.

4.1 Density Manipulation Check

First, I test whether treatment groups had higher density, defined as δ_j in Equation 11 as the share of realized ties to possible ties in community meeting j. d_j denotes the total number of realized ties between the n_j participants in community meeting j. $n_j(n_j - 1)$ denotes the total number of possible ties in the community meeting. Network relationships are directed in the sense that one person may report knowing another person even if the latter does not report knowing the former. If everyone knows everyone else in a group, the density is 1. If half of the possible relationships in a group are realized, the density of the group is 0.5.

$$\delta_j = \frac{d_j}{n_j(n_j - 1)} \tag{11}$$

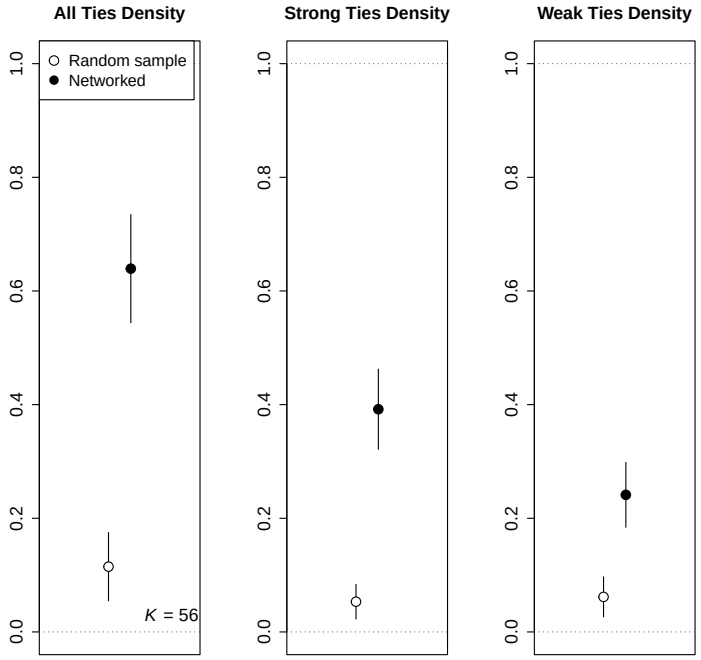
I present three metrics of participant connections. First, I measured all ties, operationalized as other people in the community meeting whom the respondent knew by name before the day of the meeting. Second, I asked respondents how many other participants they regularly exchanged visits with, a metric I refer to as strong ties. Third, I calculate a metric of weak ties by subtracting the strong-ties metric from the all-ties metric for each respondent, which captures the number of participants whom the respondent knew by name but did not regularly exchange visits with.

Figure 2A shows that the density of networked groups was much higher than that of unnetworked groups. The unnetworked groups had an average density of pre-existing ties of 11.5%, compared with 63.9% for the networked groups. Looking at panel 2, we see that the unnetworked groups had an average density of pre-existing strong ties of 5.3%, compared with 39.2% for the networked groups. Looking at panel 3, we see that the unnetworked groups had an average density of pre-existing weak ties of 6.2% , compared with 24.1% for the networked groups. The expectation that referral groups will be denser than randomly sampled groups is clearly supported by the data.

4.2 Diversity Manipulation Check

The expectation that the randomly sampled groups will be more diverse than the referral groups is based on a broad social network literature on homophily and the related empirical regularity found in many studies that people who are socially connected tend to be more

Figure 2A: Density of Within-Community-Meeting Social Ties



Note: Table version of results presented in Table 1S, Supplementary Tables (Masterson 2023).

similar than people who are not connected. There is widespread empirical and theoretical support for the idea that "birds of a feather flock together" (McPherson, Smith-Lovin and Cook 2001).

I test for a treatment effect on diversity using a standard metric of diversity, the product of the shares of units of each 'type' within each group. The measure is known as the Herfindahl index or the Herfindahl–Hirschman index (HHI) in economics and as the Simpson Diversity Index in ecology. HHI is used as a measure of market diversity or competition within industries. Simpson's Diversity Index is used to quantify the biodiversity of habitats.

When applied to firms the index measures the size of a firm in relation to its industry, as the sum of the squares of the market shares of firms within the industry. In the case of members in experimentally formed groups, I define the HHI of group j (H_j) as the sum of the squares of possible types within groups:

$$H_j = \sum_{k=1}^{\ell} s_k^2 \tag{12}$$

Across all participants, there are ℓ possible types, indexed $k = 1, 2, \ldots, \ell$. Within each group j, a share $s_j \in [0, 1]$ of the group members belong to each type k. An HHI

of 1 would correspond to a perfectly homogeneous group, and lower HHI values indicate greater diversity. In this study, type is defined with respect to a number of covariates: age, household size, arrival year in Lebanon, and ever-married status. The first three variables are continuous are dichotomized as above/below median. If two units shared all the same traits they are marked as the same type, otherwise they are a different type.

We can conduct manipulation check at the group level and at the participant level. First, when a group is recruited via networked recruitment, how much more or less diverse can we expect the group to be than if it had been recruited via random sampling? Second, if an individual is assigned to a group with their close network, how much more or less diverse can we expect the peers they interact with to be than if they had been assigned to a group of randomly sampled individuals?

The results support the expectation that randomly sampled groups will be more diverse, aligning with existing evidence and theory about homophily in social networks. Again, *increases* in the HHI correspond to a *decrease* in diversity. In both analyses there is a large decrease in diversity in the networked groups. At the group-level we see a 10.96% decrease in diversity (from 0.25 to 0.27) due to assignment to a networked group, with a randomization inference p-value of 0.3. At the individual-level we see a 10.92% decrease in diversity (from 0.24 to 0.27) due to assignment to a networked group, with a randomization inference p-value of 0.01. Although the two estimates are of similar magnitude, the uncertainty is greater at the group level, due in part to small sample size (56), compared to the individual level (227).

5 Quality Checks

I carried out thorough data quality checks. All checks suggest that the design was successfully implemented. I was in Lebanon and Jordan actively supervising research during the recruitment for and conduct of community meetings. On data-collection days a member of the research team validated essential features of the recruitment strategy and inclusion criteria before group discussion began. In calls to a random sample of participants after data collection, I asked participants about the recruitment process, how they had been contacted, whether they had been asked if they would be able to refer to three people outside their nuclear family and household to participate.

I validated transcription quality by reading a sample of English-language transcripts while listening to the corresponding Arabic-language audio recordings. I validated questionnaire data entry by reviewing samples of paper questionnaires against the dataset to ensure accuracy. Last, a number of questions were included in the participant questionnaire to test the successful randomization of experimental conditions.

6 Transcription

Community meeting transcripts were transcribed and translated by eight Lebanese transcriptionists. I conducted quality checks every week during transcription by listening to the Arabic audio files while reading recent English-language transcripts. During these quality checks I deemed one transcriptionist's work to be inadequate. That staffer was let go, replaced, and all their transcripts were redone.

I could not randomly assign transcriptionists to transcripts given the dynamic nature of which transcripts needed to be transcribed and which transcriptionists were working. Instead I gave the transcription manager three covariates—community meeting gender, treatment status, and moderator—and effectively had them use blocked non-random assignment of transcriptionists to community meetings. The transcription team make a mistake saving data about which transcriptionist handled which documents, which means I cannot test whether results are robust to controlling for transcriptionist.

7 Randomization Inference Procedures

I blocked treatment assignment on site and gender, with two community meetings in each block, and randomly assigned one community meeting to control and one to treatment in each site-gender block. Therefore, I conduct randomization inference by simulating 100,000 treatment assignment vectors under the blocking scheme. For each vector of simulated treatment labels I calculate a test statistic according to the simulated treatment assignment vector.

8 Results are Robust to Covariate Adjustment

I test for the robustness of experimental results by estimating the treatment effect with covariate adjustment, across multiple covariate sets. The results shown in Figures 3Ab and 3Aa, with point estimates and confidence intervals moving only slightly across specifications provides evidence of the robustness of the experimental results. I calculate the results in Figures 3Aa and 3Ab using OLS regression with robust standard errors, clustered at the community meeting level. I present treatment effect point estimates and 95% cluster-robust confidence intervals for all the outcomes presented in the main paper. I present the coefficients estimates from regressions including covariates including an indicator variable for moderator and an indicator variable for whether a community meeting took place in Jordan. I also present results adjusting for covariates using Lin (2013)'s covariate adjustment method of including all covariates and the interaction of treatment with the demeaned

covariates.

Figures 3Aa, 3Ab, 4Aa, and 4Ab do not show meaningful variation in point estimates or uncertainty for the paper's key results across covariate sets.

9 Tests of Alternative Explanations

9.1 Trivial Conversation

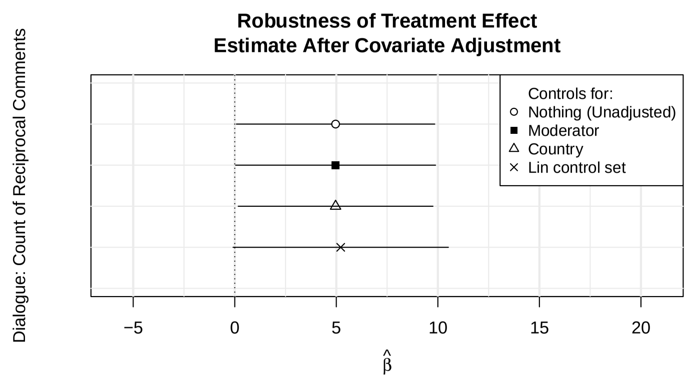
The findings on the impacts of group structure on behavior in the community meetings could be driven by trivial conversation dynamics rather than meaningful engagement with the problems. First, networked groups may simply talk more, both about trivial matters and in response to community problems. The data, however, do not support this possibility. As discussed in the results for the group-level network effect on engagement, networked groups not only exhibit a higher *number* of comments engaging with problems but also a higher *share* of comments engaging with vignettes.

Second, the resource results in the main paper in Figures 3, 4, and 5 would be trivial if randomly sampled groups discuss resources more but do not have more access to resources. In this case we would expect randomly sampled groups to make both more positive comments *and more negative comments* about resources. This would suggest that randomly sampled groups discuss resources more, and possibly are more aware of their existence, but it would undermine the conclusion that randomly sampled groups are better able to access resources. In contrast, if randomly sampled groups in fact possess more information about how to access diverse resources, I would expect them to make more positive statements about resources, as shown above, and expect them to make the same number of negative comments as networked groups (or possibly fewer negative comments).

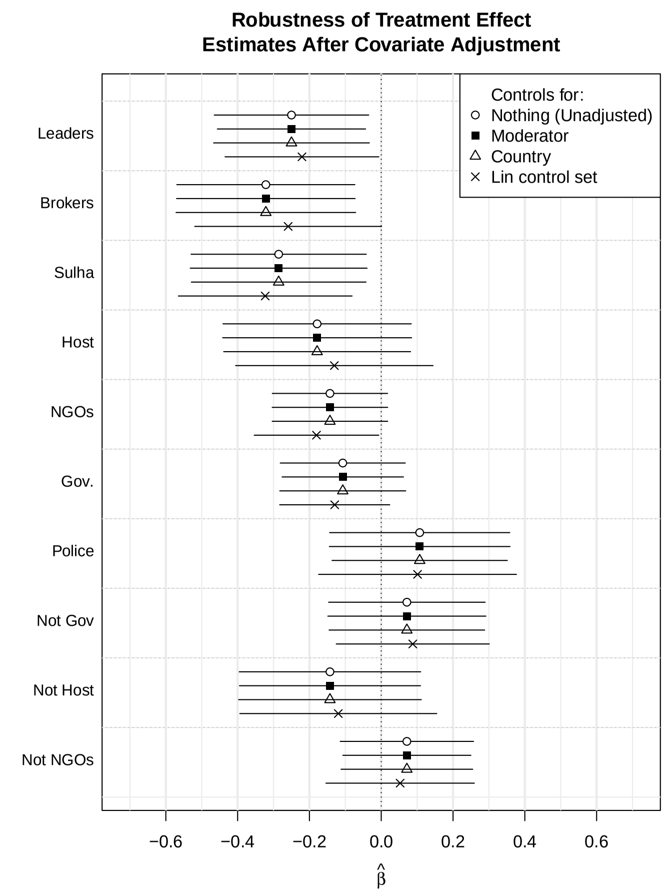
The results presented in Figures 3, 4, and 5 in the main paper are based on statements like "We can turn to brokers in response to the problem." Now I test for effects on statements about *not* being able to access resources, such as "We *cannot* turn to brokers in response to the problem." The regression results are presented in Table 1A. I do not find strong evidence of a relationship between group structure and negative statements about resources. First, columns 1-3 show that no one made statements that Syrian refugee leaders, brokers, or traditional dispute resolution would *not* be helpful. In columns 4-7, I do not find clear differences between the community meetings assigned to networked and unnetworked recruitment in terms of negative statements about resource access.

Table 2A presents results testing for an individual-level network effect on negative statements about resources. The one metric with a detectable difference (RI p-value: 0.05) is the share of seeds who discussed *not* being able to turn to the government. The lack of

Figure 3A: Robustness of Group-level Results to Covariate Adjustment, Panel (a) Dialogue, Panel (b) Resources

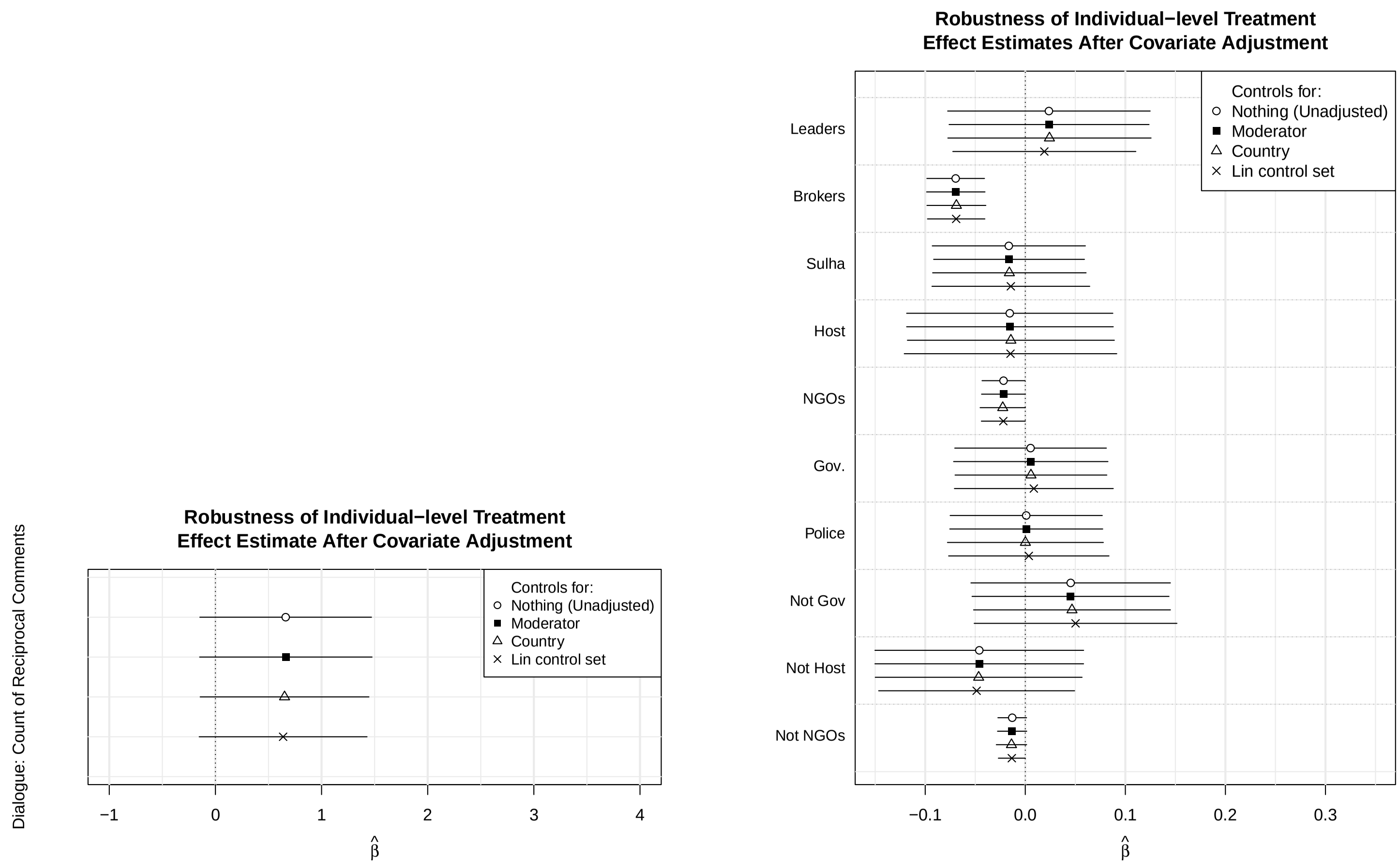


(a) Robustness of Group-level Dialogue Results Across Control Sets
Note: $K = 56$. Confidence intervals are calculated with robust standard errors. Table version of results is presented in Table 7S in the Supplementary Tables, available at Masterson (2023).



(b) Robustness of Group-level Resources Results Across Control Sets
Note: $K = 56$. Confidence intervals are calculated with robust standard errors. Table versions of results are presented in Tables 8S–17S of the Supplementary Tables, available at Masterson (2023).

Figure 4A: Robustness of Individual-level Results to Covariate Adjustment, Panel (a) Dialogue, Panel (b) Resources



(a) Robustness of Individual-level Dialogue Results Across Control Sets *Note:* $K = 258$. Confidence intervals are calculated with robust standard errors, clustered at the community meeting level. Table version of results is presented in Table 18S of the Supplementary Tables, available at Masterson (2023).

(b) Robustness of Individual-level Resources Results Across Control Sets *Note:* $K = 258$. Confidence intervals are calculated with robust standard errors, clustered at the community meeting level. Table versions of results are presented in Tables 19S–28S of the Supplementary Tables, available at Masterson (2023).

Table 1A: Group-level Network Effect: Not Able to Rely on Resources

	Not Leaders	Not Brokers	Not Sulha	Not Host	Not NGOs	Not Gov	Not Police
Control mean	0	0	0	0.71	0.11	0.18	0.61
$\hat{\beta}$	0	0	0	-0.14	0.07	0.07	-0.07
	(0)	(0)	(0)	(0.13)	(0.09)	(0.11)	(0.13)
RI p-value	–	–	–	0.29	0.69	0.72	0.43

Note: $K = 56$. $\hat{\beta}$ denotes difference-in-means estimate. Robust standard errors are reported in parentheses. Randomization inference performed with 100,000 simulated randomized treatment assignment vectors, blocked by country, site, and gender following the same structure used for actual randomization.

a relationship between group structure and negative statements obtains across the other resources. Overall, I do not find strong evidence for the possibility that the paper's results are driven by trivial discussion.

Table 2A: Individual-level Network Effect: Not Able to Rely on Resources

	Not Leaders	Not Brokers	Not Sulha	Not Host	Not NGOs	Not Gov	Not Police
Control mean	0	0	0	0.12	0.01	0.03	0.1
$\hat{\beta}$	0	0	0	-0.05	-0.01	0.05	0.04
	(0)	(0)	(0)	(0.05)	(0.01)	(0.05)	(0.07)
RI p-value	–	–	–	0.24	0.25	0.05	0.31

Note: $N = 258$. $\hat{\beta}$ denotes difference-in-means estimate. Cluster robust standard errors, clustered at the group level, are reported in parentheses. Randomization inference performed with 100,000 simulated randomized treatment assignment vectors, clustered at the group level and blocked by country, site, and gender following the same structure used for actual randomization.

9.2 Normative Obligations and Social Preferences

People in networked groups may cooperate more because they care about about each other's welfare or share a sense of normative obligation, even in the absence of any network effect on information flow. Putnam (2000) emphasizes that frequent interaction tends to produce norms of generalized reciprocity. Alesina and Ferrara (2005) propose that people may be more altruistic towards in-group members because they internalize the benefits to these people more than benefits to people outside their group.

If people have altruistic or sociotropic preferences for cooperation with their network

neighborhood, we might expect participants in networked groups to view problems as affecting the well-being of other people in the community, not just themselves and their families. To assess whether this alternative mechanism is at play, I test for treatment effects on two proxies for sociotropic concerns. First, I test whether networked groups are more likely to discuss problems as affecting the well-being of the community. Second, I test whether networked groups were more likely to discuss problems as affecting the well-being of people other than themself and their family.

Table 3A: Normative Obligations and Social Preferences

	Affects Community	Affects Other People
Control mean	0.79	0.5
$\hat{\beta}$	-0.11	0.07
	(0.12)	(0.14)
RI p-value	0.51	0.43

Note: $K = 56$. $\hat{\beta}$ denotes difference-in-means estimate. Robust standard errors are reported in parentheses. Randomization inference performed with 100,000 simulated randomized treatment assignment vectors, blocked by country, site, and gender.

Table 3A shows no detectable treatment effect on whether groups viewed the problems from a more collective perspective (left column) or through a more sociotropic lens (right column). The lack of evidence of these mechanisms aligns with existing studies (e.g., Habyarimana et al. 2009) finding that people do not exhibit greater concern for their in-group peers' welfare or prefer working with in-group members. The results are also consistent in individual-level experimental tests. Examining the same two outcomes in the individual-level experiment, Table 4A also shows no detectable treatment effects.

Table 4A: Individual-level Network Effect: Normative Obligations and Social Preferences

	Affects Community	Affects Other People
Control mean	0.08	0.17
$\hat{\beta}$	-0.04	0.12
	(0.04)	(0.13)
RI p-value	0.5	0.19

Note: $N = 258$. $\hat{\beta}$ denotes difference-in-means estimate. Cluster robust standard errors, clustered at the group level, are reported in parentheses. Randomization inference performed with 100,000 simulated randomized treatment assignment vectors, clustered at the group level and blocked by country, site, and gender following the same structure used for actual randomization.

9.3 Network Location

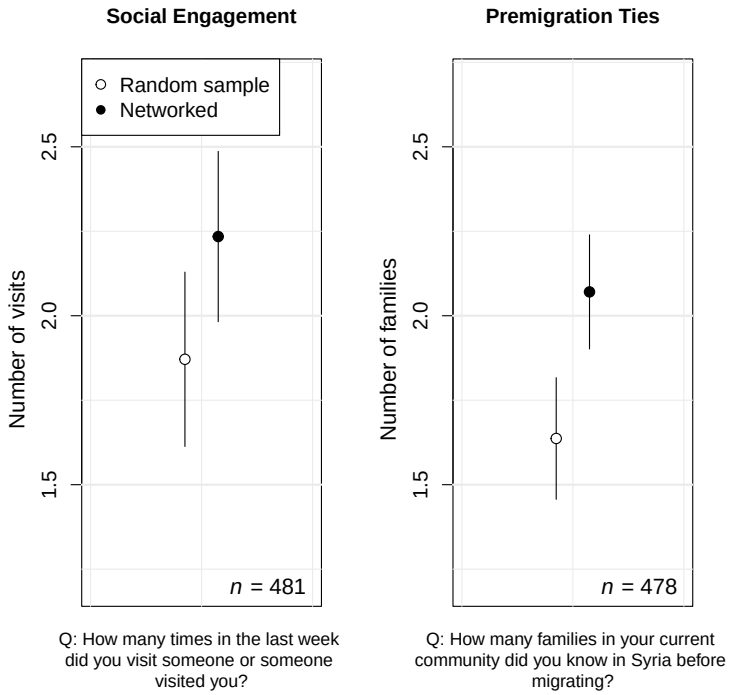
This paper focuses theoretically on the effects of group density and diversity on information flow in driving community problem solving. Although manipulation checks show clear evidence that the intervention impacted group density and diversity as intended, random variation in group recruitment creates groups that differ in ways beyond just those two network features. One relevant consideration for understanding cooperation is whether the referral-recruited participants in networked groups exhibit different network locations in their community than participants in randomly sampled groups. Recent work in political science argues that network location is an important driver of cooperation. For instance, people in peripheral network locations may face incentives to be less cooperative than more central individuals (Larson 2017).

I explore whether individuals in networked groups are more central in their communities than people in unnetworked groups. Beyond the clear evidence from the manipulation checks that networked groups exhibit greater *within-group* density, I now examine participants' relationships with people *outside* the meeting. I cannot measure network centrality directly, as mapping complete social networks was beyond the scope of this study. Instead, I test for treatment effects on two self-reported proxies for network centrality. First, I study the number of times in the last week that respondents visited someone or had someone visit them. Second, I measure respondents' long-term social relationships in their current community, defined as people in a respondent's town who they knew in Syria before migrating.

Beyond the fact that people in networked groups are more connected to each other, Figure 5A shows that they are also more connected to their community. This aligns with an intuition from Feld (1991) that sampling along edges (referral sampling) brings in more connected individuals than recruiting nodes (random sampling). Because networked groups are more connected with their community, we might expect them to have access to a wider range of information about resources. Yet the paper's results on resources show that even though the networked groups know *more* people in the community, they nonetheless draw on external resources *less* in the community meetings. This further supports the key finding about resource diversity from the parallel group-level and individual-level experiments: the resource diversity disadvantage of networked groups is driven by structural changes in the groups and not individual-level effects.

The role of network position and the network effects studied in this paper are not mutually exclusive explanations for cooperative dynamics. The goal of this paper is not to rule out network location as important, but rather to explore the importance of group density and diversity for group- and individual-level cooperative dynamics in a refugee setting.

Figure 5A: Community-Network Centrality



Note: Units are individual participants who answered each question on post-meeting questionnaire. Tabular results presented in Table 2S, Supplementary Tables (Masterson 2023).

10 Heterogeneous Treatment Effects by Gender and Country

In Supplementary Tables 29S–36S (Masterson 2023), we see that results are generally robust to analyzing treatment effects by gender and country, both at the group level and the individual level. Overall, analysis of treatment effects on women and men, and in Jordan and in Lebanon, leads to the same substantive conclusions as in the main analysis. These results, however, should be interpreted with caution. Given the sample size of 56 in the group-level experiment and 258 in the individual-level experiment, the study may not be well powered to detect heterogeneous treatment effects. The study was not designed to study sub-group heterogeneity and therefore is not well equipped to detect such differences. A study seeking to test whether the impacts of network structure are different across genders or across country contexts should select a design to detect such heterogeneous treatment effects.

11 Complete Coding Guide

Coding guide available at: `https://bit.ly/3wlhPOa`

12 Full Text of Vignettes

Below are English translations of the vignette scripts as they were played in Lebanon. The changes made for play in Jordan are discussed in Section 1.12. In each audio vignette there were two Syrian male voice actors, who I simply label 1 and 2 below. An anonymous audio recording of the vignettes with English translation in subtitles is available at this link: `http://bit.ly/vignettesaudio`.

Checkpoints	**Fight, Law and Order**
1 – Hello Abu Mustafa, did you hear that the security forces are set up a checkpoint at the entrance to the area/neighborhood? 2 – It's big problem. Most of the Syrians here don't have up-to-date residency. And we can't pass out of the area, not to work, or for any other reason. I don't have my papers, so I cannot go to work today. 1 – Yeah, me too. There are many people who won't be able to go to work today, since they cannot go out because of the checkpoint. 2 – Did you hear that our neighbor Mohammad was detained at a checkpoint a few days ago? 1 – Yeah, I know. It is not a good situation, they arrest us at the checkpoints. We cannot move. It's not a life. 2 – I know we cannot solve this problem, but isn't there some way to reduce the pressure for all these people? 1 – Yeah, how can we get to work today?	1 – Did you hear about the trouble that happened yesterday? Two young guys got into a fight. Two Syrians, one from our neighborhood and the other from the neighboring neighborhood. It's a serious problem. I'm afraid it'll get worse if we don't do anything. It's not just a problem for the young guys anymore, it's become a family problem. 2 – I don't know what we can do. There have been a lot of problems like this lately. The last fight was really serious. One of them was injured, hit by a rock, and the two were really hitting each other. The families were seeing red. A young guy from our neighborhood had to go to the hospital and needed to pay a lot of money. He might have broken a bone. 1 – I don't know who we can go to about this problem. 2 – Yeah, I know we cannot fix it, but I think we can do something to reduce the tension.

Aid Distribution	**Local Syrian-run Shop**
1 – How are you, Mohammad? Did you hear about the new NGO working in the area? They want the local residents to select the neediest families in the area to get some help 2 – Who do you mean, they want us to select? 1 – Us, the Syrians living here, they want us to choose which families will get aid and which will not. 2 – How are we supposed to do that? Everyone that you will ask will say that they need help more than the others. 1 – Yes, I know that we all need help, but we must know who really needs more help right now. We must find a way to deal with people who won't receive aid. How are going to work with this situation?	1 – Hello, Abu Ahmed, did you congratulate Moussa on his new mini-market? Good luck to him [may God help him]. His prices are good, his products are good, and he brings his bread to middle of the neighborhood so we don't have to go out when we don't have legal papers. 2 – Yes, I congratulated him but he didn't look happy. 1 – Why? 2 – Because Mohammad, the Lebanese owner of the super market, is not happy. He threatened Mousa, and said that he would force him to close his mini-market. 1 – And what he can do to Mousa? 2 – Mohammad threatened Mousa about his residency, since Mohammad has connections that could really give Mousa trouble. 1 – God protect him. What's some way to resolve this situation? 2 – We ought to do something together to mitigate the problem.

References

Aguilera, Ana, Nandini Krishnan, Juan Muñoz, Flavio Russo Riva, Dhiraj Sharma and Tara Vishwanath. 2020. "Sampling for representative surveys of displaced populations." *Data collection in fragile states: Innovations from Africa and beyond* pp. 129–151.

Alesina, Alberto and Eliana La Ferrara. 2005. "Ethnic Diversity and Economic Performance." *Journal of Economic Literature* 43(3):762– 800.

Feld, Scott L. 1991. "Why Your Friends Have More Friends Than You Do." *American Journal of Sociology* 96(6):1464.

Habyarimana, James, Macartan Humphreys, Daniel N Posner and Jeremy M Weinstein.

2009. *Coethnicity: Diversity and the Dilemmas of Collective Action*. New York: Russell Sage Foundation.

Heckathorn, Douglas D. 2011. "Comment: Snowball versus respondent-driven sampling." *Sociological methodology* 41(1):355–366.

Krishnan, Nandini, Flavio Russo Riva, Dhiraj Sharma and Tara Vishwanath. 2020*a*. "Coping with the Influx: Service Delivery to Syrian Refugees and Hosts in Jordan, Lebanon and Kurdistan, Iraq." *World Bank Policy Research Working Paper Series, No. 9326* .

Krishnan, Nandini, Flavio Russo Riva, Dhiraj Sharma and Tara Vishwanath. 2020*b*. "The Lives and Livelihoods of Syrian Refugees in the Middle East." *World Bank Policy Research Working Paper Series, No. 9327* .

Larson, Jennifer M. 2017. "How the West Became Wild: Informal Governance with Incomplete Networks." *World Politics* .

Lin, Winston. 2013. "Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique." *Annals of Applied Statistics* 7(1):295–318.

Magnani, Robert, Keith Sabin, Tobi Saidel and Douglas Heckathorn. 2005. "Review of Sampling Hard-to-Reach and Hidden Populations for HIV Surveillance." *Aids* 19(Suppl 2):S67–S72.

Masterson, Daniel. 2023. "Replication Data for: Refugee Networks, Resource Access, and Cooperation." *Harvard Dataverse* .
URL: *https://doi.org/10.7910/DVN/IZWS10*

McPherson, Miller, Lynn Smith-Lovin and James M Cook. 2001. "Birds of a feather: Homophily in social networks." *Annual review of sociology* 27(1):415–444.

Miratrix, Luke W., Jasjeet S. Sekhon and Bin Yu. 2013. "Adjusting treatment effect estimates by post-stratification in randomized experiments." *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 75(2):369–396.

Olson, Mancur. 1965. *The logic of collective action*. Harvard University Press.

Putnam, Robert D. 2000. *Bowling alone: The collapse and revival of American community*. Simon and Schuster.

The World Bank. 2017. "Syrian Refugees and Their Hosts in Jordan, Lebanon, and the Kurdistan Region of Iraq: Lives, Livelihoods, and Local Impacts." *Internal World Bank Report* .