

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Social Information and Political Action in Honduras and Ghana

Permalink

<https://escholarship.org/uc/item/2vj5b4zg>

Author

Hughes, Douglas Alexander

Publication Date

2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Social Information and Political Action in Honduras and Ghana

A Dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Political Science

by

Douglas Alexander Hughes

Committee in charge:

Professor James H. Fowler, Chair
Professor James Andreoni
Professor Scott Desposato
Professor Seth Hill
Professor Thad Kousser

2016

Copyright
Douglas Alexander Hughes, 2016
All rights reserved.

The Dissertation of Douglas Alexander Hughes is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2016

TABLE OF CONTENTS

Signature Page	iii
Table of Contents	iv
List of Figures	viii
List of Tables	ix
Acknowledgements	x
Vita	xi
Abstract of the Dissertation	xii
Chapter 1 Summary of Dissertation	1
Chapter 2 Introduction	4
2.1 Social Information	4
2.2 Plan of the Dissertation	5
Chapter 3 A Theory of Social Information	8
3.1 Norton Shores	8
3.2 Heuristic Information	11
3.3 Scope of the Problem	15
3.4 Social Heuristics	17
3.5 Example: Social Information Predicting Behavior	20
3.6 Social Network Knowledge Shapes Messages	22
3.7 Past Social Network Approaches	25
3.7.1 Political Mobilization	26
3.7.2 Political Identification and Choice	27
3.8 Hypotheses	28
Chapter 4 Social Network Data Collection in Rural Areas	31
4.1 Local Settings	31
4.1.1 Honduras	33
4.1.2 Ghana	36
4.2 Social Network Measurement Considerations	37
4.2.1 Failure to Identify Population	39
4.2.2 Forgetting Social Ties	40
4.2.3 Non-uniquely Identifying Alters	41
4.2.4 Data Entry Errors	43
4.3 Previous Solutions	45

	4.4	Consequences for the Social Network	46
	4.4.1	False Negative Edges	46
	4.4.2	False Disaggregation of Nodes	47
	4.5	Measurement in this Work	47
	4.5.1	Serving Internet in Very Rural Locations	48
	4.5.2	Dynamic Lists and Photographs	49
	4.6	Conclusion	49
Chapter 5		Political Mobilization Through Social Paths	51
	5.1	Introduction	51
	5.2	Existing Explanations	52
	5.2.1	Individual Characteristics	53
	5.2.2	Society Wide Characteristics	53
	5.2.3	Previous Explanations Identify Social Groups	53
	5.2.4	Explanations for Participation	54
	5.2.5	Network Explanations	55
	5.3	Social Information	56
	5.3.1	Information	57
	5.3.2	Influence	57
	5.4	Methods	58
	5.4.1	Statement of Causal Model	59
	5.4.2	Estimation of Causal Effects	61
	5.4.3	Assumptions about the form of interference	62
	5.5	Data	62
	5.5.1	Assignment	62
	5.5.2	Treatment Package	65
	5.6	Results	66
	5.6.1	Estimating Within Cluster Controls	66
	5.6.2	Characterization of Mobilizers in Towns	67
	5.6.3	Town Level	70
	5.6.4	Individual Level	72
	5.7	Limitations	76
	5.8	Conclusion	79
Chapter 6		Coordination on Candidates Using Social Information	88
	6.1	Overall Theory in this Chapter	89
	6.2	Examples in Social Voting in Formal US Politics	89
	6.3	Hypotheses	90
	6.3.1	Capability Hypothesis	91
	6.3.2	Candidate Connectedness Hypothesis	91
	6.3.3	Social Proximity Mechanism	92
	6.3.4	Connectedness Mechanism	92
	6.3.5	No Social Information Comparison	93
	6.4	Research Design	93

	6.4.1	Reprise: Key Network Concepts	93
6.5		Experiment Design	96
	6.5.1	Selecting Candidates	96
	6.5.2	Candidate Attributes	97
	6.5.3	Election Text	98
	6.5.4	Voting Measurement	100
6.6		Candidate Results	100
	6.6.1	Unit of Analysis: Candidate	100
	6.6.2	Capability Hypothesis	101
	6.6.3	Connectedness Hypothesis	102
	6.6.4	Vote Share and Degree	102
	6.6.5	Vote Share and Betweenness	102
	6.6.6	Vote Share Complete Model	105
6.7		Candidate-Elector Dyad Results	106
	6.7.1	Social Proximity	106
	6.7.2	Social Information	107
6.8		Robustness	108
	6.8.1	Connectedness is Not Only a Low-Quality Candi- date Screen	108
	6.8.2	Social Distance Does Not Confound Connectedness Results	109
6.9		Discussion	127
	6.9.1	Limitations	127
Chapter 7		Knowledge and Use of Social Information	129
	7.1	Problem and Proposed Solutions	129
		7.1.1 Existing Research	130
		7.1.2 Unique Contribution	134
		7.1.3 Friendship Paradox	135
	7.2	Proposed Modification	136
		7.2.1 Requirement for Proper Functioning	137
		7.2.2 Evidence Suggests Proper Functioning	139
	7.3	Criteria for Success and Evaluation	140
		7.3.1 Node Centrality	141
		7.3.2 Network Coverage	143
		7.3.3 Overlap and Non-Overlap	145
	7.4	Datasets	147
	7.5	Simulation Results	148
	7.6	Targeting Algorithm Results	149
		7.6.1 Node Centrality	152
		7.6.2 Network Coverage	153
		7.6.3 Variance Reduction	156
	7.7	Recommendations and Conclusions	158

Chapter 8	Conclusion	161
	8.1 Overview of Findings	161
	8.2 Limitations	162
	8.3 Are People Social? Are they Political?	162
Chapter 9	Resources Utilized in this Work	164
	9.1 Mobilization Script in English	164
	9.2 Mobilization Script in Spanish	165
	9.3 Election Script in English	165
	9.4 Election Script in Spanish	166
	Bibliography	167

LIST OF FIGURES

Figure 3.1: Candidate Nominations	11
Figure 5.1: Distribution of Indegree	64
Figure 5.2: Observed vs. Simulated Distribution of Mobilizer Connectedness	68
Figure 5.3: Mobilizer Connectedness and Town Population	69
Figure 5.4: Meeting Attendance By Mobilizer Connectedness	73
Figure 5.5: Heatmap of Attendance at Political Activity	77
Figure 5.6: Mobilization in Town #9	78
Figure 6.1: Social Distance Hypothesis	111
Figure 6.2: Connectedness Hypothesis	112
Figure 6.3: Town Turnout	113
Figure 6.4: Indegree Distribution	114
Figure 6.5: Betweenness Distribution	114
Figure 6.6: Vote Share Distribution, Town Population	115
Figure 6.7: Vote Share Distribution, Turnout	115
Figure 6.8: Vote Share Degree Correlation	116
Figure 6.9: Vote Share Degree Correlation	117
Figure 6.10: First Differences, Poisson	118
Figure 6.11: First Differences, OLS	119
Figure 6.12: First Difference, 2σ Indegree	120
Figure 6.13: First Difference, 2σ Betweenness	121
Figure 6.14: Marginal Effects on Vote Share	122
Figure 6.15: Social Distance and Voting for Candidate	123
Figure 6.16: Social Distance to Candidates	124
Figure 6.17: Matched Candidates	126
Figure 7.1: Best Connected Friend Performance	140
Figure 7.2: Plot of Star Graph	147
Figure 7.3: Targeting Simulations, Degree	150
Figure 7.4: Targeting Simulations, Eigenvector Centrality	151
Figure 7.5: Best Friend Targeting Against Random Targeting	154
Figure 7.6: Random, Friend, and Best Connected Targeting	155
Figure 7.7: Social Distance Histogram	156
Figure 7.8: Alters Reached By Random Alters	157
Figure 7.9: Alters Reached By Targeted Alters	159
Figure 7.10: Reduction in Variance Due to Best Connected Targeting	160

LIST OF TABLES

Table 3.1:	Social Information Hypotheses	30
Table 4.1:	Measurement Errors	39
Table 5.1:	Covariate balance check. Logistic regression predicting assignment as mobilizer.	81
Table 5.2:	Social Balance Table. Comparing covariate values of those assigned to have a social alter at distance $dist=k$ assigned as a mobilizer.	82
Table 5.3:	Ordinary Least Squares Model Predicting Mobilizer Turnout at Political Meeting	83
Table 5.4:	Linear Probability Model. Total town political activity (meeting attendance) regressed on mobilizers connections and town size. .	84
Table 5.5:	Poisson regression. Total town political activity (meeting attendance) regressed on mobilizers connections and town size.	85
Table 5.6:	Poisson Regression. Mobilizer effectiveness as a function of social connections.	86
Table 5.7:	Core causal model. Logistic regression of turnout on experimental and non-experimental factors	87
Table 6.1:	Candidate Characteristics	99
Table 6.2:	Non Candidate Characteristics	99
Table 6.3:	Votes and vote share won by candidates as a function of the candidates' extitdegree calculated two ways: among all social contacts, and among only social contacts who are friends.	103
Table 6.4:	Poisson model estimating factors associated with candidates winning more votes in each election.	104
Table 6.5:	Voting for a Candidate as a Function of Social Characteristics . . .	125
Table 7.1:	Implementation of centrality scores	143
Table 7.2:	Means and T-test of Centrality Scores. The first two rows report mean values, and the second two rows report t-test associated scores.	153

ACKNOWLEDGEMENTS

Thank you do my dissertation committee. James, Thad, Scott, Seth, and Jim, this project would not have ever come together in any form if not for your guidance, help, and pressure. My parents and grandparents have supported me through every level of education, and my brothers have been a constant reminder that if I were to slow down, others would catch up. My friends and colleagues at UCSD have been incredibly helpful in my development as a scholar and existence as a human. Thanks Mike, Micah, Devesh, Chris, Rob, Jaime, Jason, Kyle, Aaron, Scott, Zach and Nick, Adam and Drew.

Especially, thank you Laura. Our time in North County together has been the happiest time in my life. My life's most memorable experience is surfing the moon with you in Cardiff.

OGS requires a paragraph acknowledging support and coauthors of work appearing in this dissertation. While all words and analysis in this dissertation are solely my work, this project could not have been possible without the collaboration, friendship, and support of Derek Stafford. Some of the most fun I've ever had was collecting this data.

An early version of one chapter (concerning political mobilization) was presented as joint work at Midwest Political Science.

VITA

2008	B.A. Political Science; B.A. Economics University of Michigan. Ann Arbor, MI
2011	M.A. Political Science, University of California, San Diego. San Diego, CA.
2016	Doctor of Philosophy. Political Science, University of California, San Diego. San Diego, CA.

PUBLICATIONS

Hafner-Burton, Emilie M; Hughes, D Alex; Victor, David G. "The cognitive revolution and the political psychology of elite decision making." *Perspectives on Politics* (11):2, 368-386, 2013.

LeVeck, Brad L; Hughes, D Alex; Fowler, James H; Hafner-Burton, Emilie; Victor, David G. "The role of self-interest in elite bargaining." *Proceedings of the National Academy of Sciences*(111):52, 18536-18541, 2014.

Kim, David A; Hwong, Alison R; Stafford, Derek; Hughes, D Alex; O'Malley, A James; Fowler, James H; Christakis, Nicholas A. "Social network targeting to maximise population behaviour change: a cluster randomised controlled trial." *The Lancet*(386):9989, 145-153, 2015.

Kosnac, Hillary S; Cornelius, Wayne A; Wong, Tom K; Gell-Redman, Micah; Hughes, D Alex. *One Step In and One Step Out..* Center for Comparative Immigration Studies. 2015.

Shakya, Holly B; Hughes, D Alex; Stafford, Derek; Christakis, Nicholas A; Fowler, James H; Silverman, Jay G. "Intimate partner violence norms cluster within households: an observational social network study in rural Honduras." *BMC public health*(16):1, 2016.

ABSTRACT OF THE DISSERTATION

Social Information and Political Action in Honduras and Ghana

by

Douglas Alexander Hughes

Doctor of Philosophy in Political Science

University of California, San Diego, 2016

Professor James H. Fowler, Chair

Existing research argues for one of two sides of a dichotomy. Either, individuals' social connections shape behavior because information flows along social connections; or, social connections shape behavior because of influence between social actors. A theory of social information reconciles these two positions, arguing instead that social information is context-dependent. When the task social information is oriented toward is relatively low cost, then social information's primary role is to spread information; however, when the task-orientation requires relatively higher cost, then social information shifts into an influential role.

To test these propositions, I employ two field experiments where tasks require distinct amounts of effort on the part of the political actors. In one, actors need forfeit several hours' time to take political action on a weekend. In another, actors need pay no marginal cost, but instead only need coordinate their efforts

in a common direction. A theory of social information predicts that in both tasks, individuals with a large number of social connections should be more effective than individuals with a small number of social connections; but for different reasons. In the high cost case, well-connected individuals can influence their relatively proximate social alters; in the low cost case, well-connected individuals are able to coordinate the actions of others by sending a commonly observed signal.

Chapter 1

Summary of Dissertation

What leads people to take political action? What leads them to take certain types of action? The modern study of political science holds these questions at the disciplines' core. Scholars in the 1940s and 1950s sought to answer these questions from the stand point of voters' *milieu*, but for at least the past fifty years, the discipline has developed a series of answers that place the focus squarely on features of the individual: her education, political capital, past voting history, the competitiveness of her district, how her elected official reflects her policy preferences or descriptive characteristics, how successfully a campaign mobilizes her.

In the past decade, as a consequence of new developments in measurement and computing, scholars are again developing theory and examining explanations for political action that extend beyond the individual. The earliest studies in this new wave examined groups of actors as a whole, seeking to understand how groups' social capital might shape political engagement. More recent studies delve inside these groups to examine how sets of relationships between political agents shape political engagement. Both these approaches have done well to compliment the discipline's understanding of individuals' political engagement.

Yet, current theory is largely agnostic about how to combine these two sets of understanding. On the one hand, group level information faces difficulty when making individual predictions. On the other hand, individual level information does not utilize information at any level higher than an individual. In this dissertation I argue for a theory of political action that explains behavior at this meso level. Individual political actors know not only who are their own friends, but they know

their friends' friends; individual political actors know not only whether they are a group leader, but also whom they might lead; know not only whom they respect, but also whom is respected by others. In short, individual political actors possess a great deal of *social information*, and they use this social information in predictable ways to make good political decisions.

To develop this argument, I utilize political science theories about individual rationality, cognitive science theories about information processing, and sociology theories about social structure and its role in individual decision making. Specifically, I argue when people act out their everyday lives they glean information about people with whom they interact. They learn about who is friends with whom; who tends to lead groups; whom others respect. This set of information, gathered through repeated interactions with friends, coworkers and family members provide context across a range of daily tasks. In most cases this social information is not gathered for explicitly political purposes; however, when it comes time to make political decisions, individuals find this information useful to motivate others and coordinate actions.

Existing research argues for one of two sides of a dichotomy. Either, individuals' social connections shape behavior because information flows along social connections; or, social connections shape behavior because of influence between social actors. A theory of social information reconciles these two positions, arguing instead that social information is context-dependent. When the task social information is oriented toward is relatively low cost, then social information's primary role is to spread information; however, when the task-orientation requires relatively higher cost, then social information shifts into an influential role.

To test these propositions, I employ two field experiments with tasks which require distinct amounts of effort on the part of the political actors. In one field experiment examining political mobilization, actors forfeit several hours to take political action on a weekend. In the second field experiment, actors pay no marginal cost for participation, but instead only need coordinate their efforts in a common direction. A theory of social information predicts that in both tasks, individuals with a large number of social connections should be more effective than individuals with a small number of social connections; but for different reasons. In the high cost case, well-connected individuals can influence their relatively proximate social alters; in the low cost case, well-connected individuals are able to coordinate the actors of others by sending a commonly observed signal. None of this theory of social

information is possible if people cannot reliably reason about their social networks. In the final empirical chapter perform two tests. First, I test that individuals are able to reason in the terms of social information; and, second, I demonstrate how political activity might use this information to enhance the effectiveness of mobilization and I test that people are able to reason in terms of their social networks.

Chapter 2

Introduction

2.1 Social Information

My theory of how people take political action draws on the theories of low-information rationality (Popkin, 1994), sociology and the study of peoples' social networks (Berelson, Lazarsfeld and McPhee, 1954; Christakis and Fowler, 2013; Nowak and Sigmund, 2005), and I test the predictions of this theory using the tools of field experimentation (Gerber and Green, 2000; Lucking-Reiley, 1999; List, 2011).¹

I argue for the importance of studying group-based decision making. Even in electoral systems that are designed to formalize the creation of coalitions through certain vote counting rules (Cox, 1997; Lijphart, 2012), both the elites and the masses must coordinate their internal actions in order to organize their votes and win (Cox, 1997; LeVeck, 2013). This acknowledgement does not diminish the importance of the formal political system (e.g McCubbins and Schwartz, 1984; Kousser, 2005), political campaigns (Popkin, 1994; Bartels, 2006; Gerber et al., 2013), or the role of the media (Bartels, 1993), though it does re-open the process of individual and small-group decision making as a viable line of inquiry (Edgar et al., 2012; Skyrms, 2009; Lodge and Taber, 2005; Cardenas, 2000; Truman, N.d.; Pruitt, 1971; Olson, 1965).

In this dissertation, I argue that features of political actors' social lives are important for understanding how political events proceed. Specifically, I argue that

¹In other work associated with this project, but not a part of this dissertation, I use the tools of laboratory experiments (Andreoni and Petrie, 2004) to test the provision of public goods by members of groups. In particular, in this section of the examination I leverage sampling variation in group composition in a public goods game to examine how social-network dynamics shape goods provision.

through repeated interactions with other people in their daily lives, actors develop an understanding not only of who the other actors in the system are, but also the sets of relationships that exist between each person in the system.

One way to understand this argument is in comparison with long-standing theory of opinion leadership (e.g. Lazarsfeld, Berelson and Gaudet, 1944; Berelson, Lazarsfeld and McPhee, 1954; Katz and Lazarsfeld, 1955; Katz, 1957; Robinson, 1976; Zaller and Feldman, 1992; Zaller, 1990). The opinion leadership theory posits that some individuals are primarily responsible for the transmission of information to others. One way of understanding this theory of social information might be that social information is an observable, predictive feature of someone being an opinion leader. Indeed, it might be that opinion leaders and individuals who hold a large set of influential connections are an overlapping set. However, the implications of this theory of social information are further reaching than understanding the flow of information from events through opinion leaders and to the masses. This theory helps to explain include not only opinion leadership, but also coordination around candidates, decision about turning out to vote, and choices about providing public goods.

2.2 Plan of the Dissertation

My argument for a theory of social information used by political actors is in line with the renewed interest in answering political science questions from the lens of political sociology. The renewed examination, after nearly forty years of dormancy, has been driven by developments to theory and translation of theory from other disciplines. However, much of the renewed interest can also be attributed to the development computational and statistical developments that allow political science researchers to examine questions with relationships between the data observations. This dissertation makes contributions on both fronts. The rest of the chapters in this work proceed as follows.

In chapter two I provide an in depth presentation of my theory of social information. Here I draw on the political sociological literature to lay out the specifics of how actors gather social information in their daily lives and how they use this information as a very successful low-cost method for forming attitudes, behaviors and political judgements that are good enough. In doing so, I highlight the role of recent work by LeVeck (2013), Bond et al. (2012), Banerjee et al. (2013),

and Lieberman (2013). I argue that social information about who is connected to whom is used by political actors when they evaluate decisions about taking political action, whom to support when they do take political action, and what sorts of distributions to provide.

In the third chapter I introduce two cases that serve as the basis for the empirical tests: rural Honduras, and to a lesser extent, Western Ghana. Together with this introduction to the cases, I present my strategy for measuring political actors' social networks in these two locale. I contrast my measurement strategy with two other high quality, comparative social network data sets collected in India (Banerjee et al., 2013) and the Phillipines (Cruz, Labonne and Querubin, 2015). In doing so, I highlight the importance of measuring high-fidelity personal social networks, and demonstrate the expansion in the types of hypotheses that can be tested with person to person data.² In describing my measurement strategy, where necessary I note shortcomings in previous data gathering tasks, and the ways in which this data collection avoids these previous limitations.

In the fourth chapter of this dissertation I test how social network features shape the electoral fortunes of candidates in an election where I am able to fully control the slate of candidates. My theory of social information predicts that candidates who have a large number of social connections within their social network will win more votes, a prediction borne out by the data. Moreover, though, through a unique feature of this design, it is possible to measure the specific vote choices made by people in the election. With this information, I am able to assess why it is that certain candidates win a greater number of votes. The evidence suggests that individuals are more likely to vote for someone who is socially proximate. Above and beyond this proximity effect, I also find compelling evidence that those candidates who are well connected in the social network win more votes from throughout the network. This result implies that there is some feature of being well connected that leads candidates to be more attractive in an election. I argue that this feature is the development of common knowledge that arises as being more central.

In the fifth chapter I examine how political actors' ability to mobilize others to take political action is shaped by their social network. A social information theory predicts that because political action is costly, mobilization is more likely to

²In contrast, the measurements in Banerjee et al. (2013) are household to household networks and those in Cruz, Labonne and Querubin (2015) are family-name to family-name networks. As such, each of these networks encode ties between groups of individuals, rather than individuals *per se*.

occur among individuals who are socially proximate. This chapter directly informs previous findings in door to door mobilization efforts have recently been featured both as highly influential strategies for shaping who participates in elections (Gerber, Green and Larimer, 2008; Gerber et al., 2013), and also much ballyhooed for malfeasance (LaCour and Green, 2014). In this chapter, I find similarity in the importance of being well-connected in bringing others to participate. However, I also find an important difference. Whereas a considerable number of voters voted for candidates who were distantly connected in the social network, in this mobilization task, there was much less reach across the network.

In the sixth chapter, I present theory, simulation, and empirical evidence about a novel way of identifying well-connected individuals within social networks. This method, based on a known property of social networks known as the friendship paradox, combines previous knowledge together with the feature that actors in a social network are self-aware. Previous arguments have used a simple two-stage random sampling to as a method of identifying individuals in a social network who are relatively well connected. I propose modifying this task by sampling the connections of the seed node with purpose, rather than at random. In particular, rather than sampling at random from the friends of a seed node, instead I demonstrate that asking the seed node to name his or her *best connected* social alter may significantly improve the performance of this targeting algorithm. I simulate the theoretical improvement in targeting performance based on this change, and then demonstrate using data from Ghana and Honduras that (a) individuals are able to reliably nominate their well-connected alters; and, (b) individuals do not simply nominated the single best connected individual in their social network.

In the final chapter I review the evidence presented in support of this social information theory, and look forward to future tests which can continue to examine where, and in what ways these forms of social information are likely to shape political behavior.

Chapter 3

A Theory of Social Information

People are inherently social. . . But, while people may be social, citizens are not, at least according to some of the most recognized models of political decision making. Indeed, most standard accounts—including those from a rational choice or psychological perspective—focus strictly on an individually based calculus.

– Betsy Sinclair, *The Social Citizen*, xi

3.1 Norton Shores

Norton Shores, Michigan held special election to replace a city-council seat in August 2014. A council member and her family had moved and their new home, though only a few blocks down the street, was outside the city lines. By rule, she was required to give up her seat on the council.

To fill the vacancy, the council would elect an interim member to hold office for two years until the next regular election. By rule, the council had a relatively brief thirty-day window to field applications, solicit nominations, and then elect the interim-member. Complicating the council's task, on the application window's final day seven applicants filed their paperwork, bringing the total number of candidates for the vacant seat to eleven. Many of the candidates were indistinguishable on political observables: most had no prior political experience; many were successful area-business owners; party identification was a weak cue for a largely non-partisan body. Finally, rather than a campaign for the seat, candidates had only the opportunity to make a two-minute position statement to the council under normal Robert's Rules in the general meeting.

The city attorney, together with city administrative staff, collected written statements from the candidates and provided a dossier to council-members in an attempt to enrich the information environment. However this document was furnished a scant 24-hours prior to balloting. In an effort to keep the decision making process in the public light, a formal rule was set that prohibited back-channel communication between council members. All told, the large quantity of candidates and small quantity of political information created a political coordination problem: when it came time for balloting, which of the eleven candidates would the council stand behind?

Despite the dearth of formal political information, there was rich informal, social information at play in the election. Several members were approaching their second decade on the council and had developed friendships and close working partnerships with other members. Each member knew with whom she tended to work, but, importantly, each also knew who the other council members worked with. In short, each of the members was aware of their own social ties, but were also aware of the broader, council-wide *milieu*. The question at hand in this dissertation is whether, and to what degree this social information might shape the outcome of the election.

The election was to proceed as follows: each of the sitting council members would publicly nominate one candidate to stand for election from the applicant pool. After nominations, a round of voice voting would take place. If a candidate received a majority of the votes, that candidate would be seated in the council. In the event that no candidate won a majority of the votes, the candidate or candidates with the lowest vote total would be eliminated. Balloting would continue in this fashion until a single candidate held a majority of the votes. Because both the nominations and voting were to be conducted by voice, each of the council members had knowledge not only of who had nominated each candidate, but also for whom each council member had voted.

Council members nominated seven unique candidates stand for balloting. That seven of the eight members nominated a unique candidate suggests the rule limiting outside communication had been effective and there was limited, if any, pre-election field-clearing or coordination. Nominations included the owner of a light-manufacturing machine shop, a yoga instructor, several retirees, and a teacher.

In the first round of voting, the votes broke in favor of one candidate, although the total was not decisive. One candidate received four votes, one candidate

received two votes, and two candidates received one vote each. The candidates who had not received any votes were eliminated, and a second round of balloting took place among the remaining four candidates. The second round of balloting was decisive – the frontrunning candidate received an additional two votes and was seated as the interim member of the city council. How did the council come to agreement on a candidate with such rapidity? What tool did the Norton Shores' city council use to agree on a candidate?

Several considerations could have been in each members' calculus. Council members may have cast votes for candidates who matched their political leanings, or who matched their age, or whose careers were in a similar industry, or whose ethnic background were similar. Each of these observable indicators are low-cost ways for council members to assess the skill, viability, or preference overlap of each candidate. Yet, a review of the characteristics of the candidates suggests that these features were not at play. Instead, it seems that the council members used another readily available information shortcut: the social information of those on the council.

In fact, this city council can be roughly split into two generations of members, one older and one younger. The winning candidate had been nominated by one of the older generation of council members, a council member well regarded by each other member of the older generation. As is represented in Figure 3.1, each of the members of the city council held respect relationships with others on the council. And, the winning candidate had been nominated by the candidate who was best regarded within the largest bloc of council. To be certain, this city council example is not dispositive of the role of social connections, although it is instructive.

In the absence of clear preferences for one option over another, if each council member held some weak preference for being a part of the winning coalition then each council member's task reduces to a coordination task. Winning depends on making good predictions about the behavior of others. These predictions become very easy when there exists some institution that allows for the credible signalling ahead of the vote, but without such an institution, the predictions are quite difficult (Schelling, 1960). The contention in this fable is that without an institution to signal preferences, instead council members used what tools they possessed to reach a decision.

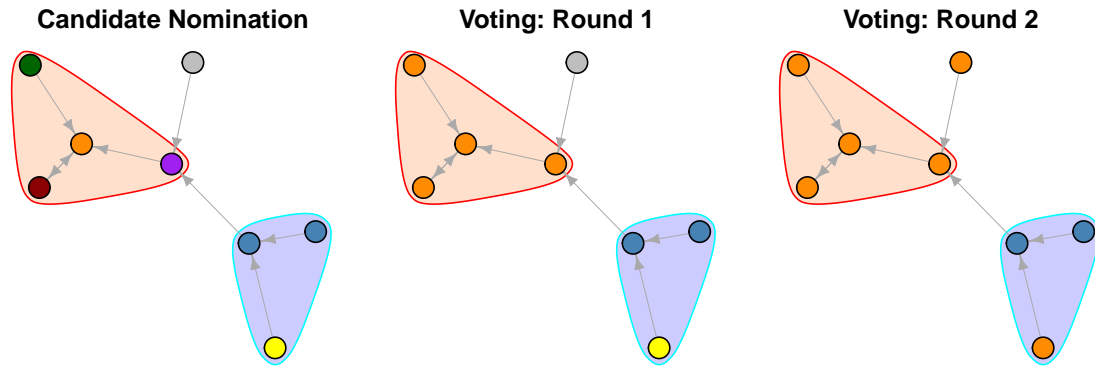


Figure 3.1: Nominations for candidates to city council. Each circle represents a seated council member, each arrow represents a directed relationship between council members in response to the question, "For whom on the council do you have respect?" Colors of the circles correspond to candidate nominations; each color represents a unique nomination.

3.2 Heuristic Information

Popkin (1994) builds on the framework of Simon (1955) and Downs (1957) to argue that voters use easily deployed criteria to evaluate candidates. This framework, rationalized by positive political theorists, but whose core insight is drawn from the earlier Columbia studies, argues that "Voters are not always aware of what the government is or could be doing," (p.13 Popkin, 1994, citing Downs (1957)), and yet the government continues to function in a more or less successful way. The core contribution of the Popkin (1994) work is to identify the information market within which voters exist. Recall, the core analogy that Popkin uses is that of the voter as an investor. Rather than a voter purchasing a good as if in an insurance market, instead the voter is trying to make a long-term bet on an mechanism that has some uncertainty about the return. Popkin argues that if voters can obtain two pieces of information that equally allow them to discriminate between predictions about future performance the voter is likely to prefer the less costly signal.

Less developed in Popkin (1994) is the dichotomous feature of mass political decision making. Popkin's theory of a reasoning voter is clearly located in the context of national two-party elections in the US. Thus, a tacit assumption of Popkin's theory is that at every point voters are making a single, dichotomous choice. In the primary phase, voters can choose to support the frontrunner of the party, or not; in the general phase, voters can choose to turnout or not; if they turnout to vote, voters can choose to support the Republican, the Democrat. At

each step, for those voters who use a heuristic, that heuristic operates in a decision that is constrained to be between two choice. Many elections in other contexts share this same dichotomous feature. The lack of breadth of this choice, together with the forced election-day timeline, means that it is the role of campaigns to stack diverse factions onto a single, energized and action-ready dimension.

For the voter, the forced timeline and dichotomous choice mean that for those relying on heuristic decision making, very little information can sway a vote. As an example, Popkin identifies Presidential hopeful Gerald Ford's admonition to, "Always shuck your tamales." In the run-up to the 1976 Republican presidential nomination, Ford was touring through San Antonio, TX. When food was passed at the gathering, Ford took a tamale from a plate and took a bite, husk and all. Harmless as it was, the implication of this gaffe for Hispanics, Latinos and Mexican-Americans is that someone who was unfamiliar with how to eat *tamal* must know so little about the preferences of the group of voters that he must have been the wrong candidate.

While existing formal institutions clear the field in many cases, a considerable amount of political decision making occurs that is not simply a dichotomous choice. Indeed, many decisions where the output is black and white are argued in shades of grey by political sophisticates and broadcast to relative novices in dichotomous form (Zaller and Feldman, 1992). These cases might be about decisions for funding levels for a block grant, numbers of troop to deploy to a peacekeeping mission, time to allocate to a community project. Minimally, at least, there are location in the political system where coordination needs to occur, but the set of decisions has not been cleared by some institution that facilitates coordination. For example, consider the voting example in Norton Shores: in the first round of selection, council members needed to complete the difficult task of coordinating on one of a large slate of options with little information. The second round of selection presented a less difficult task because of the field-clearing and signaling in the first round.

In this chapter, I argue that *social information* is a prime candidate for shaping electoral considerations. Especially when established, formal institutions do not influence decisions, commonly held information by political actors may shape decisions and outcomes. To make this argument, first I locate the importance of the question in the behavioral political science literature. This begins with the Columbia School, but continues into an active programme aided with technological developments and richer data sources. I draw on a growing body of research

in neuro-psychology. This research, perhaps called "social neuro-psychology", or simply social psychology, argues that one of the core features of human psychology is in our capacity for processing social information. Indeed, some even argue that "We are wired to be social," (pp. ix, Lieberman, 2013). If the social psychology hypothesis is correct, and one of the key features of human cognition is social processing, then, I argue that social information might be thought of as a *very* low-cost type of information. I combine this theory of low-cost information acquisition with formal theory which predicts high prominence (social well connected) actors are uniquely capable to shape group behavior to make predictions about political behavior.

I build upon information transmission and norm-based theories and propose a complementary mid-range theory: the social network itself conveys information to those who are embedded within it. I refer to this notion that the social networks contain information as a theory of "social information." Social information, I argue, is the product of nuanced interactions between social actors. Following sociological arguments, I reason that this information is context-dependent, meaning that the question at hand shapes the form of social connections that are salient (Gee and Jones, 2016). Finally, and most critically for this work, I argue that social information is task-oriented; depending on the task to be completed, actors use social information in distinct ways.

The desire to form and hold social connections are strong individual motivations (Baumeister and Leary, 1995), and falls in the first half of Maslow's *Hierarchy of Needs* (Maslow, 1943). Social information is generated through repeated interactions between individuals. Prominent theories of tie formation identify the role of similarity between individuals in the formation of friendships and social connections (Kandel, 1978; Mcpherson, Smith-lovin and Cook, 2016). Individuals with shared interests, similar personality traits and personal histories tend to form social connections with one another in a process termed *homophily*.

The development of friendships between similar individuals, together with repeated social interactions among these groups of friends facilitates the possibility that individuals hold broad *social information* about others actors in the social network. The residents of the Columbia Studies' New Haven and Elmira frequently reported discussing political events with others in their churches, neighborhoods, and social groups. Huckfeldt and Sprague's account of political life in South Bend identify a considerable number of reciprocal connections between political

discussants. Due to the repeated nature of interactions between actors, actors can come to know the general landscape of their social networks.¹

First, well connected actors likely have attained such status by their actions in past exchanges (Csányi and Szendroi, 2004; Acemoglu and Jackson, 2015). Knowing this, the actions of these highly connected actors might serve as a useful heuristic for other political actors to satisfy and reach a decision (Simon, 1955). Second, and in parallel, friends and family members are more likely to satisfy representativeness heuristic criteria. As a result, the actions of closely related political actors might serve as a complementary decision criteria. Actors may also make intentional use of social information to make coordination and distribution decisions, especially when payoffs are conditioned by beliefs about the actions of other actors. Third, actors may simply hold a preference for taking actions similar to those with whom they share social bonds. Although it is an important question, the core contribution of this proposed dissertation is not to adjudicate between theories about why or how social networks actionable information, but rather to suggest that this information is at play in ways that shape important political outcomes even in the absence of communication.

Even if actors do not have an explicit knowledge of the aggregate network structure, they do have heuristics to intuit their social system's important structural elements. For instance, children in a classroom know which other children are popular; politicians in closed door discussions about military action know who is well connected in the Democratic organizational committees (DCCC); citizens discussing conservative politics at local meeting places (e.g. Washtenaw Dairy, Walsh (2004)) or out at the local surf break know whose opinions are likely to stick within the group. In this way, popularity and network centrality measure the same phenomena; both concepts identify social information that makes an opinion or action likely to be observed. These actors also understand social cleavages: children understand that some are jocks while others are nerds; politicians understand who are hawks and doves; surfers know who are long-standing locals and who is not. These distinctions outline community structures identified by algorithm in social network graphs. Finally, these actors can even integrate the connectedness and community structure concepts. Children can identify the most popular jock and the nerdiest nerd; the most influential hawk; the best connected local.

¹See, especially the targeting and use of social information chapter of this volume, but also Enemark, McCubbins and Weller (2012).

In each case, these actors are a part of a social system that shapes the likelihood an action is observed. A class of political outcomes, most notably coordination tasks, are shaped by the beliefs of actors. That is, while actors may have preferences for a particular outcome, equilibrium reasoning requires they make some assessment of the preferences, cognitive structures, etc. of other actors in the game, and it is these beliefs about the actions of the alter that in fact shape eventual outcomes. For a child deciding which clothes to wear, if she holds a preference for a blue shirt but also preferences to coordinate her shirt color with others, then her beliefs about the shirt color of others shapes her eventual choice. Similarly, a citizen discussing conservative politics might have a preference for antagonistic tax policy, but also a preference to not be the first mover in her social group to identify with the Tea Party because she dislikes dis-coordination.

3.3 Scope of the Problem

An important strain of political science research identifies the role of the social environment to shape political outcomes. Berelson, Lazarsfeld and McPhee (1954) anchor the sociological political science tradition and squarely rest their account of political life in Elmira on the *social milieu* – the set of relationships and norms voters hold. “To a large extent, political discussion follows the composition of friendship groups,” (Berelson, Lazarsfeld and McPhee, 1954, p.103). This work laid the foundation for later scholars to measure how civic-mindedness affects political participation (Verba and Nie, 1987), assess how discussion networks shape identity formation (Huckfeldt, Johnson and Sprague, 2004; Huckfeldt and Sprague, 1995), and examine how shared norms (Rosenstone and Hansen, 1993; Sinclair, 2012) and conditional cooperation (Rolfe, 2012) explain variation in individuals’ levels of political activity.

Although political sciences’ sociological tradition identifies a natural place for social context to influence citizens’ behaviors, many past empirical works in the paradigm have not measured important parts of this social context or have relied on imprecise approximations or somewhat clunky aggregations. Two influential books published in the last year develop extensive theory about how social context shapes Americans’ participation in electoral politics (Sinclair, 2012; Rolfe, 2012). Sinclair (2012) argues that a preference for conflict avoidance leads voters to form preferences that match those of political discussants. Rolfe (2012) develops a differ-

ent theory that comports well with Sinclair; she argues that voters take conditional action based on the actions of those political actors with whom they interact. This conditional behavior immediately calls to mind the work of Gerry Mackey and behavioral norms. However, because data on the social relationships their theories implicate does not exist, both scholars instead rely on imprecise geographic proxies reasoning that living in the same Congressional district increases the likelihood that two voters share social ties. As a result, in both Sinclair and Rolfe present results that are suggestive of a role for social context, though most of the hypothesized relationships were not borne out by the data at traditional levels (e.g. Sinclair, 2012, p. 68).

Previous theories have argued that social relationships serve as an important conduit for information (Almond and Verba, 1965; Robinson, 1976; Huckfeldt, 1979a; Huckfeldt and Sprague, 1987; Kenny, 1992; Rosenstone and Hansen, 1993; Walsh, 2004; McClurg, 2004, 2006; Klostad, Sokhey and McClurg, 2013) and form the basis of norm based behavior (Wolfinger and Rosenstone, 1980; Spitzer et al., 2007; Sinclair, 2012; Goette, Huffman and Meier, 2006). Rolfe (2012) notes, “Citizens who are asked to vote are more likely to do so”; in 1920s, multi-ethnic Chicago, Gosnell (1927) found that women’s social-groups affected women’s probability of turning out to vote. In a nationally representative sample, Rosenstone and Hansen (1993) find higher turnout among members of social organizations, and Verba, Scholzman and Brady (1995) argue that “congregational churches” and other civic organizations train their congregants and members in the civic skills necessary to be effectively engaged in politics. McClurg (2004) finds that party contact of political discussant pairs can spill-over from one discussant to another, changing the substance of what these pairs discuss. These theories hold that information largely flows along social channels, and that understanding the contours of these channels might help us to better understand political and behavioral outcomes (Mutz, 2002; Friedkin, 1998).

Much of this literature studies the flow of information between actors, and so focuses on the transmission and diffusion of information explicitly related to electoral politics (Huckfeldt and Sprague, 1987, 1995; Walsh, 2004). The most recent work in the area has focused on the related notions that social networks engender norm-based (Sinclair, 2012), conditional (Rolfe, 2012) responses to stimuli. and has contributed to our understanding of how political opinions and identities are formed in mass politics.

Sinclair and Rolfe (independently) argue that voters' decisions to turnout to vote or participate in local-level political activism are shaped by individual predispositions of voters and how those predispositions change when voters interact with one another. Sinclair argues that broad political outcomes can be well described by a conflict-avoidance mechanism whereby voters choose positions that are in line with the positions of their social connections. In this model, because one selects social alters that are likely to share tastes, and beliefs, and, indeed genetic material (Christakis and Fowler, 2014), taking the position of social alters both does a good job of taking a position near one's own ideal point, but also minimizes the conflict between social alters who take different positions.

Rolfe arrives at a similar conclusion, but argues that political action is the result of conditional-cooperation filtering through a socially-connected world. This view takes a tit-for-tat, game theoretic, view of political action, and argues that there are a small number of actors in society that will take costly political action independent of the actions of others, but that others' actions are conditional on being forced into action by political active members of society. While Rolfe identifies conflict avoidance as one of several mechanisms driving this social effect, she is not interested in specifically engaging the mechanism. Rather, Rolfe's modeling focuses on the effects of broader social-structure rather than interpersonal influences.

3.4 Social Heuristics

Social cognitive neuroscience is the study of cognitive function of people in the social world. This burgeoning discipline is a brackish waters between neuroscience and social psychology (p. 143 Lieberman, 2010). In this role, social cognitive neuroscience (hereafter SCN) combines the tools common to neuroscience – the fMRI chief among them – with the tools of social psychology – the laboratory experiment – to examine how the brain functions in social interactions. My aim in this section is not to convince the reader that every test of theory in SCN point toward a specific role for social-cognitive function in political decision making; others make this argument as it relates to politics (Lieberman, Schreiber and Ochsner, 2003; Fowler and Schreiber, 2008). Instead, my goal is that it is quite likely that social cognition and therefore social information are likely to arise through daily activities.

Social Cognitive Neuroscience

Social cognitive neuroscience provides two useful structures as it concerns this dissertation. First, the brain seems to evince some sort of measurable activity when placed in social circumstances. Second, it seems to be the case that when the brain is not otherwise tasked, similar neural regions are being utilized as when individuals are actively thinking about social context. This set of cognitive architecture is termed the Default Mode Network, as it the architecture that is active when subject are not tasked with positive stimulus, when subjects are in their default state (Lieberman, 2007; Uddin et al., 2007; Buckholz et al., 2008; Izuma, Saito and Sadato, 2008; Lieberman, 2010; Boksem et al., 2011; McDermott, 2011).

The first useful structure provided by SCN is the identification of a cohort of brain regions that tend to activate when the brain is not under positive stimulus. Second, the SCN has identified that the unlike a large number of other tasks—mathematics, memory recall, emotion, and the like—when behavioral subjects are asked to undertake *social cognition* subjects' tend to keep the default mode network activated. In concert, SCN researchers suggest that these findings imply that social cognition utilizes the same low-cost neural architecture, and that this might mean that out brain has evolved for such social awareness.

Second, after clear patterns of brain-activity had been established, researchers began to examine questions about the brains' function in the absence of positive-task orientation. Researchers examined how the brain functions while sleeping, while meditating, and most-importantly for this work, while actively doing nothing. Shulman et al. (1997) presented the first of this work, which was later followed up with research by Mckiernan et al. (2003) who termed the complex of brain regions that are active under no positive stimulus the *Default Mode Network*. This second, core finding lays the empirical basis for the claim that social information is *low-cost* information for humans and citizens to obtain and process.

Social Cognitive Neuroscience's Limitations

The chief concern which SCN faces is that although the neither the measurement of brain activity though fMRI nor the manipulation of circumstances to prime response² are controversial, drawing a connection between stimuli that activate similar brain regions is challenging. In particular, for all the imagery produced by

²This is the core behavioral experimental paradigm

functional neuroanatomy, the ability to strongly attribute a similarity in purpose for brain regions cannot be supported as a causal relationship. That is, if both stimulus A and B both prime a response in region Q, there is no logical implication that A and B should be related. This concern, while levied against SCN in particular and functional neuroanatomy in particular, is not of concern for my argument of social information.

Limitations Do Not Apply

Previous work has examined neural correlates with the DMN. In particular, Lieberman, Schreiber and Ochsner (2003) finds that for those who are political sophisticates, the evaluation of political events requires little cognitive engagement beyond the default state network. That is, those for whom political reasoning is easy, there is relatively little task-lift that is necessary to process the events. Similarly, the research of Lieberman, Schreiber and Ochsner (2003) finds that evaluating political tasks involves little task-lift from the DMN. While tasks like math, physics, or purposively empathizing with an alter require the activation of neural regions that are not under load at rest, the processing of social information does not require such lift.

It is important to acknowledge that it is possible that there are some measurement or experimental difficulties that are leading to the difficulties with examining the task-lift experienced by subjects when asked to perform social tasks. Maybe processing the social world does require some cognitive feature that scholars have not yet identified. Even if this were the case, then either this feature must be so low cost that it is not being detected with current instrumentation; or, it must operate in a way sufficiently distinct from any other previously understood high-load or low-load processing tasks. In sum, the evidence suggests it is quite likely that *to the brain, the processing of social information is easy.*

Social Information Is Like Other Heuristics

If the processing of social information is relatively easy or is associated with other brain regions that are active when the brain is under easily processed tasks, then it seems reasonable to expect that social information might also be utilized as a low-cost, low-information tool for individuals to make judgments about political actions. Simply stated, it might be that social cognition is a useful information

short-cut that political actors can use to take political action.

That political actors might use data gathered by others to reduce the search cost when making political decisions is not new. Indeed, this is the core thesis of the two-step flow of communication (Lazarsfeld, Berelson and Gaudet, 1944; Berelson, Lazarsfeld and McPhee, 1954; Lazarsfeld, Berelson and Gaudet, 1968; Zaller and Feldman, 1992), and the subsequent work in this paradigm. The theory of social information, however, makes a somewhat different claim. It isn't that the others are providing information; but rather that the easily provided social context can *itself* serve as information that political actors use to form decisions, opinions and take action.

3.5 Example: Social Information Predicting Behavior

Consider, a classic example of the two step flow of communication: the transmission of information from scientists to the mass public. To make the example concrete, further establish the information under consideration to be information regarding anthropocentric climate change and that the transmission is occurring at a dinner party. Under this construction, the transmission would likely contain the following facets. First, there may be a small set of individuals who are specifically knowledgeable in the subject; these individuals maintain an awareness of the current climate forecasts in general science journals, understand the bargaining dynamics, and for whom the annual *Conferences of the Parties* is a news worthy event. Second, there may be a larger set of individuals who participate in a conversation, and whether knowingly or not, adhere to social norms surrounding the discussion; they listen attentively while others speak, form their own beliefs, and the present those beliefs in an appropriate manner.³

At this party, one informed individual might quip, "Climate change is the single greatest political challenge humanity has faced. We should give up our polluting cars, avoid travel by airlines, and forego having children. Only a reduction and global awareness of this scale can save our humanity." While this might certainly seem to be a bleak statement, for many at the party who have only a passing understanding of the problem, this might seem an extreme, but informed opinion. Another, equally abreast individual might respond to the doomsayer the

³To expect that all in attendance would present their opinions in a generally agreed upon, socially acceptable manner, must certainly indicate that this gathering is not a gathering of graduate students.

feasibility of a more moderate position, instead suggesting that a more moderate stance, together with a sensible strategy for reduction, abatement, and innovation might also solve the problem.⁴ Again, the interested party-goers might agree with this position.

These are two positions that reasonable people might disagree about. If party-goers have only these two statements to begin to form an opinion, on what criteria are they likely to make this decision? Or, if we, as political scientists, were to predict the opinion of each of the party goers, what information would we use in our prediction?

If the two experts possessed clear party affiliations, then we might expect this label to solve much of the information problem for party-goers. Those who identify as Democrats would be more likely to agree with the statement of the "expert" Democrat and those who identify as Republicans would be more likely to agree with the statement of the "expert" Republican. Alternatively, if for some reason descriptive representation held decision-relevant information, we might expect party-goers to identify with the "expert" who more closely matches their demographic profile, be that old or young, majority or minority.

However, absent the existence or activation of such strong institutions it would be difficult to predict for whom the attendants of the dinner party would agree. We might posit a null model whereby individuals randomly partition themselves between supporting the position of the hardliner or the pragmatist. We could likely improve our model with information about the party identifications of party attendants, but relatively little more might be added to the model if individuals do not have domain specific knowledge.

A theory of social information might serve as another mechanism by which individuals sort themselves into camps of support. Perhaps some of the attendants are good friends with one or another of the individuals making the climate statements. Or, perhaps someone known to be a non-friend takes a position that sides with one or another of the experts. These seemingly simple positions might go a long way toward forming a clear division of support among the groups.

To this point, the examples have focused on simple cases of low information decision making. However, the usefulness of social information is not limited to such tasks. Hafner-Burton, Hughes and Victor (2013) argue that complex information environments may lead even highly experienced decision makers to

⁴This very circumstance occurred while I was writing this section of this dissertation.

pursue learned, successful heuristic means of processing information. Domestic and international politics present several contemporary examples of where social information might serve as a useful heuristic: Republican delegates, if the nomination should come to a brokered convention will be faced with a dramatic example of a coordination task which might be more easily solved using social information; Democratic party superdelegates have a strong incentive to coordinate their actions on one of the two possible nominees for the Democratic ticket; Supreme Court justices, when arriving at a decision within the short-staffed court have an incentive to make unanimous decisions to protect the authority of the court. In any of these information-complex, high-politics issues, there is ample opportunity for social information to be used by decision makers.

3.6 Social Network Knowledge Shapes Messages

A theory of social information makes a more specific prediction than previous information processing theories. After acknowledging the information processing of political actors, a social information theory identifies an additional, systematic component of information evaluation: other actors' positions in the social network. Two features can be evaluated under this social information theory: first, actors evaluate whether the sender of information is socially proximate or a close connection; second, actors evaluate whether the sender of information is core or peripheral to the group of actors.

Concretely, a theory of social information implies that for some political action, be it mobilization toward a public good, the selection of a preferred candidate, a position on an emerging issue, or which Republican nominee to support, it would not be enough to know what someone has said, without knowing where in the social network they were when they said it. That is, the literal content of an information or preference statement is modified by the sender of the information. The opinion leadership literature acknowledges this relationship in elite-mass communication, identifying either mass deference (Converse, 2006; Lodge and Taber, 2005; Tomz and Sniderman, 2005; Bartels, 2005) and elite persuasion (e.g. Richard F. Fenno, 1977) as means whereby the messages sent from elites to masses generate greater response than might messages sent from masses to masses.

To argue by analogy, political actors' social networks provide actors either a "social soapbox", or a "social sandbox". For those who are well connected to others

in the town, the messages they broadcast are benefitted by being well connected to others. Conversely, for those who are poorly connected, broadcast messages are likely to find only a limited audience. The soapbox and sandbox analogy highlights the differences in ability of actors to broadcast actions and messages; however, more than this, it also helps to identify a more nuanced point. The effectiveness from standing on a soapbox, at least as it was practiced in town squares of yore, benefitted the speaker not only by raising the speaker above the crowd, but also through the reciprocal nature that all in attendance could observe the speaker, and knew that others in attendance were also in observation of the speaker. Not only is what they say more likely to be heard, or actions they take to be observed, but also there is a reciprocal effect where others in the network are also aware of this increased likelihood of observation.

The argument for preferential learning from individuals is not new to this work. Indeed, the dual inheritance theory of human evolution posits two pathways through which humans have developed: first, the commonly identified genetic mechanism, and second, the less commonly identified cultural mechanism. Henrich and Henrich (2007) present this model in this theories' most clear terms, drawing on the pioneering work of Boyd and Richerson (1985) (Lewens, 2013). Dual inheritance theory holds that individuals are in complex information environments (Simon, 1955, 1956) and as a means of reducing search costs, actors may seek to mimic the patterns of other successful actors in the system, even those who are non-kin. Mimicry can lead to increases in fitness within a single generation; however, due to the complexity of the information environment,⁵ it may not be clear to actors whom they should mimic (Boyd and Richerson, 1985). The task to each actor, then, is to develop a learning rule that mimics behavior that is fitness-beneficial without copying behavior that is not fitness beneficial or even fitness detrimental.

The argument for this mimicry is perhaps best developed by example. Suppose an individual were recently settled in San Diego, and, taking a job with a biotech firm in the area, aimed to build a custom home to settle a family. What architectural aspects should this new home possess? In what neighborhood should the home be built? The biologist, though an educated person, is likely to have spent little time considering such things as egress, coastal protection laws, and perhaps

⁵Here, complexity is used in a manner consistent with complex systems research. As such, complexity means the possibility of unclear relationships, *ex ante* between input values and output values due to the possibility of feedback loops and unobserved order.

worst of all, NIMBY neighbors. The biologist may move to hire an architect to aid in the design and construction of a home, but even this process is difficult with no prior information about what makes someone a good architect. How then, should the individual come to reach a decision? As a social learner, the biologist might consider company peers and simply mimic their choices. This decision might cause the individual from committing the obvious mistakes — living too far from campus, living in a neighborhood blighted with foreclosures, Pacific Beach — but this social learning may not readily identify an optimal solution.

Two separate theories propose an additional refinement to the social learning process. The first, *prestige biased learning* begins from the complex information position that leads to social learning in the first place, and additionally posits that individuals use *prestige*—the cues of individual success provided by others in the system—to overcome the additional task of whom to mimic.

“Figuring out who possesses the most adaptive skills, strategies, preferences and beliefs is no straightforward task. To achieve this, people rely on a range of cues related to skill, success, and prestige. . . . Once the psychological machinery that makes use of competence- and success-based cues for targeted cultural learning has spread through the population, highly skilled and successful individuals will be in high demand, and social learners will need to compete for access to the most skilled models,” (p.12, Henrich and Henrich, 2007).

That is, if social learning is commonplace, those individuals who are most skilled will be in demand from others who seek to mimic their behavior. As such, some market must exist between those who are being mimicked and those who are mimicking. The contributions that a *social information* conceptualization make to this existing theory of prestige biased learning are two: social information provides a theoretically motivated conceptualization and operationalization of prestige, namely centrality measures calculated using social connections between individuals; and, relatedly, a justification why a mimicry “mentor” may choose mimicry “mentees.”

The second theory that refines social learning targeting is that of *prominence biased transmission* due to Acemoglu and Jackson (2015). Working with the coordination framework of social norms, the authors argue that some actors—those who are socially prominent—are uniquely capable of shaping norm-based outcomes because the norms serve as frames of reference that coordinate agents’ expectation about

behavior. In this theory, prominence determines the probability that an actors' actions are observed by others. Highly prominent actors are actors whose actions are observed with certainty by all others in the system, while non-prominent actors are those whose actions may not be observed by all. This framework is commonplace in models of evolutionary game theory (e.g. Mailath and Samuelson, N.d.), and has been profitably utilized to address female genital mutilation (Mackie, 1996), turtle hunting (Smith and Bird, 2000), as well as political party cues about the economy (LeVeck, 2013).

In many ways prominence biased transmission is the complimentary mechanism to prestige biased learning. Whereas prestige biased learning points individuals to take prestige as a signal of successful strategies and therefore an individual who may be usefully mimicked, prominence biased transmission suggests that the opinions, behaviors and actions of prominent individuals are more likely to be commonly observed. As such, when actions require coordinated action—as in norm formation and elections—the actions by prominent individuals are more likely to be taken up by other members of the groups. However, notably missing from the theory of prominence biased transmission is a measurable characterization of what features might make an individual prominent. As with prestige biased learning, the contribution of this *social information* theory is to provide a working operationalization of prominence. Specifically, actors who hold a large number of social connections in a social network are prominent actors; actors who do not hold a large number of connections are non-prominent actors.

3.7 Past Social Network Approaches

The recent acknowledgement of the connections between actors have helped bring about the rapid development of both theoretical and empirical work. However, with a few notable exceptions, many of these scholars have conceptualized network effects in one of two ways that are, at their core, un-networked. The first type of simplifying assumption is that of a group-level examination (e.g. Putnam, 1993, 1995; Sinclair, 2012); the second type of simplifying assumptions are those of peer-effects (e.g. Sacerdote, 2001; Fowler and Christakis, 2008; Fafchamps, Vaz and Vicente, 2013). Both the group-effects and peer-effects conceptualizations fail to utilize the core insight of the social network approach, and, as a consequence, neither approach is able to make sufficiently specific predictions about political

behavior as a result of the actions of social actors' political behavior. In this section, I present these concerns through the lense of political mobilization and vote choice.

3.7.1 Political Mobilization

First, consider the case of political mobilization. A cohort of scholars have identified group-level evidence for social considerations in the choice to take political action. Fowler (2005) argues that a model with minimal social mimicry and close social connections may lead to group-level externalities for a single individual choosing to vote. Cox, Rosenbluth and Thies (1998) argues that rational mobilizers (campaigns) "should target supporters plugged into wider and more tightly knit social networks, in the hope of producing favorable secondary mobilization," (Cox, Rosenbluth and Thies, 1998, p. 448). They find that rates of turnout correlate with district-level measurements termed "social capital." Gerber, Green and Larimer (2008) find that the most effective mobilization primes social pressure from a group of neighbors; high levels of turnout among elderly Japanese women can be partially attributed to the *social obligation*. Fowler (2005) predicts that individual decisions to turn should have group-level consequences, at least when there is some incentive for individuals to behave like others in the social network, and the social network is sufficiently small that it is possible to observe others' behavior. Sinclair (2012) finds that congressional districts with large numbers of donors—districts argued to have high levels of social capital—also donate more to political campaigns than districts with lower levels of social capital.

Collectively, these scholars surround a position that increased numbers of social connections are associated with an increased likelihood of individual political mobilization. However, two limitations hamstring each of these studies: the authors are neither able to make a strong claim that increasing social connections causes the increase in mobilization; nor can any of this set of authors make a specific prediction about individual level behaviors. For example, consider the most specific prediction possible under the empirical results of Sinclair (2012). A district identified have a large number of coincident voters—that is, one identified as possessing high social capital—is predicted to raise more donation money than one with lower levels of political capital. But these predictions exist only at the group level. Under these results, knowing some specific individual has donated to a campaign is uninformative about whether another is likely to also participate.

The lack of specificity in prediction of mobilization outcomes continues among conceptualizations of social networks as collections of peers. This conceptualization, rather than measuring group-level outcomes, instead, focuses on the specific relationships held between individuals. A leading example of this work, which also demonstrates the limitations of the programme, is that of Bond et al. (2012). In this work, the scholars use an excellent experiment that manipulates the information environment of voters, randomly exposing voters to information that social contacts had taken political action. In contrast to the group-level programme, studies of this class can make limited predictions about behavior at an individual level, however, the specificity of these predictions are limited only to average predicted changes as a result of the myopic evaluation of alters' one degree separated. More specifically to the point, these studies make the strong assumption that the effect of *one particular* social alter are same as the effects of *any possible* social alter. This simplifying assumption does not jibe with the empirical reality. There are both social and non-social reasons to suspect that the actions of one social contact convey a different meaning than the same actions of another social contact.

The limitations faced by Bond et al. (2012) as a result of the dyad-collection view of a social network is also faced by others. Nickerson (2008), in an inventive field experiment, demonstrates the spillover of door-to-door canvassing. However, in this study the design collapses all types of cohabiting adults into a single category. This type-imprecision might be solved with a straightforward model that classes social alters; however, the dyad-collection view essentially pushes off the reductionist, rational actor assumption by one layer. Rather than assuming the core unit of analysis is the individual, the peer-effects conceptualization acknowledges the existence of extra-individual considerations, but it does not actually model the broader sets of connections. In effect, the peer effects framework fails to fully address the problem that it sets out to solve; although the peer-effects scholars acknowledge the role of system level influences, the models in this paradigm do not meet the requirements of these system-level considerations.

3.7.2 Political Identification and Choice

Second, consider the case of vote choice and political identity. Questions of this time are both durable, with well-known connected explanations for vote choice extending back as far as Berelson, Lazarsfeld and McPhee (1954).

Candidates have a clear incentive to try and coordinate the behavior of the people who are voting. If they can lead voters to hold similar views and valences on the issues, candidates can reduce the variance between the opinion of the median voter in their win-set and the opinions of all the actors in that set. Even more, and of primary importance, need to coordinate the actions of voters to vote them into office.

It is perhaps less clear why the voters would like to be coordinated by a candidate. This type of coordination might mean changing attitudes and beliefs that are loosely organized (Campbell et al., 1960; Converse, 2006) or drawn from a distribution of attitudes (Achen, 1975; Zaller and Feldman, 1992). In this paper, by design there is neither communication between candidates and voters, nor communication between voters. As a result, I assume that voters' beliefs are unchanged when presented with a slate of candidates. Consequently, the design leaves voters with only two primary considerations. First, the voters would like to select a candidate who will do a better job. Second, for both psychic and intrinsic considerations, voters would like to be among the group of voters who supported the winning candidate.

3.8 Hypotheses

From a theory of social information, it is possible to develop a series of discrete, observable tests of individual behavior. In these paragraphs, I present a number of these hypotheses. While the hypotheses presented herein hue toward those that will actually be tested in this work, included as well are other hypotheses that might be tested in future work. In particular, I present a series of hypotheses about distributions of goods that might be tested in future analysis using data already collected as a primary part of this dissertation research. Finally, where it is possible, I highlight the difference between the predictions that results from a theory of social information and more standard theories of rational and behavioral political actors.

Two main empirical tests are brought forward in this work. One test is a test of political mobilization wherein a random sample of individuals are chosen from within the residents of villages and are provided an incentive to recruit as many individuals to take costly, public political action. The second test is of political choice: in these same villages, a separate assignment procedure randomly assigned

individuals to stand for representation in a village council. Both tasks require the transmission of information from the seeds assigned to hold it to the others in the village. However, distinct between the two tasks is the particular set of incentives for the task outcome. This difference — coordination in the election experiment and influence in the mobilization experiment — provide the distinct context to evaluate the functioning of social information.

Despite the differences in the task incentives, there are common predictions from a theory of social information that are not made in a theory that does not acknowledge the possible role of individuals social relationships. First, and foremost, actors who have a greater number of social relationships are predicted both to receive a greater number of votes in the election task, and also to mobilize a greater number of peers to take political action in the mobilization task. This hypothesis has been addressed in the context of organizational leadership (e.g. Mehra et al., 2006; Balkundi and Kilduff, 2006), though little of this research has been incorporated into the political science literature. Second, because social networks form the basis of communication in the study population—and, indeed, all populations—political actions taken by subjects should be observably related to their relationships. However, this general prediction is modified by task-orientation. In cases where either action or communication are costly, actors are predicted to be more successful at shaping behavior among relatively close, or proximate social connections. Examples of costly action range from cosponsoring legislation, to borrowing money, to participating in a protest event; costly communication might involve holding an extended conversation about a politically contentious topic or communicating a lengthy policy platform. For clarity, these hypotheses are also presented in Table 3.1.

Table 3.1: Social information hypotheses are listed for both mobilization and election tasks. Distinct between the two tasks are the cost of subjects' participation.

Task	Hypothesis
Mobilization	<ol style="list-style-type: none"> 1. More social connections → more alters mobilized 2. Costly nature of a attendance → socially proximate alters mobilized
Elections	<ol style="list-style-type: none"> 1. More social connections → more votes 2. Voter preferences for representativeness OR rents → socially proximate votes 3. Voter pref. for winning (coordination) → socially well-connected votes from diverse network locations
Knowledge	<ol style="list-style-type: none"> 1. Individuals can reliably answer name generator questions. 2. Individuals understand social network <i>meso</i>-structure

Chapter 4

Social Network Data Collection in Rural Areas

4.1 Local Settings

The persistent question throughout this dissertation is the measurement of how political behavior is shaped by the connections between political actors. The measurement of the connections between individuals is remarkably difficult: ties between individuals are frequently context specific (Sokhey and Djupe, 2014), determining just who should appropriately be included in the conceptualization of a social network is frequently unclear, and errors in this determination generate highly unreliable estimates of social effects.

Recent developments in how individuals interact with one another through computer and internet mediated fora have, for the first time ever, facilitated the analysis of connected data. However, the form of social interaction through computer mediated platform leads to the creation of a large number of social connections that are not actually utilized for interpersonal interaction. While scholars have developed methods for determining the strength of online relationships (e.g. Jones et al., 2013) and these methods have been increasingly implemented in facilitating and targeting online interactions, a research strategy utilizing these technology would necessarily have both a computer and a predictive model between the researcher and the measurement of social network information. Other scholars have used clever measurements strategies to measure the social networks of individuals in India (e.g. Banerjee et al., 2013), and families in the Philippines Cruz, Labonne and

Querubin (2015).

In this work I argue for the intuitive, nuanced use of social network information. A key component of this work is the careful measurement of real-world relationships between individuals. These relationships are the kind of social connections that come to mind when we think of the community to which we belong, the workplace connections that we value, and the information networks that we trust. In short, rather than networks of weak-ties (e.g. Gee and Jones, 2016).¹, these are the ties that we might consult when considering an important decision or hosting a gathering of friends. These networks, it is taken in this work, are networks of understanding and influence that are built through repeated interactions; reputations and the meanings and functioning of social ties are nuanced and difficult to measure on a massive scale or reasonably generate in an experimental settings through structured interactions.

These measurement properties lead this project to collect data in small-scale settings. These settings, Honduras and Ghana, help solve many of the most difficult problems in the measurement of interpersonal connections. The size of towns makes talking with all community members possible; the separation between towns makes for natural break-points in the community (in fact) and in the measurement of the community (in practice).

These small scale communities possess the features of all communities. Individuals interact with one another and regularly making political, business, social, and religious decisions. The individuals choose with whom to associate, to whom to look for advice, and for whom to advocate. That these features are constants in the lives of residents of small-scale and developing societies comes as no surprise. However, even in large-scale, digitally-connected, and developed societies individuals have close social contact with only a limited number of social alters (Gonçalves, Perra and Vespignani, 2011). Perhaps, San Francisco's bustling neighborhoods change the cultural dialogue by bringing diverse groups of individuals into contact with one another; but, while these groups may mix and develop new connections among members, remarkably the number of social alters

¹The Paradox of Weak Ties in 55 Countries.

4.1.1 Honduras

I obtained consent and followed data collection practices according to the guidelines of the Human Subjects review board at the University of California, San Diego.

Honduras is a relatively poor nation by comparative Latin American standards, and La Union is comparatively poor within Honduras. Greater than 60% of the residents of La Union are in the lowest quintile of Honduran Income distribution. The residents of La Union have less access to potable water, more residents without access to latrines, lower television penetration, and fewer residents who continue to secondary education. Despite these characteristics—or perhaps as a result of them—the region is home to a vibrant set of social relationships. Both Christian and traditional holidays are celebrated, farmers are organized into collections of cooperatives and aggregators, school-aged children participate in local club sports. Religious life, while for many in the area a core part of social events, is not a monolithic social force. Indeed, in most every village both Catholic and Evangelical adherents participate together in daily life, and a non-trivial number of residents report that they do not participate in any religious services. Therefore, in many ways, at a high-level the residents of this region might be typical of those in other parts of Honduras, Latin America, and rural residents broadly.

The sample for the study was drawn from a census-penetration sampling frame in the thirty-two villages that surround La Union. The set of villages selected to be a part of the sampling frame represent a choice of thirty-two drawn from more than sixty in the area. The choice of village was non-random, as villages that were more readily accessible from the county seat were chosen. These villages range in population size from as few as twenty to as large as six-hundred fifty. For example, on village, *Los Perdomos*, literally “The Perdomos”, is a town in which every resident is a part of an the organized Perdomo family group. In contrast with Los Perdomos, San Bartolo is a village that houses a large coffee production center and is home to more than five hundred residents. Across the entire set of villages included in the study, nearly ninety percent of the residents of the village were surveyed.

Data collection in the towns surrounding La Union proceeded in the following four steps. First, the researcher and research assistants met with residents of the village and together canvassed the village to generate a map of the physical locations of all buildings in the village. Although there was little ambiguity about whether a particular building was a residence or a storage unit, it was somewhat

more difficult to ascertain whether a residence was presently occupied or of the residents of the building has moved. The more difficult case to identify as a residence was when a building was being constructed and was partially completed. In the case that the research assistants thought the building was potentially inhabited, the research assistants marked the location on a map and scheduled it for a return visit. The reasoning for this enumeration follows that of the US census and other population representative samples: everyone living in a town needs either a permanent or temporary residence, and so by enumerating the buildings, all individuals could be located. In addition, in each village we spoke with residents in the town to query about whether there existed any residents of the town who lived without a permanent structure.

Second, after generating a map of the buildings in each town, the research team returned to each structure to speak with the individuals who lived at the structure or on the property. In this stage of the enumeration, we spoke with an adult living on the property and asked the adult to list, by name, the individuals who lived at the building, or on the property. Because there is a concern that subjects with limited recall and large families might not accurately recall each individual living on the property, we employed an age mnemonic to aid respondents; we asked respondents to start with the eldest individual and work toward the youngest individual. For each listed individual, we also queried whether that individual had a partner who lived in the same location. In this manner, by completing the same procedure at every building in town, the research team was able to create a list of every person living in town.

Third, and concurrently with the enumeration of the individuals living at each residence, we snapped a photograph of each individual to be keyed onto that respondent's database record. To take this photograph and facilitate the inclusion on the database, we asked each resident to hold a small white board on which we noted the "address" of the building and a record for the individual.² The reasoning behind taking this photograph merits some further description.

A key component of this data collection, and one that is described in detail later in this chapter, is the ability to measure high-fidelity social network relationships in our dataset. In preliminary interviews in the area, the research team identified two features which were problematic for the successful completion of

²The "address" was little more than the code we generated for the residence on our preliminary, map-making stage of the data collection.

this task. First, because of naming conventions in the area, there exist a relatively small number of given names and family names. Second, and relatedly, individuals living in the area developed a series of "called-names" to circumvent this problem. The difficulty with this called-name, or nickname convention is that nicknames themselves were neither unique to a single individual, nor was a single nickname always sufficient to identify someone. That is, two individuals in a single town might be called "Chuck", while at the same time, one individual might be known as "Chuck" with one group of friends while being known as "Charlie" to a separate group of friends. This convention is problematic for a researcher with the aim of measuring ties between individuals. By taking a photograph and pairing this with a high-recall search, the research team was able to quickly, and effectively link town-residents with database records.

The fourth, and final step in the primary data collection was returning to administer the survey to each resident enumerated on the population list. This process was a multi-day process, and involved survey enumerators spending considerable time in the town. Because the ambition of this data collection was to elicit social network information from respondents, very high response rates were required. As such, rather than setting a stopping condition beyond which a targeted respondent is considered unreachable, as is commonly used in population representative sampling, in this data collection the enumerators continued to work in a town until they reached ninety percent completion. As a result, those individual who were enumerated in the initial canvassing of the town, but are not included in the primary data collection, are a non-random set of town members.

While this data collection task was not able to gather direct evidence about the individuals who were not surveyed, semi-structured interviews with these missing individuals' social contact suggest that typically these individuals were more likely to be male adults working at locations far from home. Importantly, it was sometimes the case that when we returned to conduct primary data collection, that we would ask when *Potential Respondent A* might return home for the enumerators to speak with them that the residents of the house would respond that the individual was living in the capital city and would not return for several months. In the event that an individual had not stayed at the home she was identified as living in for the month prior, and was not anticipated to return in the coming month, we removed this individual from the population list. If the individual was to be present in the village within a month prior or post the data collection, the individual was

left in the population list. This decision is consequential, as individuals who are identified as non-residents from the town are unable to be nominated as a social connection by other residents of the town.

4.1.2 Ghana

I obtained consent from all participants according to the guidelines of the Human Subjects review board at the University of California, San Diego.

This work took place in the coastal fishing community Ehonle Nwianzo. Ehonle is home to fifty-one children and scores of children, and is located on a thin strip of beach between the Atlantic Ocean and a brackish lagoon. It is surrounded by two towns, one to the east and another to the west. One half-mile to the west is the coastal, agricultural community of Princess Town. In the recent past, Princess town was a bustling fishing village that was home to nearly three thousand residents, but recently residents have transitioned away from their fishing heritage and there has been significant out-migration. Currently, two thousand people live in Princess Town; most are farmers. One quarter mile to the east of Ehonle, is the town of Akatakya (pronounced *Ak-ahh-tetch-ee*). Akatakya is a fishing village of about six hundred individuals and is the port for twenty large fishing boats, each crewed by a team of between ten and fifteen men. These outboard driven vessels typically motor thirty minutes along the coastline to productive fishing locations where the fishermen use a combination of netting and line fishing practices.

In contrast with Akatakya and Princess Town, Ehonle has neither access to calm mooring waters nor productive farming land. Instead, fishing in Ehonle is primarily beach-seine netting with supplemental cast-net fishing in the off seasons. In the Ewe beach-seining, a canoe of seven men rows three hundred meters to sea and drops a net that floats on foam-bobs on the top with a low side is weighted to sink. This presents a vertical net face to gather the harvest. At either end of net, two ropes run the distance back to shore where they are maintained by members of the village. Once the net is sorted, the members of the town haul in the net by pulling the ropes the three hundred meters to shore. Once ashore, the haul is separated from the jellyfish by-catch and fish are distributed equally among each of the family units in town. Additional catch beyond primary subsistence is sold to female merchants from nearby Princess Town. Like the catch, receipts from the sale of fish are distributed equally among all the residents of the Ehonle. All

told, the process of setting, hauling, and divvyng takes around three hours and is undertaken twice daily, first at eight in the morning and again at two in the afternoon.

All of the towns in the region are hereditary-chief centered communities. As such, the chief owns the land upon which the community members live. In addition to a village chief, in most towns there is a chief-fisherman who is the arbiter of maritime conflicts. The chief-fisherman has no formal authority granted to him³ by the chief or the villagers. The decisions of the chief-fisherman are followed because the lack of following these edicts would preclude one from bringing suit against another fisherman in the future – if one does not follow the rules, he then has no recourse to sue another who fails to follow the rules.⁴

4.2 Social Network Measurement Considerations

While researchers have become increasingly aware of the cognitive and psychological components of how survey respondents answer survey questions (e.g. anchoring biases, surveyor demand effects, respondent fatigue), little systematic research has assessed how a survey respondent might be expected to respond to questions about social network connections.

The workhorse survey instrument for the elicitation of social network data is the *Name Generator* survey question. Name generators often take the following form, “Consider {some context}. Are there any individuals in this context whom you typically talk to?” Leading the way in the development of this methodology have been Bob Huckfeldt and John Sprague, who pioneered this with the 1984 South Bend Study (Huckfeldt, 1979b). The scholars followed with the 1996 Indianapolis -

³They are without exception male.

⁴As an aside, this system presents an interesting set of incentives with regard to conservation. The chief-fisherman has meaningful authority in matters that pit one fisherman against another – for example the destruction of one’s fishing gear by another fisherman. However, the chief-fisherman’s sanctioning authority lacks the power to shape outcomes that all the fishermen agree upon. If the chief-fisherman wanted to change fishing behavior from an environmentally-degrading practice toward some other less degrading practice, and if the fishermen came to a consensus that they were not interested in this change, then the chief-fisherman would have no recourse. In this way, the legitimacy of the chief-fisherman is fragile. Indeed, even a finding of fault against a popular fisherman has the potential to undermine his authority. And, advocating for unwanted changes would almost certainly undermine his authority. The frequent refrain in structured interviews with these chief-fishermen was that they, and the fishermen, were aware that their practices were damaging the long-term health of their stock, but that none of the stakeholders had the combination of sanctioning authority and conservation-will to bring about environmentally sustainable changes.

St. Louis Study (Beck et al., 2002; Huckfeldt, Johnson and Sprague, 2004). Indeed, the “political discussant” line of inquiry in American politics is based on name generator questions of this form.

In the South Bend study, political discussants were identified using the following name generator:

“We are interested in the sort of political information and opinions people get from each other.

Can you give me the FIRST names of the three people you talked with most about the events of the past election year? These people might be from your family, from work, from the neighborhood, from church, from some other organization you belong to, or they might be from somewhere else.”

Name generators of this type have been included on the ANES in multiple waves of surveying, beginning with a pilot implementation in 1998. The 1998 ANES version read:

“From time to time, people discuss government, elections, and politics with other people. I’d like to know the people you talk with about these matters. Apart from the people in your immediate household, can you think of anyone?”

If the response was yes, the interviewer solicited the first name of the discussant, and followed up with,

“Is there anyone else you can think of?”

for up to three discussants for each respondent outside the household. In 2000, Bob Huckfeldt and Ronald Lake published a retrospective evaluation of the reliability of this form of instrumentation, concluding that the instrument was reliably measuring political discussants, and moreover, the discussants identified by the measure were generally better informed than the initial, seed respondent. This form of snowball, or *ego-centric* sampling responses in a social network can provide powerful additional data, especially about social support and some types of other information. However, because it includes only one or two political discussants, this type of data cannot provide structural information about the network. That is, it cannot provide information about respondents’ friends’ friends; cannot provide information about groups of individual that tend to associate with one another; cannot provide information about whether a respondent is core to the social group or peripheral.

To generate this type of network data, as opposed to just discussant data, a broader conceptualization of the data—a conceptualization like that taken in this work—collection need be taken. In the following sections, I outline three distinct practical issues which might arise in the creation of social network data: (1) Failure on the part of the respondent to identify all relevant social ties; (2) inability of the alter to uniquely identify social alters; and, (3) data entry errors on the part of the enumerator. This typology of errors follows from the work of Wang et al. (2012), who conduct simulation research on the robustness of calculated network properties in the face of measurement errors. The Wang et al. (2012) error typology is included in Table 4.1, with the addition of the error type identified for the following sections.

These issues do not exist in most computer mediated social interaction; the handling of identifying individuals, persona, or avatar has necessarily been accomplished at the point a human enters the mediated system. However, for the researcher whose goal is identifying real-world interactions, these considerations become salient.

Table 4.1: Wang (2012) Measurement Errors. The location of the error, the type of error, and a real world example of the error are presented.

Error Level	Type	Example
Node	False Positive	Boundary Specification
Node	False Negative	Boundary Specification
Node	False Aggregation	Non-Unique Identification
Node	False Disaggregation	Nickname Problem
Edge	False Positive	Data Entry Errors
Edge	False Negative	Forgetting Social Ties

4.2.1 Failure to Identify Population

Reliably identifying the relevant social population is, in large part, the responsibility of the researcher prior to beginning the data collection task. In some cases the referent social group may be easy to identify—the set of individuals in a Karate Club, the body of undergraduate students at a university, or the group of Democratic Party superdelegates—but more frequently, group membership is more fluid. Indeed, the neighborhood participation task of Huckfeldt (1979a). If the goal of this work were to measure the entire neighborhood social network, the

researcher must make an *ex ante* specification of who does and does not belong to the neighborhood group. In some locale this might be easy, but if the neighborhoods in question were in New York or San Francisco, two cities with a shared affinity for naming (and re-naming) city blocks, then the determination in edge cases becomes considerably more difficult. The difficulty identifying population becomes increasingly difficult as egos live near the boundaries; someone living in one of San Francisco's *Painted Ladies* houses sits on the border between the neighborhoods of Alamo Square and Hayes valley. For this person, any neighborhood parsing by the researcher may cleave social contacts, and lead to inaccurate measurements of social network alters.

4.2.2 Forgetting Social Ties

Subjects cognition operates under a state of limited cognition and recall (Simon, 1955, 1956). This much is well known, and considerable meta-research, survey design, and instrument-design research has investigated methods of improving subject's accuracy in performing survey tasks. Validation of performance has typically be through test-retest validation whereby subjects are asked to perform some survey task (denote this task "Task A"), proceed through other survey tasks ("Tasks B – k"), and then return either to Task A or a simliar task ("Task A"). Ideally, test-retest reliability would be high on stable attributes.

Empirically, this is not the case. Subjects asked to perform even very simple tasks – recalling the year their parents were born, the state capital of New Hampshire – perform with substantial disaccord in repeated trials. What is more, as subject cognitive capacity decreases or the limitations on the ability to fully process prompts increase, subjects performance further drops. These are formidable challenges when ascertaining the social connections of actors in poorly-educated, high constraint environments.

It is well demonstrated that open-ended questions demonstrate the least reliability in test-retest validation. Even fuzzy matching, de-duping, and other best-practices to standardize text responses are left with low reliability. This is, it isn't simply that responses that ought be scored as identical are not being scored as such; rather it is the case that subjects are, in fact, giving non-identical responses. Furthermore, the broader the range of the response set, the lower the test-retest reliability. For example, test-retest when asking a subject to type her single favorite

color might be high; increasing the response set to measure the subjects' five favorite colors will be significantly lower. Then, the task of asking a subject to recall an exhaustive list of social connections is Herculean task in light of the open-ended, high-dimensional response space.

4.2.3 Non-uniquely Identifying Alters

Here, I expand this concept to describe the concept of "Researcher Uniqueness" which is the requirement that for social network measurement, there is a *known* one-to-one correspondence between a bit of empirical, or real-world data into the database entry for that data. Researcher uniqueness is a problem unique to social network studies where the population of respondents need to be able to correctly identify other database records corresponding to real-world individuals. In contrast, typical cross-section survey work undertaken in the social sciences does not require that any particular row of the data hold contextual knowledge of any other row in the data. Even in repeated panel studies, typically the researcher holds a contact file that contains the contact information which is linked with a survey key on the database record of the data. This form of data storage is inadequate in social network surveys because a reliable, certain (aka non-probabilistic) mapping from one database row to another is required. In this way, both survey responses and subject meta-data need to be linked, in place, in the survey architecture.

When working in developing democracies, it is not uncommon for a researcher to be limited in how she can uniquely identify study respondents. In particular, if a researcher were establishing a study in the US she might use respondents' email addresses, cell phone numbers, street addresses, date of birth, or some combination of these features to reliably identify subjects. For example, Hobbs, Christakis and Fowler (2014) are able to uniquely match 98.3% using only *first name*, *last name*, and *date of birth* as identifying variables. Typically, very few of these features are available in developing democracies. As such, even in relatively small populations, it can be quite difficult for a researcher to make a dispositive database record that links a person with her data.

One example of recent work that has faced this problem is that of Cruz, Labonne and Querubin (forthcoming) who examine vote buying strategies in the Philippines. The authors leverage a remarkable artifact of the Spanish colonization of the islands – family names were assigned systematically to large portions of

the Philippine population. As a result, even in present day contexts there may be only a limited number of family names in a particular village. Compounding this problem is that there is a relatively short list of first names chosen by parents when they name their children. Consequently, within a small village, there may be several males who share the same first and family name. To be certain, upon further inquiry, the researcher might be able to leverage middle names, or mothers' names to uniquely identify a respondent for the purposes of researcher uniqueness in the database.

Using a specific bit of data to uniquely identify an individual for the purposes can guarantee researcher uniqueness, but often times that unique data is not commonly known among that individuals' social contacts. In the Philippine case, the combination of first, mothers' and fathers' last names, together with DOB would almost certainly create a unique hash of an individual in a village. However, when the elicitation task requires that another actors living in the village is able to use that information to identify an alter, this information is useless if the alter does not also know that information.

One seemingly sensible option might be to use the alternative name, or "nick-name" conventions. In areas where names are relatively similar, common practice is to create alternative names to provide to people. These alternative naming conventions are, in fact, developed to overcome exactly the problem that the research is also addressing! However, in practice, relying on this alternative naming convention can be problematic because many of the alternative names are context specific. That is, depending on the part of the community that the researcher is addressing a particular individual might have distinct, unrelated names.

The difficulty with this alternative naming conventions bites down when the alternative names are not themselves unique, and also when there is limited recall on the part of subjects for all the names they are called. Subjects under limited attention and recall are unlikely to accurately reproduce every alternative name they are called; this is the same problem as was introduced in the "limited cognition section." As such, the researcher loses the unique identifier she wanted to use between data records.

To solve this uniqueness problem in this work, the team took photographs that were keyed to a database entry. In this way, not only could researchers be certain that they were associating survey answers with a particular database entry, but the researchers could also use the same mechanism to draw connections between

real-world survey respondents and those survey respondents' corresponding data-entries in the database record.⁵

4.2.4 Data Entry Errors

A reality for many researchers under-taking field work in developing countries is the process of building, fielding, and eventually scoring and entering a mountain of paper surveys. In one salient example, a colleague surveyed 300 members of the Ewe tribe in the Volta Region of Ghana. This survey, impeccably constructed with built in randomization and shielded responses, after fielded occupied her *entire* graduate student office in the Social Sciences Building on the campus of University of California, San Diego. While this was a problem outright, the greater difficulty came in transporting her data, in paper format, back from Ghana so that it *could* occupy her office. The process of shipping, tracking, and confirming that her dissertation data was complete, not lost, and not subject to ransom at a checkpoint in Accra was a monumental task that many researcher in democratizing locale understand and appreciate.

But, it need not be this way. The data that we possess as field researchers is not so data intense that it need occupy an ENIAC computer rooms' worth of space. Indeed, the entire dataset for this dissertation, when compressed can be fit onto a 3.5" floppy drive; the style of data storage from 1992. The difficulty comes in the translation from the questionnaire into a completed survey, before the coding into a digital space occurs.

The process of this encoding is, itself, fraught with problems. Inclement weather poses a particularly ironic challenge. If a researcher is working in a developing or democratizing area—precisely the areas that paper surveys are the mode *de rigueur*—the probability that he or she is exposed to inclement weather potentially spoiling those surveys is higher than if the researcher were working in a developed area. Minimally, in well-developed areas, in the case of inclement weather, the ability of a researcher to find shelter in a building that is leak-free is greater than

⁵The link between survey responses and identifying record data presents somewhat of an ethical challenge. Practically, in the survey utilized in this dissertation the enumerators did not ask any questions that might be considered even remotely sensitive; but more generally, linking identities and survey responses in place in a computer database should provide the researcher occasion to pause and carefully consider the security protocol in place. Indeed, Netriks has been used locations and among populations where there was considerable damage that could be caused if the data were to be held by an non-research party.

in democratizing areas. Why the concern with inclement weather? Ink bleeds on surveys when exposed to moisture.

This author has experienced, first hand in Mexico, being caught in a down-pour and watching the responses from a paper survey being administered by an enumerator bleed to the point of illegibility. However, conditional on the spoiling of the data, this was actually an optimal circumstance because the enumerator could be encouraged to begin the survey anew. A worse occurrence is when the survey is spoiled but the researcher does not realize the spoilage until he or she returns back to her research institution and unboxes the precious cargo of data. At this point, on several projects that I have worked it has been common for the researcher to have item non-response on as many as five percent of the *completed* survey items due to inappropriate data marking or the results of corruption of the data from the paper surveys.

The transcription of surveys from paper copies to digital data is a repetitive, monotonous task. Typically, the thing furthest from the researchers' goals upon returning to his or her own home and own bed after a week, or month or year in the field is, "Hurrah. Now I can sit down and code these thousands of pages of documents. Instead, more frequently her reaction is, first let me take a short reprieve from the task of this work; and second, let me get to the data as quickly as possible. A further reality of the field data collection process is that there tend to be limited budgets that remain at the conclusion of the field research endeavor. Together, the combination of monotonous task, little remaining zeal to perform the data entry herself, and little budget are a perfect storm which lead to the data crashing against the rocks of undergraduate RA help.

Undergraduate research assistants are both inexpensive, and at least initially, eager to help a bushy-eyebrowed professor with their research. After all, the recommendation of a faculty member or a graduate student is the only piece of the puzzle missing from their admission packet to an Ivy League law school. This zeal is, of course, waning. When the first round of assigned readings come in, or the first Greek-life event, or the first "Bio" midterm, or really anything, come down the gauntlet toward undergraduates, our data is the first thing to drop off their plate.

Certainly, errors in the data transcription at this point can be reasonably conceptualized as purely random noise pumped into the data. In the case that this noise is associated with the outcome variable, any estimates will be unbiased but measured with higher variance. If the noise is associated with a righthand-side

variable, it will not lead to bias in a particular covariate of interest, it will decrease the precision with which that covariate is measured, which will in turn lead to attenuation bias in the estimation of relationships between covariates and outcome variables.

4.3 Previous Solutions

This project is not the first to acknowledge the difficulties in measuring the social networks of actors. Indeed, (e.g. Brewer and Webster, 2000; Brewer, 2000) devised an early solution to this problem whereby he prompted respondents with a name generator, for example "Who are the people in this town that you attend religious services with?" and then before accepting nominations from the subject, provided the subject with a previously gathered population list. This method solves the problem of failing to identify the entire population, and somewhat reduces the cognitive load on respondents. Consequently, it considerably reduces errors of forgetting social ties and boundary problems.

Despite the advances that Brewer's method makes, it still leaves several problems unaddressed. Specifically, the Brewer method does not address uniquely identifying social alters, data entry errors, and in the way it is implemented may lead to considerable false-positive identification of social alters. Under the population list elicitation mechanism, the threat to unique social alters is still present; a long list of similar names, while it is likely to reduce recall errors, does not solve the inability of respondents to parse between two or more highly similar names. Research by Corin Apicella, who studies the Hadza of Tanzania⁶ modified the Brewer method to photograph the population. The population list was then presented to survey respondents at the time of elicitation and the secondary uniqueness method was utilized to ensure accurate nominations.

The Brewer list method, especially when paired with the additional uniqueness check of cross-referencing the nomination with a photograph of the alter, reduces issues of cognitive load on survey respondents but it introduces the possibility for additional errors. Surveyor demand and social desirability concerns are likely to lead subjects to overreport their social connections. Specifically, there is a (weak) norm toward being a part of a community and this norm lead subjects

⁶The Hadza are one of the few remaining hunter-gatherer tribes in the world. This nomadic tribe's range covers a broad swath of Tanzania.

to nominate more social alters than they hold in actuality. As a consequence, the measured social networks, in overcoming a concern of underreporting, instead are likely to lead to over-reporting compared to the unmeasured truth. It should be well noted, however, that the absence of ground-truth measurement means that this claim cannot be empirically adjudicated.

An additional concern about the Brewer list method, also driven by social desirability bias leads to the over-reporting of connections with prominent, socially-well connected opinion leaders in the community. Similar to the previous concern that social desirability will lead to subjects over-reporting connections with social alters with whom they do not, in fact, have a social connection; even a weak norm to want to know those who are well connected in town will lead to an over-reporting of ties to these actors. That is, people who are well connected in town are likely to receive more friendship nominations than they would have under a measurement strategy that would instead accurately measure the social network. Consequently, comparisons of the node-level network properties of well-connected against poorly connected individuals within a social network will be relatively overstated, potentially leading researchers using empirically measured social networks to underestimate a bivariate relationship that uses any measure of social connectedness as an explanatory righthand side variable.

4.4 Consequences for the Social Network

The problems identified in the previous section can be immediately mapped onto issues present in the graph-level data utilized in the eventual network analysis. The errors most likely to occur under unscrupulous data collection in this case are: (1) False Negative Edges; (2) False Disaggregation of Nodes; and (3) Poor performance (precision and recall) on node identification.

4.4.1 False Negative Edges

The limited search capacity of respondents will lead to a reporting of the total number of social relationships that is uniformly biased downwards from the true value. Failing to uniquely identify alters will lead to random "rewiring" of edges between nodes. Data entry errors will further lead to rewiring of edges between nodes. And finally, failing to properly identify the population will lead to both

boundary specification problems and missing node problems. I address each of these concerns in turn in the following section. Each error in the reporting of these social network components can be thought of having both a local and a global effect, which I address for each concern.

The limited cognition of subjects will lead to downward bias in the "outdegree" network for individuals. This mechanism could occur in at least two ways. Subjects may fail to remember a close social relationship, for example a sibling; or, subjects may fail to remember more distance social relationships. At the local level, either of these mechanisms will be associated with a downwardly biased estimate of the social connections for each individual; individuals will have fewer outgoing social connections reported internal to the survey instrument than they actually hold in reality. This might lead to inappropriate estimates of social support or social engagement. Likely more damaging, however, is the problem that these decreased outgoing connections have on the global properties of the network.

One of the important properties of social networks, indeed, the property that is at the heart of the "friendship lever" (Jason Jones), is the vast trove of information that is stored in weak social connections. While strong connections are the most important for daily activity and social support, it is weak connections that allow for some of the remarkable properties of the network small-world effect (e.g. "Six Degrees of Kevin Bacon", and Duncan Watts' mailings). When subjects are unable to locate these relatively weak social connections, it would be impossible to demonstrate or understand these small world and friendship lever phenomena.

4.4.2 False Disaggregation of Nodes

4.5 Measurement in this Work

To overcome the limitations identified in the previous sections, for this work I developed new software with the help of my computer-science roommate from my undergraduate education. This software, which we called Netriks and which we designed to be implemented on open-source Linux distributions, allows for the precise measurement of social networks and other survey items in areas where there is very limited or no internet connectivity.

Netriks is a bare-bones implementation of a standard Linux web-server stack

commonly referred to as a LAMP server.⁷ In this section, I very briefly describe the technical setup, though I leave any of the technical details to the documentation materials for the program. The web-server nature of Netriks means that it is relatively platform- and architecture-independent. This independence does come at the significant cost that the setup of the software requires measure of technical proficiency. Netriks uses the commonly implemented Apache web-server⁸ which is bundled with virtually all modern operating system distributions and is nearly twenty years old. The Apache documentation claims that the server is currently the most popular web-server. MySQL is the worlds most popular SQL database engine. It can be rapidly deployed inside virtually any *X distribution, and provides the data-base backend handling for the Netriks LAMP stack. Finally, php⁹ which parses user defined code into web-ready *html* script, serves up enumerators the end-user survey content.

Because Netriks is a web-hosted survey tool, if a researcher has access to secure internet connections, Netriks can be deployed without any additional setup on enumerator supplied hardware. To implement this over existing internet connections, the researcher simply implements the server side setup their own machine, and authenticates into other machines in a standard web-interface. In this case, enumerators can simply be provided login credentials in a way that any sysadmin or network admin might be able to construct, and the enumerator can begin the survey process. The primary benefit of this method is that the enumerator retains ownership, both of data and hardware, in a controlled system. Especially in areas where sensitive information is being collected and there is some expectation that non-researcher entities may be interested in accessing this information, the end-to-end ownership is an important benefit.

4.5.1 Serving Internet in Very Rural Locations

With little additional configuration the LAMP server can be install on each client device (utilized by the enumerator) and so can be deployed in locations where there is no internet connectivity. Pragmatically, the most straightforward implementation for this is on low-cost chromebooks and netbooks. In particular, the operating systems that are shipped with mobile devices are somewhat more

⁷The LAMP acronym stands for Linux-Apache-MySQL-php.

⁸<http://httpd.apache.org/>

⁹PHP is a recursive acronym for: "*PHP: Hypertext Preprocessor*".

difficult to establish as a working implementation of Netriks (if locally hosting the internet is required).¹⁰ However, local hosting of a web-server on each machine creates the circumstance where at the end of each work day machines hold a unique database record from any other machine. In essence, even if machines started with a current database record at the beginning of the day, at the end of the day new surveys completed by a survey enumerator are unique records that are not stored in the database of record.

To address this issue, Netriks was written to sync these out of database entries for inclusion into the database of record. So, at the end of a research endeavor, every enumerators' machine can be returned to the central stack, the MySQL records dumped¹¹ onto the central record, and then this version-current database can be pushed back onto each enumerator machine. In this way, the survey team begins each survey task with a current record, moves out of sync through the day, and then re-syncs at the end of each day.

4.5.2 Dynamic Lists and Photographs

The key feature of Netriks combines features of both the photographic identification used by Apicella in the Hadza with Brewers' population-list elicitation. However, given the dynamic recall that is possible with the computer-facilitated recall, the Netriks method is less susceptible to norm-driven over reporting of social connections. In particular, Netriks allows the enumerator to prompt the subject with a name-generator and allows for free-response nomination by the subject. However, the free-response is immediately and dynamically matched against a population list keyed with photographs of each individual.

4.6 Conclusion

In this chapter I have described the data collection procedures used to collect data in Honduras and Ghana. In addition, I have briefly describing the technical procedures of this data collection at each site. Accurately detailing and recording this information in this chapter makes it possible for future researchers collecting

¹⁰This low-power laptop is how Netriks was deployed in gathering the data for this dissertation, and how it has also been deployed in Oaxaca, Mexico; Ghana; and Tanzania.

¹¹This data transfer is designed to be possibly either wirelessly across an ad-hoc network at the researchers' base of research, or via a USB drive.

social network data to understand the identification, enumeration, and ultimately social network generation techniques that have produced this data. What is more, this chapter argues that the data collected and utilized in the remainder of this dissertation work is among the highest quality data of face-to-face social networks collected to date.

Chapter 5

Political Mobilization Through Social Paths

5.1 Introduction

Why do some citizens choose to participate in political and civic activities while others do not? Most previous explanations have focused either on individual characteristics like education and past voting history, or on large-scale, societal characteristics like civic-mindedness. In this chapter, I test a prediction from the theory of *social information*: political mobilization occurs through actors' social networks. Even more, because of the costly nature of political mobilization – the individual who is mobilized to take action is paying concentrated costs for what are likely to be diffuse benefits – a theory of social information predicts that mobilization is most likely to occur in local networks.

I make the following contributions in this chapter. First, I build on recent work concerning drawing causal inference in the presence of spillover (Bowers, Fredrickson and Panagopoulos, 2013; Sinclair et al., 2014; Baird et al., 2014; Hudgens and Halloran, 2008). Whereas previous scholars have largely developed methods to cope with spillover as a nuisance parameter by estimating average differences between clusters not subject to spillover (e.g. Sinclair et al., 2014; Baird et al., 2014), I demonstrate that with spillover-pathway data (e.g. social networks), it is possible to recover causal estimates, at least in some cases. Second, I argue that previous two-stage randomization procedures advocated for inference in the presence of spillover are not a *necessary* precondition for inference, especially if within-cluster spillover

can be evaluated. Third, and finally I make several substantive contributions to the understanding of political mobilization drawn from a unique dataset and intervention fielded in Honduras. Using between-village variation in the connectedness of randomly-assigned political mobilizers, I identify that targeting well-connected mobilizers causes a marked increase in town-level political activity. Then, I examine the two plausible mechanisms leading to this effect: either well-connected mobilizers hold sway across the entire network, or well-connected mobilizers hold sway in their (relatively larger) local networks. On the whole, I find that the evidence supports the second of these propositions.

5.2 Existing Explanations

Existing explanations for political mobilization behavior fall into four classes of explanations. The most common explanations in the current literature are individual characteristics, typified by the work of Rosenstone and Hansen (1993) and Brady, Verba and Schlozman (1995). The other four explanations that I identify are supra-individual explanations; these explanations, while recognized in the current literature, play a relatively minor roll compared to the individual characteristics explanations. I describe, in detail, these explanations in subsequent subsections, but briefly they are society-wide characteristics, social group characteristics, and social network explanations.

Perhaps the best, most succinct characterization of these classes of explanations is due to Brady, Verba and Schlozman (1995), who opine:

“Why do citizens participate in political life? One way to think about this puzzle is to invert the question and ask why people don’t take part in politics. Three answers immediately suggest themselves: *Because they can’t, because they don’t want to, or because nobody asked,*” (Brady, Verba and Schlozman, 1995, p. 271).

The features identified in this subsection comport with the categories of motivation identified by Brady, Verba and Schlozman. However, the mapping is not one-to-one. Specifically, individual characteristics may affect both capacity and willingness; in the same way, supra-individual characteristics may affect any of Brady, Verba and Schlozman’s categories.

5.2.1 Individual Characteristics

Existing, large-sample observational studies have identified a range of characteristics that are predictive of individual political behavior. These features can be grouped into the broad classes of time, money, and civic skills (Brady, Verba and Schlozman, 1995). While consensus seems to have emerged around these predictive features, for quite some time, this participatory question was at the heart of both American and Comparative behavioral political science.

5.2.2 Society Wide Characteristics

5.2.3 Previous Explanations Identify Social Groups

Participation and mobilization are distinct concepts. While participation is a stable, long-term, equilibrium outcome that has typically been explained using equally long-term independent variables. Accordingly, many scholars explain variance in political participation with characteristic variables like race, level of education, feelings of personal efficacy, and years in residing in community which are unlikely to shift over small periods of time and as such are fixed in the short-run (Rosenstone and Hansen, 1993). Brady, Verba and Schlozman (1995) expand the range of explanatory variables, but maintain a focus on long-term variables of civic skills and resources. For instance, an individual may accrue resources over time, and those may precipitously increase or decrease from exogenous shocks, but typically there is little movement in these *characteristics* between any two consecutive elections.

Conversely, mobilization is an attempt to push an individual's action off of the equilibrium determined by the aforementioned fixed participation variables; mobilization is an external push off an equilibrium. The "calculus of voting" model Riker and Ordeshook (1968) suggests that voters undertake a simple cost benefit analysis: if the probability weighted benefit (pB) of voting outweighs the structural and cognitive costs (C) of casting a vote, the voter turns out and casts a ballot (Downs, 1957; Riker and Ordeshook, 1968; Gerber, Green and Larimer, 2008). Although Riker's model is useful for analyzing equilibrium comparative statics, the model is not trying to accurately predict any particular voter's decision. Downs (1957) and then Riker and Ordeshook (1968) attach an unmeasured valence (D). The term functions identically even though the scholars have different name for the

valence: Downs (1957) believes this term is a voters' preference for the long-run function of democracy, while Riker and Ordeshook (1968). The calculus of voting is represented as

$$pB + D > C \rightarrow VOTE;$$

and thus, attempts to mobilize are an attempt to alter one of the short-term parameters of the calculus.

Mobilization is distinct from exogenous shocks to participation variables in that mobilization requires strategic action taken by an external third party. For example, if an actor wins the lottery, she may be more likely to participate because of the increase in resources that stem from the lottery. In this case winning the lottery would of course be an exogenous shock. If however there is a change in electoral registration laws which cause short-term changes in the participation of voters, then the legal changes could be exogenous socks but it would depend on whether the officials who changed the laws did so with the intent to influence participation rates. For mobilization to occur, the citizen must be intentionally moved off their equilibrium behavior and as such, after a mobilization effort ceases, the likelihood an individual participates should return to the pre-mobilization likelihood unless having participated increases one's likelihood of further participation. In which case, a new equilibrium would be established. Finally, mobilizers are often opinion leaders who increase the salience of a political issue with the intent to alter the likelihood of participating for a particular group (Bartels, 2006, 2008).¹

5.2.4 Explanations for Participation

In the canonical argument for the effects of social economic status on participation, Rosenstone and Hansen find that income, education, unemployment, internal and external political efficacy, Party Identification, church attendance, and mass-elite strategic mobilization form a model that correctly predicts the decision to turn-out to to vote in 75% of cases (Rosenstone and Hansen, 1993, p. 275). Brady, Verba and Schlozman (1995) updates Rosenstone and Hanson's argument to include a broader conception of SES, and concludes that people do not participate for three reasons: they are not able to, do not want to, or are not asked.

¹See Also: Brooks and Manza 1997; Manza and Brooks 1999; Legee et al. 2002; Frank 2004; Shor, Bafumi, Park, and Cortina 2008.

Gerber, Green and Larimer (2008) and Gerber and Green (2000) investigate two central unexplored questions from Brady, Verba and Schlozman (1995); *if* voters are asked, does it matter *who* asks them and *how* they are asked? Although these are very broad questions, the authors contend that yes it matters who asks and how citizens are asked because different attempts to influence will activate different social norms. First, they find that the choice of media used to mobilize matters. Calls from a phone bank, mass mailings and house-calls by campaign activists do not have the same effect on participation (Gerber and Green, 2000). Second, the authors find social pressure applied from a neighbor has a stronger effect in changing the likelihood of an experimental subject than social pressure applied from an unknown researcher and even state “Exposing a person’s voting record to his or her neighbors turns out to be an order of magnitude more effective than conventional pieces of partisan or nonpartisan direct mail,” (Gerber, Green and Larimer, 2008, pp. 34).

5.2.5 Network Explanations

Among others, Fowler and Christakis have linked obesity, smoking, and cooperation to social networks (Christakis and Fowler, 2007, 2008; Fowler and Christakis, 2010). While this research, together with work by Sinclair (2012), Rolfe (2012), Siegel (2013), Klostad, Sokhey and McClurg (2013) has brought the question to the attention of political scientists in the modern epoch, the thinking is by no means unique to political science, nor this era. Earlier work has shown that medical innovation adoption (Marsden and Podolny, 1990), contraceptive choices (Valente et al., 1997), and adolescent smoking (Bearman, Jones and Udry, 2000; Alexander et al., 2001) can all be correlated with the behavior of peers. Social influence is also an important determinant in the decision to participate in politics. This is a non-controversial claim, but to date little empirical has studied the phenomenon (see e.g. Gerber, Green and Larimer (2008)). This paper demonstrates two distinct, but related, ways that social networks shape political participation: First, political information spreads through interpersonal relationships, and more importantly some agents in social networks are more influential than others because of their structural position within the network. Taken together, this research aims to fill a lacuna in the political mobilization literature noted by Brady, Verba and Schlozman: “*There are three reasons that individuals fail to participate in politics; they are unable to,*

they are uninterested, or nobody asked." This paper answers the final possibility; are some people better at asking than others?

To date, reliable inference for questions about social network influences have proven difficult. "Given the number of confounding factors and some of the data requirements, it may be prohibitively difficult to substantiate the role of social networks...through survey methods alone" (Valente, 2005). Indeed, due to the difficulty gathering whole-network data, most empirical research has either focused on using either egocentric data from sampled individuals – data that samples individuals at random from a population and queries about social alters, possibly following up with these alters in a snowball sampling fashion – or repurposed, found, sociocentric network data (e.g. Fowler and Christakis, 2010).

An additional problem that arises in the evaluation of social network explanations of behavior is that the nature of social networks, indeed, the very hypothesis being tested, is that behavior contains a component that is socially influenced. In the case of observational data, this social relationship leads to an exceedingly difficult problem to adjudicate: is the similarity in behavior between socially proximate individuals², or instead are observed correlations in behavior a consequence of the causal factor under study. Increasingly, scholars are advocating the use of experimental intervention on whole-network populations as a means of separating, and identifying specific causal quantities (e.g. Valente, 2005).

5.3 Social Information

How, specifically does a theory of social information fit into an explanation of political behavior? What are the mechanisms that might lead either an *information* mechanism or an *influence* mechanism to really take hold? This is a remarkably subtle question to adjudicate with data, since developing a test that makes differential predictions between the two among real-world networks challenging.

Consider, for example, the case that a scholar can field only a single experiment. Further, presume that the scholar has an interest in making a causal statement, have measured a large number of social connections, and has the resources to run an intervention. Furthermore, presume that the specific content of the outcome

²A phenomenon termed "homophily"; drawn from self-love, or selection into others who are like oneself.

measure and the intervention material are largely immaterial – the researcher is solely interested in designing a test capable of clearly identifying that one mechanism is responsible and not another. Even in this fanciful case, a finding of a causal effect that well connected individuals are more effective, on *some* dimension, would struggle to identify which of the mechanisms was at play.

Indeed, the best designed tests to distinguish between these two effects would run several trials with different costs on system "alters". What does this mean? In one trial, the researcher would imbue system egos with stimulus that is highly contagious; that is, a stimulus with relatively low-cost to spillover. If it is the case that the primary network effect in all circumstances is an *influence* effect, then even in the case of this highly-contagious stimulus that (a) highly connected individuals are the most effective, and (b) that the bulk of the network effect occurs in the highly connected individuals' local social networks. If it is the case that the primary network effect in all circumstances is an *information* effect, then even in the case of a low-contagious stimulus we would expect highly connected individuals to be the most effective mobilizers, but we would expect that the catchment of uptake would be more broadly represented in the across all the actors of the network.

5.3.1 Information

5.3.2 Influence

For Gerber, Green and Larimer (2008), when a neighbor is *asking*, it increases the pressure for conforming to social norms. When a stranger asks the subject to participate, the same social norms are present, but the pressure is nonexistent. What differentiates the neighbor from the stranger is that the neighbor has a shorter social distance to the subject. In the future, the subject will interact with the neighbor again, and that neighbor may communicate with other neighbors about the subject's willingness or refusal to conform. Whereas the subject will never see the stranger again. Therefore, the neighbor has a greater potential to increase the cost of defecting on and the rewards of conforming to social norms. If this social cost-benefit analysis is part of the decision to participate, then the next logical question is which neighbors are capable of exerting more social influence than others?

Social influence begins with the premise that people all people are seated in a social environment. This environment structures group pressures and the socialization of party identification (Berelson, Lazarsfeld and McPhee, 1954; Camp-

bell et al., 1960). However, distinct from characteristics, without activation social influence does not necessarily hold a direct influence on behavior (Dasgupta and Serageldin, 2001; Lin, 2001; Smith, 2005). Similarly to mobilization, social influence requires intentional activation to alter behavior. However, unlike mobilization efforts, the effect of social influence need not attenuate through repeated use. Indeed, social influence may manifest positive feedback loops that build toward increased effectiveness through repeated use (Milgram, 1974).

Moreover, the results from Gerber, Green and Larimer (2008) results probably underestimate the magnitude of social influence. The use of neighbor-ness is an intuitive proxy for social influence precisely because it differentiates the set of people with whom a person has iterated contact from the set of people with whom a person will never see again. While insightful, neighbor-ness captures a minute portion of the complex relationships that are embedded in the larger social environment. “The social in social cognition research is largely missing” (Kuklinski, Luskin and Bolland, 1991). The perceiver in this literature is a “passive onlooker, who... doesn’t *do* anything – doesn’t mix it up with the folks he’s watching, never tests his judgment in action or inaction. He just watches and judges,” (Neisser, 1980, pp. 603-604, emphasis in original). But, we are social beings, and theories of social cognition must, eventually take account of that fact (Krauss, 1981). Accordingly, this study intervenes within empirical communities and the experimental design allows precise operationalization and measurement of influence, and also allows for causal effects to be estimated. Although this study does not capture all of the complex relationships within the social environment, it should represent a significant step forward.

5.4 Methods

To examine the causal quantities, I build on the modeling framework most closely similar to Bowers, Fredrickson and Panagopoulos (2013). I also utilize the very useful quantities proposed by Baird et al. (2014), and the conceptual contributions of Sinclair (2012).

5.4.1 Statement of Causal Model

Following Bowers, Fredrickson and Panagopoulos (2013) consider a $n \times n$ adjacency matrix labeled \mathbf{S} that records network relationships between individuals. $\mathbf{S}_{i,j}$ contains a value of 1 in the i, j position if there is a social network connection between individuals i and j and 0 if there is no relationship between i and j . For undirected networks—networks where a connection from one individual to another implies a reciprocal connection—the adjacency matrix \mathbf{S} is symmetric across the diagonal. In this chapter, I propose to use only undirected networks.

Further, let treatment, when assigned at an individual level, be recorded in a vector \mathbf{Z} . Conventionally, before assignment, \mathbf{Z} is a random variable that can obtain one realization of the sample space Ω defined in \mathbf{Z} . The sample space, Ω has size k^n , where k is the number of treatment levels and n is the number of subjects possibly treated. After assignment, let \mathbf{z} be the particular realization, and $z_i = 1$ if subject i is assigned to treatment, and $z_i = 0$ if i is assigned to control. Let $\dot{\mathbf{z}}_i$ be the vector of treatment assignments for all individual not i ; for compactness in writing, whenever i is explicitly indexed in a quantity, let $\dot{\mathbf{z}}_i \equiv \dot{\mathbf{z}}$. For example, in the statement $Y_{z_i=1, \dot{\mathbf{z}}_i} \equiv Y_{z_i=1, \dot{\mathbf{z}}}$. Finally, use the prime notation ($'$) to define a distinct vector, meaning that \mathbf{z}' indicates that for at least one z_i indexed in \mathbf{z}' , $z_i \neq z'_i$. Similarly, $\dot{\mathbf{z}}'_i$ indicates that for some $z_{j \neq i}$ in $\dot{\mathbf{z}}'_i$, $z_j \neq z'_j$.

I use the potential outcomes framework to formalize the notion of a particular treatment causing an individual change (Rubin, 1974). Under this framework, then for every individual i , prior to the assignment of treatment there are two potentially observable outcomes: the individual's outcome if she receives treatment and her outcome if she receives control. In this notational scheme, let $Y_{i,\mathbf{z}}$ represent the potential outcome observed for individual i under treatment regime \mathbf{z} , and $Y_{\mathbf{z}}$ represent the potential outcomes to treatment \mathbf{Z} for all subjects.

Standard reasoning—which assumes non-interference—would at this point define an intent to treat effect for individual i as the difference between the potential outcomes when i is assigned to treatment and when i is assigned to control,

$$ITT \equiv E[Y_{i,z_i=1}] - E[Y_{i,z_i=0}].$$

In the presence of full compliance, this ITT is an unbiased estimate of the average treatment effect (ATE), the average of all the causal effects of treatment. However, in the presence of interference – alternatively called spillover in the

literature – the schedule of potential outcomes for each individual is defined not only through the assignment status of the referent individual, i , but also the assignment statuses of other individuals. Hence, the necessity of the \mathbf{z} and $\dot{\mathbf{z}}$ notational system. It is useful to note at this point that there is no *ex ante* prediction for which direction interference will bias the two group difference from the true causal estimand. The relationship depends on the particular relationship of the potential outcomes to treatment and control, the nature of spillover, and the possibility of displacement of subjects.

Baird et al. (2014) use this conceptual system to define causal estimand in the presence of interference. Particularly useful are identifying the quantities of the **Intent to Treat (ITT)**, **Spillover on the Non-Treated (SNT)**, and the **Total Causal Effect (TCE)** in the face of spillover.

Let the **Intent to Treat (ITT)** be the difference between expected potential outcomes of individuals when they are assigned to treatment compared to the case when they are assigned to control, holding all other assignments constant.

$$ITT = E[Y_i(z_i = 1, \dot{\mathbf{z}})] - E[Y_i(z_i = 0, \dot{\mathbf{z}})]$$

This indicates that the ITT effect for individual i is the result of setting her treatment status while holding the vector of other treatment assignments constant. This might mean changing the assignment of z_i from control to treatment while holding all other assignments at control ($\dot{\mathbf{z}}_i = 0$), or treatment, ($\dot{\mathbf{z}}_i = 1$), or some other vector value ($\mathbf{z} = \{1, 0, 1, 1, 0, \dots\}$).

Let the **Spillover on the Non-Treated (SNT)** be the difference between the expected potential outcomes for individuals assigned to control as the result of distinct treatment assignments.

$$SNT = E[Y_i(z_i = 0, \dot{\mathbf{z}})] - E[Y_i(z_i = 0, \dot{\mathbf{z}}')]$$

Finally, the **Total Causal Effect (TCE)** is the difference in overall expected potential outcomes, unconditional on identifying the treatment status of a particular individual i . Then, the TCE is

$$TCE = E[Y(\mathbf{z})] - E[Y(\mathbf{z}')]$$

Note, here, that because there is no individual-level indexing in the statement of TCE, any comparison in TCE from a comparison of \mathbf{z} and $\dot{\mathbf{z}}$ must occur at the

whole network level.

Two classes of \mathbf{z}'_i vectors hold particular interest. The first is the vector corresponding to all individuals are assigned to control, $\mathbf{z}_i = 0$, for all z_j . The second is a \mathbf{z}'_i vector with known network characteristics. Examination of the second class of vectors highlights that there is a potentially very large set of interesting potential outcomes comparisons to be made. For example, two particularly interesting examples of network characteristics are, (a) all one degree alters, $\mathbf{S}_i = 1$, being assigned to treatment; and (b) the five individuals with the largest number of social connections are assigned to treatment.

5.4.2 Estimation of Causal Effects

In practice, while estimation of causal quantities could proceed by group-means estimation, in the case that there exists some right-hand-side, pretreatment covariate \mathbf{X} such that the sum of the covariance between the potential outcomes and the outcome variable is greater than the variance of \mathbf{X} , then either rescaling outcomes or adjusting by covariate adjustment using \mathbf{X} will yield a more efficient estimate of the causal effect than an estimate that fails to use that covariate. That is, if

$$Cov(Y_i(0), X_i) + Cov(Y_i(1), X_i) > Var(X_i),$$

implying,

$$\frac{Cov(Y_i(0), X_i)}{Var(X_i)} + \frac{Cov(Y_i(1), X_i)}{Var(X_i)} > 1$$

then, including that \mathbf{X} indicator will improve estimation. Colloquially, a pre-treatment variable that supplies more predictive signal about the outcome variable than noise encoded in the pre-treatment variable will usefully improve the efficiency of causal estimand. This opens regression analysis as a straightforward estimation technique, not only because it is familiar to all empirical scholars, but also because under the design considerations highlighted above, and some reasonable regression assumptions, regression is a maximum efficiency, unbiased estimator of the treatment effect.

5.4.3 Assumptions about the form of interference

Common in the analysis of spillovers is to place some bound beyond which spillover is assumed to be zero. Termed *Stratified Interference* by Hudgens and Halloran (2008), the typical assumption is that within some identified cluster of subjects—a media market, a city, a classroom—interference exists, but that between these clusters there exists no interference. These assumptions are often quite easily met in the data.

Important under the *Stratified Interference* assumption is an assumption concerning the form of spillover within spillover clusters. Baird et al. (2014) assume a random effects form of interference. Under this random effects interference, for any treatment saturation—the number of individuals assigned to receive treatment from the fixed size population—the particular \mathbf{z} vector of individuals assigned to treatment does not effect any individual Y_i potential outcomes. That is, for some saturation π , built of treatment assignment vector \mathbf{z} ,

$$E[Y_i(\pi, \mathbf{z})] = E[Y_i(\pi, \mathbf{z}')]]$$

A consequence of this assumption is that within spillover clusters, exactly *who* is treated does not shape outcomes. Perhaps, this assumption is tenable in a spillover cluster that consists of co-habitant, intimate partners: whether one partner or another is treated, the causal effect is unchanged. However, in many other types of relevant spillover networks, this assumption does not meet well with substantive understanding. For example, within a network of staffers for political office, it would seem to be highly relevant if a treatment is assigned to the legislative intern or the director of staff.

5.5 Data

Data was collected as a part of the RSNS in Honduras. The data collection and process is described at length in chapter 4.

5.5.1 Assignment

The empirical distribution of social connections is plotted in Figure 5.1. Consistent with theory (Williamson, 1975; Sidanius and Pratto, 2001) and past

measurements (e.g. Apicella et al., 2012), the distribution of social connections in this population is heavily right-skewed. Because estimation of the effect of social connectedness relies on building experimental variation in the connectedness measure, and because a simple random assignment mechanism would assign a relatively large number of poorly-connected mobilizers, care was taken to build a criteria-stratified assignment mechanism. In particular, to ensure adequate variation in the hypothesized causal variable, I employ a stratified random sample with an intuitively derived stratification heuristic—distance from the center of town. In each town, the enumerator assigned approximately 40 percent of the mobilizers from the 20 percent most centrally located homes. Distance from the center of town was expected, *ex ante* to serve as a relevant instrument for social connectedness; this instrument was necessary, given the design, because in some villages assignment occurred prior to the completion of primary data-collection. In this case, individuals who had not been surveyed would be ineligible for assignment into treatment.

In each village, an enumerator was tasked with assigning individuals to a mobilization role. This enumerator was provided a laptop computer loaded with a population list, a randomizer, and a map. The enumerator would select a button in the survey software and a unique identifier was presented. The unique identifier contained information about the randomly selected subjects' house number; this house number was matched against the map of the town made by surveyors. A decision about how many mobilizers to assign was made by the researcher prior to the enumerator starting his task, and was based on the size of town. The smallest towns were assigned either three or four mobilizers, while the largest towns were assigned to receive between four and eight mobilizers.

At this point, the criteria-stratification was evaluated. Specifically, enumerators were instructed to select between one and two mobilizers from the houses located "near" the center of town. In the event that the criteria stratification had not been met, the enumerator would seek out the house and the individual inform the subject that he or she had been selected as a mobilizer. The mobilization script is included in the appendix to this chapter. In the event that the criteria stratification had been met – two or more mobilizers had already been selected from near the center of town, then upon drawing a home within the center of town, the enumerator would simply re-select another individual. Likewise, in the event that the individual could not be identified—he or she was at work, visiting another town, or otherwise unable to be located—the enumerator would select another random

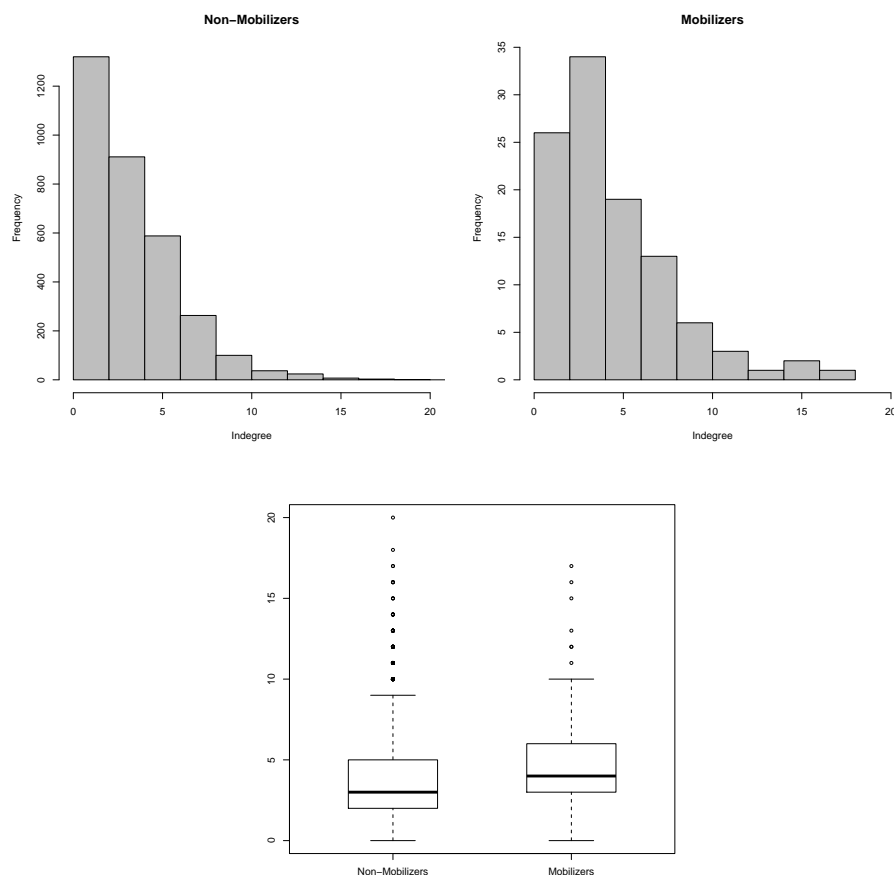


Figure 5.1: Histogram of the distribution of indegree connections. The distribution is right-skewed. A large number of individuals have a relatively small number of social connections while a small number of individuals have a relatively large number of social connections.

draw using the town software.

Care was taken to attempt assignment after the men and women had returned home from work to mitigate the threat that reached mobilizers were systematically different from unreached mobilizers; the initial recruitment was typically successful, with the enumerator estimating that he contacted approximately eighty-five percent of the primary targets.

An examination of the covariates of those selected to be mobilizers suggests that the mobilizer corps is largely similar to the population not assigned to be mobilizers. These results are reported in Table 5.1. Consistent with the targeting strategy, mobilizers were better connected than non-mobilizers. In addition, mobilizers were more likely to carry covariate profiles predicted by being better connected – they are

more likely to be married, and more likely to have taken part in civic activity in the past. Importantly, there are no difference between mobilizers and non-mobilizers on covariates that are not predicted by being well connected – namely age and gender.

Recall, however, that the core causal claim under examination in this paper is not that assigning individuals to be a mobilizer makes those individuals more likely to take political action. Indeed, the targeting used in the assignment mechanism explicitly unbalances the group of individuals assigned as mobilizers from those not assigned. Instead, the core question under examination is whether the assignment of a social alter as a mobilizer increases a subject's likelihood of taking political action. Critical, then, is establishing that, of those not assigned to be mobilizers, there are no pre-treatment differences between individuals assigned to have a social alter serve as a mobilizer and those assigned to have no social alters assigned as mobilizers. To examine this question, I compare mean values, and test for a difference of means on observable characteristics. The mean and sem of this comparison are reported in Table 5.2.

5.5.2 Treatment Package

Individuals assigned to be mobilizers were informed of two facts. First, mobilizers were informed that the study team planned to hold a town meeting two-days in the future where the team would thank members of the community, explain our research, and provide information about a microfinance organization that intended to start work in the region in the coming months. Second, mobilizers were instructed that they had been selected at random help the study team bring individuals to this meeting, and that we were going to provide a form of compensation for their help. Specifically, we informed each assigned mobilizer that at the subsequent meeting we would hold a raffle; for each individual who the mobilizer convinced to attend the meeting, we would enter a ticket in that raffle to win 100 Lempira.³ Our prior experience in the region informed us that this concept of a raffle was well understood, a fact confirmed in our enumerators' conversations with each mobilizer. The specific language read to mobilizers is included in section 9.1.

Among those who were assigned to receive the treatment, seventy percent

³At the time, the conversion rate between Honduras Lempira and US Dollars equated this sum to be approximately \$5.50. At the time of the study, this quantity was equal to the federally set daily minimum wage.

attended the meeting. The because better connected individuals were explicitly targeted through the center-of-town heuristic, a two group comparison between those individuals assigned to receive treatment and those assigned not to receive treatment does not hold a causal a design-based causal interpretation. Indeed, as reorted in Table 5.1, those assigned as mobilizers are married at greater rates, have previously been more politically active, and, by design, are hold a greater number of social connections. Among the set of subjects assigned as mobilizers, however, little predicts the mobilizers' eventual turnout at the meeting. Table 5.3 presents these results, and shows that there is little evidence of a systematic relationship between covariates and meeting attendance.

5.6 Results

5.6.1 Estimating Within Cluster Controls

To facilitate comparisons between treated and untreated individuals in the pursuit of causal estimands, it is necessary to identify the units designated as control. In a design where spillover of assignment status is possible, it furthermore necessary to identify those individuals who maintain the apples-to-apples comparison of balance on potential outcomes, but who also receive no spillover from individuals assigned to treatment. One promising avenue for this estimation, though it is not pursued in this design, is the creation of individuals to receive "ghost-treatment", or effectively individuals randomized into the treatment group but whom *by design* do not receive treatment (Johnson, Lewis and Nubbemeyer, 2015). Developed in the context of poorly-defined comparison groups in the online-advertising context, the concept may also be usefully applied to estimating causal estimands in the context of interference. In essence, once a model for interference has been identified, it is possible to use the individuals who receive *ghost-treatment* and the social alters of these *ghosts* to form a comparison group for those who do, in fact, receive treatment. The comparison then, would be the difference in outcomes between those individuals spilled over from a treatment individual and the outcomes from those individuals (not) spilled over from a ghost-treatment individual.

A first alternative approach, and the one pursued in this work, attempts to identify non-spilled over individuals through estimation rather than design. One such method would identify subjects within a treatment cluster who are social

isolates – individuals who have no social alters – and therefore are unlikely to have received spillover from treated individuals. These individuals can be used to identify the baseline "hum" or "buzz" about treatment that cannot be attributed to a systematic (e.g. peer-to-peer, or institution-based) spillover mechanism. Once concern with this approach is that those individuals do not hold social alters are unlikely to be broadly representative of the individuals randomized into either treatment or possible interference conditions. In the case that the researcher has *ex ante* expectations that the systematic differences will cause a two-group comparison to underestimate (biased downward, attenuation biased) the true causal effect, this *social isolate* strategy may still be profitably used to estimate a minimum-causal effect, though this strategy clearly comes at the expense of potentially failing to reject the null hypothesis of no-interference when in fact interference does exist.

A second alternative approach would use *ex post* outcome data to estimate the systematic components of spillover, and then use those identified as unlikely to be spilled-over-to as control individuals. The procedure might take the following form:

1. Randomize individuals into treatment;
2. State a model for systematic spillover, e.g. "spillover occurs as a function of social distance";
3. Estimate the degree of spillover, e.g. "evidence is found for spillover at one and two degree relationships, but not three-or-more degree relationships";
4. Compare individuals identified as being unlikely to receive spillover against those who receive treatment (identifying the ITT), or against those likely to receive spillover (SNT).

5.6.2 Characterization of Mobilizers in Towns

The key, causal variable in this analysis is the connectedness of the mobilizer corps in each town. Recall this mobilizer corps is the result of a single assignment vector, \mathbf{z} . Figure 5.2 characterizes the distribution of social connectedness realized by the realized assignment vector in blue, and plots this realization against a simulation of 10,000 alternative draws. The use of the location-based caliper appears to have successfully targeted slightly more well-connected mobilizers than would

be expected by chance. Indeed, in the realized mobilizer corps, there are fewer very-poorly connected mobilizers, and slightly more well-connected mobilizers. However, it is important to note that there is good coverage across the entire range of mobilizer connectedness.

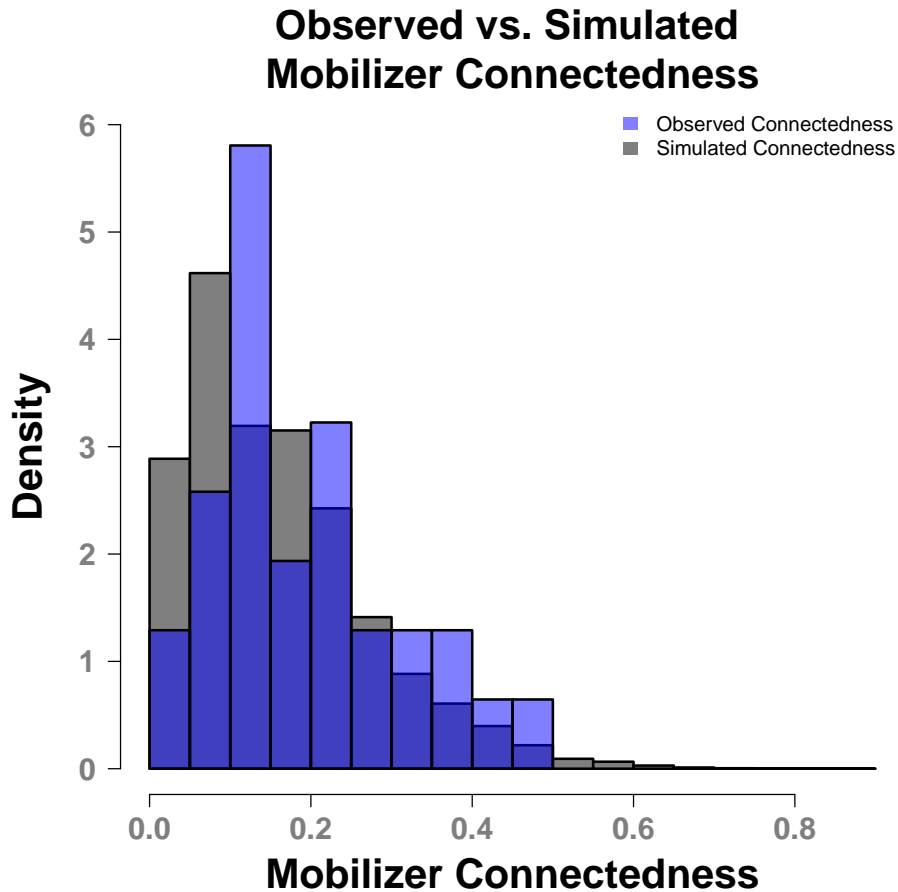


Figure 5.2: Mobilizer connectedness, measured as each mobilizers' *indegree* scaled by the total population in the village. The blue histogram plots the observed values of mobilizer connectedness, and the grey histogram plots the histogram of connectedness from 10,000 simulated draws of mobilizers within towns. The assignment mechanism performed well at covering the range of possible assignments while targeting assignment of subjects to be mobilizers who were slightly better connected.

Figure 5.3 further examines the distribution of mobilizer connectedness by breaking down the distribution of connectedness based on the size of the town. Towns are binned into population bins of size one hundred, and the connectedness measure reported is the per-mobilizer total number of connections. In small towns,

with population less than one hundred, the median number of connections held by a mobilizer is just fewer than four. In larger towns, with populations greater than one hundred, the median number of connections is between 5.5 and 6.5 ties per mobilizer. A correlation test does not find evidence that mobilizers in larger towns are better connected (Spearman correlation test, $\rho = 0.28$, $p = 0.13$).

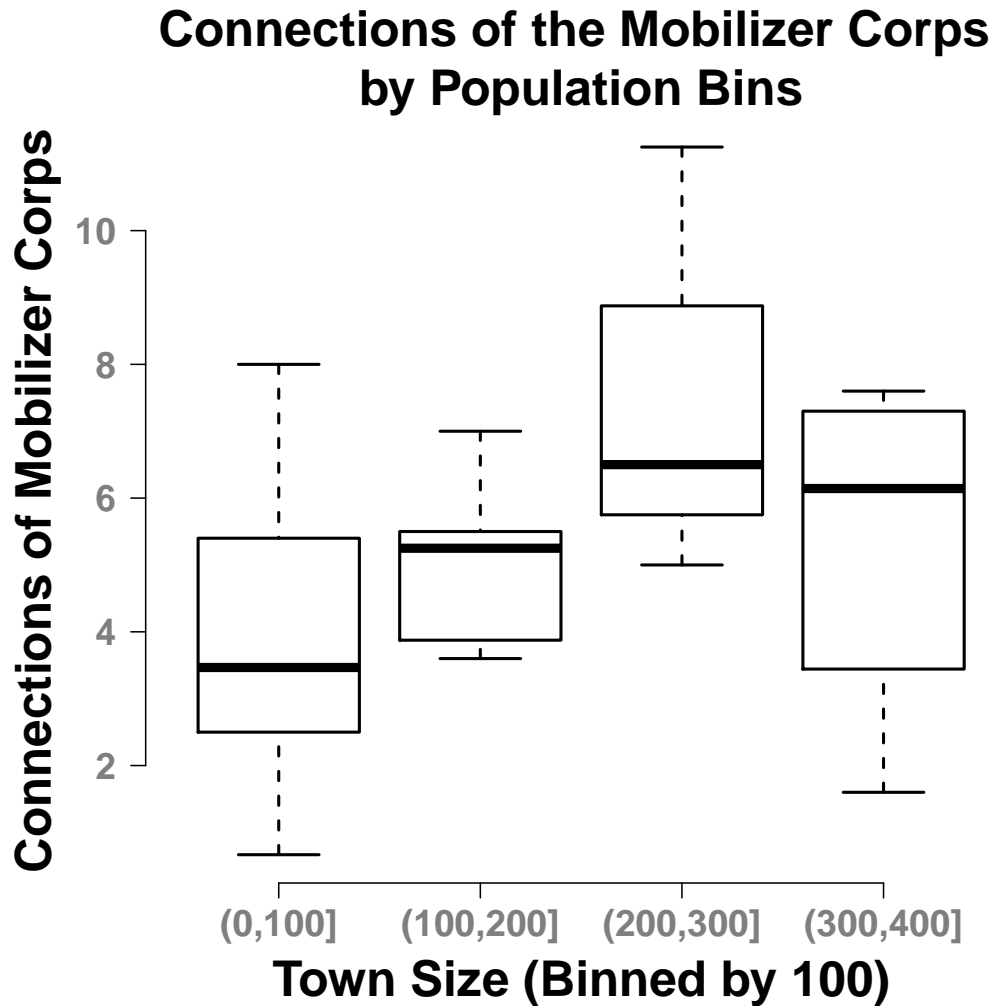


Figure 5.3: Mobilizer connectedness plotted against mobilizers' town size. The y-axis plots the total number of connections held by the mobilizer corps, divided by the number of mobilizers. Mobilizers living in larger towns have slightly more connections, though there is no difference in between any town with at least 100 residents.

5.6.3 Town Level

I begin the analysis of this mobilization experiment by estimating the **Total Causal Effect** of a well-connected mobilizer corps. To do so, I use the exogenously generated, between-town variation in the connectedness of mobilizers. Recall the TCE causal model presented in subsection 5.4.1,

$$TCE = E[Y(\mathbf{z})] - E[Y(\mathbf{z}')].$$

As previously noted, the potential outcomes for the TCE are not indexed in i , and estimation comes as a result of between-cluster comparisons of treatment regimes.

A useful reference for \mathbf{z}' would be the vector where no subjects are assigned to treatment: the pure control case. Then, the average causal effect could be estimated as

$$\begin{aligned} E[TCE] &= E[Y(\mathbf{z})] - E[Y(\mathbf{z}(\mathbf{0}))] \\ &= E[Y(\mathbf{z})] - 0, \end{aligned}$$

Although useful to assess the turnout that would arise as a result of residents' curiosities and then estimate the turnout in treatment towns compared against this baseline, in the study fielded for the RSNS, assigning clusters (villages) to pure control could not be completed for logistical reasons. Creating pure control clusters might best be achieved through a stepped-wedge design where some proportion of data receives treatment, an observation is made of all data, and then those units that did not receive treatment in the first intervention stage receive treatment in the second intervention stage. Because the group of enumerators recruited for this were students — either college students in Michigan or highschool students in Honduras — there was insufficient time in the summer break between sessions to allow two intervention periods. Additionally, and even more to the point, the concept of a "pure control" in this particular case is somewhat misguided. Indeed, to create a set of village as pure control would have meant surveying a entire village for baseline demographic characteristics and then holding a political activity without informing any of the village members such an activity was scheduled.

Rather than pursue this pure control strategy, as an alternative inferential strategy I estimate the marginal changes in turnout resulting from variation arising in the connectedness of the mobilizer corps. Provided the assignment mechanism of mobilizers is not correlated with features of the town units, this strategy retains the causal interpretation warranted by the previous estimate of the TCE, but with the benefit of retaining individuals in treatment in all units.

To estimate this relationship, I estimate the following model:

$$Y_v = \alpha + \tau_{TCE} * D_v + Z_v\beta + \epsilon_v,$$

The dependent variable in this model, Y_v , is a count of the number of individuals who attend the village meeting in village v ; τ_{TCE} is the marginal increase in attendance at the village meeting as a result of a marginal increase in the exogenously assigned mobilizer connectedness; β is a vector of non-causal estimates of the relationships between predictors, Z_v that potentially improve the efficiency of the causal estimand; α is an estimate of the turnout when a mobilizer corps has zero social connections (and other covariates are zero); and, finally ϵ_v is the residual between the count of village meeting attendance and the fit model.

The results from a linear probability model (OLS regression) for the relationship between mobilizer connectedness and town-level turnout are presented in Table 5.4. The results from a poisson regression are presented in Table 5.5.

To test for the possibility that the relationships between mobilizer connectedness operates differently in towns of varying size, Table 5.5 presents the regression estimates for the total causal effect broken down by town population. Table 5.5, column one presents the estimates for towns of all size. Here, as reported in the previous paragraph, for every extra social connection in the mobilizer corps approximately 1.5 more individuals attended the village meeting. In the hypothetical case that the mobilizer corps held *no* social connections with others living in the village, this model predicts that 4.5 individuals would turn out to vote. While in general interpreting the intercept of a regression holds little substantive meaning, in this case, the predicted number of attendees at the village meeting when mobilizers hold no social connections is remarkably similar to the 4.17 mobilizers assigned on average in each town.

Table 5.5, columns two through five examine the possibility that the relationship between mobilizers' social connectedness and village-level attendance at the meeting are different by subsetting the data on village size and running separate

regressions. While there is *very* little data in any one of these regressions, in each, the relationship between mobilizer connectedness and turnout is positive, and remarkably stable around the whole-sample estimate. As a formal test for differences in the relationship conditional on town size, Table 5.5, column six interacts the causal variable with town size.

While there is little data for a model of this complexity to fit the results presented in Table 5.5 are notably stable, and suggest a positive relationship between mobilizers' connectedness in all size towns. Indeed, fitting a model on the full dataset suggests that a two standard deviation change in mobilizer connectedness⁴ causes the predicted attendance at the village meeting to more than double, from about twenty-two residents in attendance to about forty-seven (95% prediction $CI_{low} = [19.23, 23.91]$; 95% prediction $CI_{high} = [43.18 - 50.28]$). These data also suggest that the causal effect of better connected mobilizers is more than twice the magnitude in this sample's larger towns than the smaller towns. To form this comparison, compare the estimated interaction coefficients in Table 5.5, column (6). In every case, the interaction term for treatment in each of the larger town indicators is larger in magnitude than the baseline effect estimated in the smallest towns. However, it bears mention that these heterogeneous effects are both fundamentally non-causal and also the product of estimating a flexible model on a very small dataset.

Figure 5.4 plots the bivariate relationship between the dependent variables, meeting attendance at the village level, and the causal variable, mobilizer corps connectedness. Window (a) plots this relationship without rescaling by the size of the town; figure (b) rescales these raw numbers by total town population. Importantly, there is little change in the causal relationship as a result of this rescaling. For every two additional connections held by the mobilizer corps, nearly three more village residents turned out to participate in the village meeting ($\beta = 1.44$, $SE_{\beta} = 0.30$, $p < 0.001$).

5.6.4 Individual Level

How does the mobilization cue spillover through political actors' social networks? In the previous section I presented evidence about the positive relationship between mobilizers' social network characteristics and turnout at a village-wide

⁴From the mean minus one standard deviation to the mean plus one standard deviation.

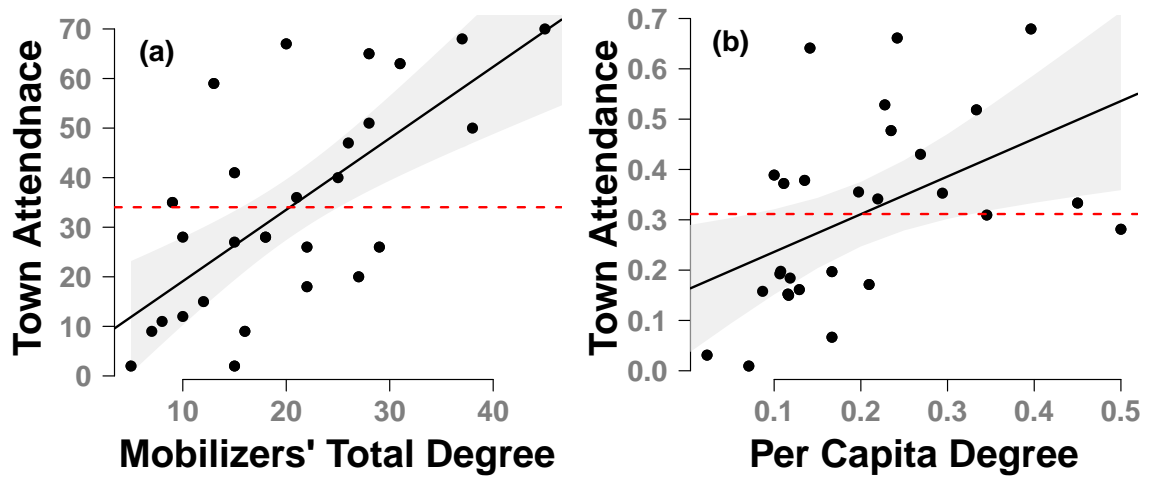


Figure 5.4: Meeting attendance by mobilizer connectedness. Each plot displays the causal variable, mobilizer connectedness, on the x-axis and the dependent variable, town meeting attendance, on the y-axis. Data realizations are plotted as points. The solid black line is the best fit OLS regression line, and the dashed red line is the best fitting null (intercept only) model. The grey region is the 95% confidence interval for the estimated best fit OLS regression line. Panel (a) presents these results without rescaling by town population; while panel (b) rescales both variables by town population.

political event. The more social connections held by a cohort of mobilizers, the more individuals turned out to participate in the political activity.

In this section, I continue to examine the features leading to the political turnout, paying particular attention to difference in turnout attributable to spillover from treated to non-treated units. A theory of social information predicts that in the case of costly action, individuals will be most effective at spreading information and influence among alters who are socially proximate. In Figure 5.5 I present evidence that those subjects with a greater number of socially proximate alters assigned as mobilizers are more likely to take political action, and in Table 5.7 I present complementary, model based evidence for this causal effect. I expand upon each in turn.

Table 5.6 presents that results of a non-causal regression predicting the number of individuals a mobilizer is responsible for bringing to the political meeting as a function of mobilizer characteristics. Because mobilizers are assigned at random, the number of individuals turned out by a mobilizer holds a limited causal

interpretation; indeed, this quantity holds a clear relationship with the spillover on the non-treated causal quantity previously described. However, because the *traits* of mobilizers are not randomly assigned, it is not possible within this framework to identify which trait of the mobilizer is responsible for increasing the mobilizers' effectiveness. Stated another way, it is possible to identify causal effects in this setup, but it is not possible to establish, with certainty, causal mechanisms. Even with this caveat, examining the features of mobilizers that are associated with differential performance in the mobilization task can lend useful circumstantial evidence to the argument for social information.

Table 5.6 presents the results of a poisson regression of the number of individuals mobilized to take political action on the characteristics of each mobilizer. In this model, observations are individuals assigned as mobilizers, meaning there are 104 observations (99 that hold full covariate data) in the regressions. Consistent with the predictions of a theory of social information, and specifically with a prediction of increased effectiveness for costly action among individuals who are socially proximate, mobilizers who hold a larger set of social connections are more successful at bringing individuals to attend the political activity. Indeed, a two standard deviation change in the number of social connections held by a mobilizer predicts a doubling of the number of alters the mobilizer is predicted to bring to the political activity, from a predicted value of 4.8 to a predicted value of 9.8. These results are similar to those presented in Figure 5.4 which aggregated this relationship up to the town-level. Additional features that predict a mobilizer being more able to perform the mobilization task are living in the center of town, ($\beta_{M(4)} = 0.121$, cluster robust $SE = 0.03$) and being older ($\beta_{M(4)} = 0.012$, cluster robust $SE = 0.006$). There is little evidence in this data mobilizers who have previously taken political action, or are married have greater mobilization capacity. One possibility for this null finding is that because connectedness is not randomly assigned, these other effects are operating through other channels (e.g. *Connectedness* and *Probability of Assignment*).

Recall Table 5.1 and Table 5.2 which presented estimates of covariate balance between mobilizers and subjects who were assigned to have a mobilizer at varying degrees of social distance, respectively. With the exception of a difference in the probability of being married, subjects assigned to have a mobilizer at a social distance of one⁵ were indistinguishable from subjects assigned to have a mobilizer

⁵A friend was assigned as a mobilizer.

at a distance of two.⁶ These individuals one and two degrees from a mobilizer represent more than sixty percent of the observed data; those one, two, or three degrees from a mobilizer represent more than 95% of the observed data.

To estimate both the direct causal effects of being assigned as a mobilizer, as well as the spillover effects from mobilizers to social alters, I estimate a model of the following form:

$$Y_i = \alpha + \tau D_i + \phi E_i + \delta \sum_{j \neq i} (D_{j,S}) + \gamma \sum_{j \neq i} (E_{j,S}) + \beta X_i + \mu T + v_i$$

In this model, the dependent variable, Y_i , is the binary outcome for whether individual i attended the town meeting. α is the baseline turnout rate when all other variable are zero, D_i is the assignment of individual i to serve as a mobilizer, and τ is the difference in the rates of turnout between those who were assigned to be mobilizers and those who were not assigned to be mobilizers; ϕ estimates the difference in the probability of each being assigned as a mobilizer, under the stratified random assignment regime. The term $\sum_{j \neq i} (D_{j,S})$ is the sum of the number of alters, j , assigned as mobilizers under an assumed spillover model, S . In Table 5.7 the spillover model examined is spillover across social networks, with the flexibility to estimate differential spillover across levels of social distance. Then, δ is the causal effect of having social alters assigned to be mobilizers, the SNT. The term $\sum_{j \neq i} (E_{j,S})$ estimates the effect of the expected number of social alters of individual i assigned as mobilizers; this expectation is the product of two features, the social location of subject i , specifically the number of social alters he or she holds, and the stratified randomization scheme. Then, γ serves as a control variable bringing into balance social connections between individuals having social connections according to S_i .⁷ X_i is a vector of individual covariates that may improve model fit and contextualize estimated causal relationships. β then is a vector of coefficients associated with non-experimentally assigned individual-level covariates; T is a town-fixed effect to remove unmeasurable town-level difference, and v_i is the vector of residuals from the fit regression. This model is intentionally similar to the model estimated in

⁶A friend of a friend was assigned as a mobilizer.

⁷The expected number of mobilizers is calculated for each ego by summing the probability each alter was assigned to be a mobilizer, by the number of alters. For example, an ego with three friends, each of whom have an *ex ante* chance of 0.25 of being assigned as a mobilizer has a $3 \times 0.25 = 0.75$ expected number of distance one alters as mobilizers.

Bond et al. (2012), which estimated a similar effect on a network of internet-based, social network users. Importantly, Bond et al. (2012) simulated the properties of this estimator; these MCMC simulations found little bias in estimates, and found Type-I and Type-II error rates in line with specified rates (Bond et al., 2012, SI p. 12).

Across all fit models reported in Table 5.7 the effect of being assigned as a mobilizer ranges between a four and five times increase in the probability of taking political action if assigned as a mobilizer ($OR = [4.05, 5.47]$, lowest 95% $CI = 2.69$, highest 95% $CI = 10$). Turning to examine the effect of assigning an individual's social alters as a mobilizer, there is clear evidence that assigning both "friends" and "friends of friends" as mobilizers increases the likelihood an individual turns out to take political action. Indeed, Table 5.7, Model (6) estimates that for each distance one social alter assigned as a mobilizer, an individual is approximately one and a half times more likely to take political action ($OR = 1.56$, 95% $CI = [1.18, 2.07]$). At estimated values then, the marginal effect of assigning an individual as a mobilizer is the same as assigning 2.5 friends as mobilizers. Even more, there is evidence across all models for the effect of assigning an alter a social distance two; the odds ratio for a single alter is 1.38 (95% $CI = [1.21, 1.52]$).

Figure 5.5 presents a non-model based representation of the meeting attendance rates of individuals. Individuals with neither a distance one or distance two mobilizer turn out at the meeting at a rate of 21%. Individuals with a single distance two mobilizer (but no distance one mobilizers) attend at slightly higher rates, about 35%; a single distance one mobilizer (and no distance two mobilizers) attend at a rate of 41%. Those with a single one degree mobilizer and a single two degree mobilizer turn out at the meeting at a 50% rate. Figure 5.5 presents the remainder of the comparisons.

5.7 Limitations

Despite the strength of the random assignment procedure and the inferential strategy there continue exist some limitations in this study. The primary inferential limitation in this study is drawn from the form of political activity being monitored. That is, the chief limitation is the lack of designed pure-control villages. This was a decision evaluated in the design phase, where it was decided that costly data collection in an entire village, while not applying treatment in that village was infeasible.

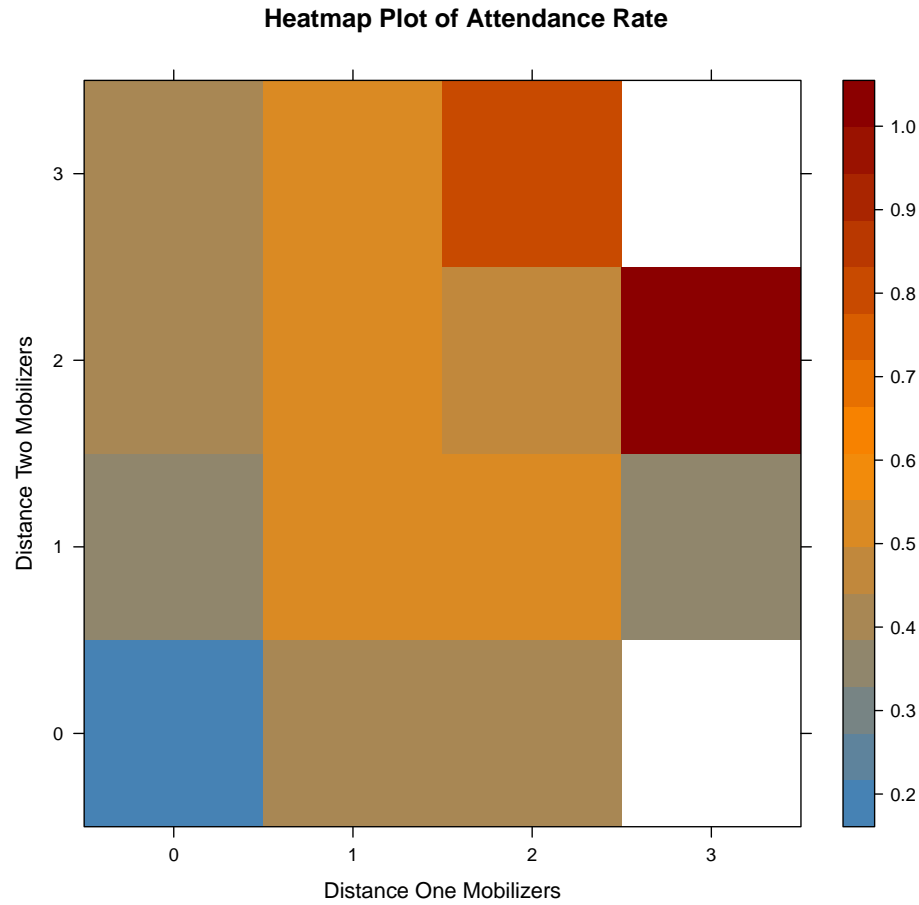


Figure 5.5: Heatmap plot of attendance at political activity. Blue colors are the lowest likelihood of attending the political activity, red are the highest probability. Probability is a data-based (i.e. not model based) simple probability that an individual with Distance 1 and Distance 2 mobilizers of a particular number attended the political activity.

Despite this limitation, this design is able to estimate rates of change in turnout that have a causal interpretation, even if this estimate is somewhat less straightforward on first impression.

The second limitation concerns the how generally these results might be mapped. Indeed, this study was both undertaken in a unique global setting—rural Honduras—and also explicitly incentivized subjects to spillover behavior to non-treated alters. To be certain, the first of these limitations poses considerable challenge; should one expect that a similar social spillover mechanism might exist among political activists in the United States, or among political violence reduction campaigns in east Africa, or in maternal and child health outcomes in Latin

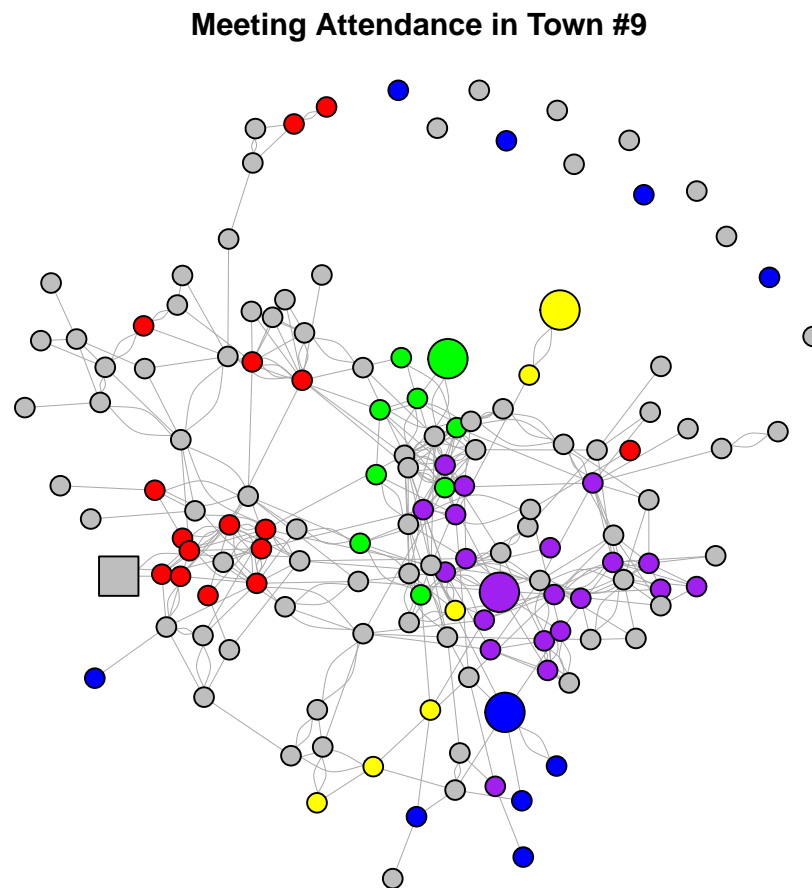


Figure 5.6: Social network plot for village 9. Nodes in this plot represent individuals; lines represent social connections. The large nodes are those individuals who were assigned as mobilizers, and the small nodes are all other residents. Grey nodes did not attend the meeting. Colored nodes attend the meeting, and are colored to match the color of the mobilizer who brought them to the meeting. Note that the large grey square was assigned as a mobilizer (and would have been colored red), but did not attend the meeting.

America? These are valid concerns, though the concerns do not uniquely apply to the research in this volume, but instead apply to laboratory, quasi-laboratory, and even field experiments at any time they are run. To the extent that what has been demonstrated in this chapter—that individuals seek to influence those with whom they are socially proximate, and that individuals are, in fact, subject to this influence—is a general trait of humans as a social species, then these results hold broad import into other domain. As noted in the theory development in an earlier chapter in this dissertation, the major scoping that is theorized to increase or decrease the importance of social ties are the costs borne by up-taking subjects.

Finally, the design-choice was made to specifically incentivize subjects to spill the behavior over to untreated subjects. Indeed, this choice almost certainly ensures that the magnitude of the spillover observed in this data is greater than the magnitude that might be expected due only to social observation of behaviors; the magnitude of spillover for a voter education campaign, a decision to support a controversial candidate, or an emerging policy-position will very likely be more subtle than the results presented here. Yet, a large part of political action involves creating consensus among a coalition of actors. When subjects have incentives to form as large a group as possible for some campaign, then these results suggest that they may be most successful among those with whom they hold social connections.

5.8 Conclusion

In this chapter I have argued that when a researcher holds data about the pathways of spillover it is possible to estimate causal quantities in a field experimental setting. I built upon the causal framework presented in Baird et al. (2014) to measure the Intent to Treat Effect, Spillover on the Non Treated, and Total Causal Effects of a randomly assigned incentive structure that was presented to three thousand residents of a rural, Honduran system of towns. As in Baird et al. (2014), statements of causal models influence the specification of regression-based models for the purposes of drawing inference. However, unlike the previous findings in (Baird et al., 2014), in this intervention where subjects assigned to receive treatment were specifically incentivized to spillover treatment, I find clear and robust effects of spillover on the non-treated. In the finding for spillover to the non-treated, the results in this chapter are most similar to those of Bond et al. (2012) and Fafchamps, Vaz and Vicente (2013). However, several features distinguish these results from

previous findings. First, in this chapter I find evidence that is robust to the level of inquiry that when political action is seeded with individuals who are highly connected within their social networks that program uptake or spillover are greater than if the political action had been seeded with individuals who are not highly connected within their social networks.

There are several implications for this work, both theoretic and practical. I take each in turn. The findings in this chapter closely comport with the theory of social information presented in the theory chapter of this dissertation. The theory of social information predicts that the content of messages between individuals is conditioned by both the person who is sending the message and also by the person who is receiving the message. In particular, in this case when the sender of the message about political activity was better connected, the message was more readily taken up by social alters. Additionally, social alters were more likely to take up a message if the sender of that message was socially proximate.

Even with the acknowledged limitations for generalizability identified in the previous section, these findings hold the possibility for considerable change to be made.

Table 5.1: Covariate balance check. Logistic regression predicting assignment as mobilizer.

	Assigned as Mobilizer (1)	Assigned as Mobilizer (2)
Age		0.002 (0.002)
Married		0.422 (0.462)
Years Education		−0.001 (0.005)
Female		−0.223 (0.300)
Voted		0.539 (0.342)
Connectedness		0.147*** (0.037)
P(Assign)		16.790*** (3.880)
Constant	−3.480*** (0.115)	−5.887*** (0.534)
N	2,609	2,576
Log Likelihood	−350.600	−226.340
AIC	703.210	468.690

*p < .1; **p < .05; ***p < .01

Logistic regression predicting assignment of individuals to be a mobilizer. Connectedness is the total number of social nominations from other individuals. 'P(Assign)' is the exact probability of assignment according to the targeting procedure.

Table 5.2: Social Balance Table. Comparing covariate values of those assigned to have a social alter at distance $dist=k$ assigned as a mobilizer.

	Mean.d1	SE.d1	Mean.d2	SE.d2	Mean.d3	SE.d3	Mean.d4	SE.d4
Age	40.30	1.55	42.03	1.60	38.18	2.07	34.20	4.11
P(Female)	0.48	0.02	0.52	0.01	0.60	0.02	0.65	0.03
P(Vote)	0.70	0.02	0.63	0.01	0.52	0.02	0.38	0.03
P(Married)	0.88	0.01	0.83	0.01	0.73	0.01	0.65	0.03
P(Assign)	0.04	0.001	0.03	0.001	0.03	0.001	0.02	0.001
N Obs.	594		1,357		958		218	

Notes: Comparison of means and standard errors for distances from assigned mobilizers. 'd1' are subjects who are one degree from a mobilizer, 'd2' are subjects who are two degrees from a mobilizer, and so on. Distance is calculated as shortest-path distance along all social network ties.

Table 5.3: Ordinary Least Squares Model Predicting Mobilizer Turnout at Political Meeting

	<i>Dependent variable:</i>	
	Mobilizer Attended Meeting	
	(1)	(2)
Age	0.001 (0.02)	0.02 (0.02)
Married	1.47* (0.77)	−0.42 (1.27)
Female	0.58 (0.50)	0.94 (0.69)
Voted	−0.66 (0.61)	−1.29 (0.87)
Probability of Assignment	9.82 (8.04)	−13.02 (19.60)
Intercept	−0.66 (1.24)	20.51 (5,315.00)
Town FE	No	Yes
Observations	99	99
Log Likelihood	−54.92	−34.99
Akaike Inf. Crit.	121.84	127.99

Note: *p<0.1; **p<0.05; ***p<0.01

Table 5.4: Linear Probability Model. Total town political activity (meeting attendance) regressed on mobilizers connections and town size.

	Meeting Attendance		
	(1)	(2)	(3)
Mobilizers Connectedness	1.43*** (0.24)		1.30*** (0.45)
Town Population		0.09* (0.05)	0.03 (0.05)
Constant	6.03 (6.01)	22.90*** (7.03)	5.00 (6.63)
N	24	24	24
R ²	0.50	0.21	0.51
Adjusted R ²	0.48	0.17	0.47
F Statistic	22.11*** (df = 1; 22)	5.69** (df = 1; 22)	11.11*** (df = 2; 21)

*p < .1; **p < .05; ***p < .01

Notes: OLS regression of total number of individuals attending village political meeting on total indegree connectedness of mobilizers. Column (1) reports the regression with only mobilizer indegree; column (2) reports the regression with only town population; column (3) reports the regression with mobilizer connectedness and town population; and, column (4) tests for a differential relationship conditional on town size.

Table 5.5: Poisson regression. Total town political activity (meeting attendance) regressed on mobilizers connections and town size.

	Meeting Attendance						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Mobilizer Connectedness	0.037*** (0.003)	0.025*** (0.008)	0.054*** (0.013)	0.021** (0.010)	0.048*** (0.009)	0.035*** (0.005)	0.025*** (0.008)
Town Pop: 100-200						0.296*** (0.090)	-0.338 (0.354)
Town Pop: 200-300						0.111 (0.146)	0.465 (0.407)
Town Pop: 300-400						0.095 (0.115)	-0.516 (0.330)
Connectedness * Town Pop: 100-200							0.029* (0.015)
Connectedness * Town Pop: 200-300							-0.004 (0.013)
Connectedness * Town Pop: 300-400							0.024** (0.012)
Intercept	2.731*** (0.086)	2.840*** (0.139)	2.501*** (0.325)	3.305*** (0.382)	2.323*** (0.299)	2.669*** (0.096)	2.840*** (0.139)
Data Subset	All	0-100	100-200	200-300	300-400	All	All
N	24	12	6	2	4	24	24
Log Likelihood	-149.250	-79.294	-38.779	-5.890	-15.096	-143.260	-139.060
AIC	302.490	162.590	81.558	15.780	34.192	296.520	294.120

*p < .1; **p < .05; ***p < .01

Notes: Poisson regression of total number of individuals attending village political meeting on total indegree connectedness of mobilizers. Column (1) reports the regression with the whole dataset. Columns (2) - (6) examine the possibility that the relationship is conditionally different based on town population. Columns (2) - (5) subset the data by town population, and column (6) interacts a town-size indicator with the causal variable. Having well connected mobilizers in small towns increases turnout; having well connected mobilizers in larger towns has roughly double the effect.

Table 5.6: Poisson Regression. Mobilizer effectiveness as a function of social connections.

	Number of People Turned Out				
	(1)	(2)	(3)	(4)	(5)
Connectedness	0.119*** (0.025)	0.120*** (0.026)	0.127*** (0.026)	0.121** (0.048)	0.121*** (0.032)
P(Assign)		0.843 (3.294)	2.909 (3.305)	8.123 (7.270)	8.123 (6.245)
Age			0.011** (0.006)	0.012 (0.009)	0.012* (0.006)
Voted			0.296 (0.281)	0.197 (0.348)	0.197 (0.305)
Female			−0.112 (0.238)	−0.112 (0.365)	−0.112 (0.260)
Married			0.858 (0.639)	0.600 (1.003)	0.600 (0.769)
Intercept	1.452*** (0.156)	1.406*** (0.240)	−0.205 (0.782)	−0.784 (1.416)	−0.784 (1.044)
Town FE	No	No	No	Yes	Yes
N	104	104	99	99	99
Log Likelihood	−531	−531	−475	−402	−402
AIC	1,067	1,069	965	865	865

*p < .1; **p < .05; ***p < .01

Notes: Poisson regression predicting count of individuals turned out to meeting. DV is number of people at meeting identifying a mobilizer as responsible for their attendance. Connectedness is the total number of social connections of a mobilizer. Models (1-4) use Huber-White heteroskedastic-consistent errors. Model (5) uses Huber-White heteroskedastic-consistent errors clustered at the town level.

Table 5.7: Core causal model. Logistic regression of turnout on experimental and non-experimental factors

	Individual Attends Meeting		
	(1)	(2)	(3)
Mobilizer	1.516*** (0.248)	1.696*** (0.259)	1.705*** (0.314)
Distance 1 Mobilizer Alters	0.261*** (0.086)	0.466*** (0.122)	0.451*** (0.151)
Distance 2 Mobilizer Alters		0.323*** (0.096)	0.306** (0.125)
Distance 3 Mobilizer Alters		0.158* (0.091)	0.108 (0.113)
Distance 4 Mobilizer Alters		0.083 (0.081)	0.077 (0.097)
Age			-0.001 (0.002)
Married?			0.338** (0.133)
Years Edu			0.002* (0.001)
Female?			0.775*** (0.102)
Voted Last Election?			0.385*** (0.107)
Connectedness			0.003 (0.027)
Intercept	-0.334 (0.225)	-0.564* (0.299)	-1.709*** (0.371)
Town FE	Yes	Yes	Yes
Propensity Score Mobilizer?	Yes	Yes	Yes
N	3,346	3,346	2,604
Log Likelihood	-1,856	-1,846	-1,369
AIC	3,768	3,760	2,818

* $p < .1$, ** $p < .05$, *** $p < .01$

Notes: Logistic Regression of individual turnout to political activity. The outcome variable is measured turnout. "Distance 1 Mobs" are the number of individuals assigned as a mobilizer at a social distance of 1. The same is true for distance of 2, 3, and 4. $P(\text{Assign as Mob})$ is the probability an individual was assigned as a mobilizer. All models include a term for the expected number of mobilizers given the randomization scheme. All models use Huber-White HC errors.

Chapter 6

Coordination on Candidates Using Social Information

Do political actors use social information to coordinate their actions? If so, how? Across contexts as diverse as bucolic Honduras, the US Congress, the *Ewe* tribe of fishers in Ghana, city councils in Michigan, and the Republican Primary election in the US, humans take political action in a context where the people selecting those for leadership have ongoing, direct contact with those being selected for the leadership role. In spite of the fact that such a large number of political actions are taken in this context, very little is known about the rich set of social information and social histories might shape outcomes.

In this chapter, I argue that *social information* plays an important role in shaping political outcomes. Social information deals with who knows whom and who dislikes whom; where the cleavages between group members exist; sets of actors who frequently take the same action; and other queues easily collected and processed in routine, daily interactions. This social information is commonly shared by all members when the groups are relatively small and interact frequently. Indeed, in many decisions made by small groups of actors, social information is the sole piece common knowledge.

Preview of Results To test whether political actors use social information in their decisions, I construct a novel data-set that collects political, demographic, and social data from more than 4,000 residents of rural Honduras. With this data in hand, I partnered with a burgeoning social development and micro-finance non-profit in the region to randomly assign 113 of these 4,000 to stand for election. In all, the

elections placed representatives in 32 year-long governance positions in with the non-profit organization. Although many of the leadership positions have since turned over, the non-profit continues to work actively in the area.

I find that better socially connected candidates did better in this election, even when controlling for observable quality indicators. Specifically, well connected actors receive fifty percent more votes than poorly connected actors, and actors who bridge clusters of actors in a community receive a nearly 25% boost in vote share over and above the effect of being well connected. In addition, this research design allows me to examine the social information mechanism voters use to choose candidates. The results are consistent with a theory that voters use social information to coordinate group-based actions.

6.1 Overall Theory in this Chapter

6.2 Examples in Social Voting in Formal US Politics

Already, at press-time for this chapter (early March, 2016) the Republican presidential primary election has been a uniquely interesting political event. Even when compared to the sensational rise of the populist, tea-party Republicans in the 2010 and 2012 election cycles, and the first case of a potentially viable progressive socialist candidate in Bernie Sanders on the Democratic side, the Republican primary has been remarkably newsworthy. Driving the news coverage is the presence of controversial hopeful Donald Trump. Magnifying the coverage is the inability of a motivated Republican National Committee, despite its clearest desires, to organize primary voters behind a viable alternative to the Trump primary campaign.

But, what is the relationship between elections for representation in small-scale societies where individuals know one another and the monster, impersonal political machine that is the US national electoral politics? If the Republican should fail to have any hopeful presidential candidates reach the threshold necessary to secure the nomination in the first vote then by RNC rules, the next round of voting would release some delegates from their delegate voting requirements and allow them to vote freely. Voting rounds subsequent to the first re-vote would liberate increasing proportions of delegates from dedicated responsibilities.

In the event that delegates are not bound to cast votes for the presidential

hopeful elected by the primary election or caucus, how then should they decide upon a hopeful? The convention rules sequentially relax the restrictions on delegate binding; an unbound delegate interested in casting a vote for a winning candidate has a strong incentive to coordinate her actions with other delegates as quickly as possible.

In the event that delegates caucus to discuss whom they would like to elect, how will they choose with whom to talk and for whom to vote? For the former consideration, a reasonable *beginning* of where delegates will discuss their votes is among other delegates with whom they share a social connection — a common alma mater, a common state, or a common past electoral experience. And, since conversation is costly (in terms of time), it is not likely that any one delegate will converse with the universe of other delegates to poll these individuals, further increasing the drive for delegates to find relatively low-cost conversations to have.

An additional consideration — beyond simply conversing with other delegates — is developing a sense for the direction the other delegates are leaning. Understanding the intention of the other delegates is a dynamic process that closely resembles the discussion and voting process envisioned in Rolfe (2012). Some delegates, once de-committed, will hold a strong preference for one candidate over another. For these delegates, there is no action by others that would lead them to change their vote. Other delegates, however, may not hold such steadfast conviction about one or another of the presidential candidate hopefuls. For these candidates, being a part of the group of delegate who selected the eventually winning candidate might be the primary, salient decision criteria. Indeed, being a part of the delegate corps that selected the eventually winning hopeful to serve as the party nomination for the Presidential ticket, if that candidate should eventually win Office, would almost certainly be a desirable "feather in the cap" of a delegate and in some circumstances might even lead to tangible benefits to the delegate in terms of future positions within the delegation, or in another state or national office.

6.3 Hypotheses

A theory of social information leads to the following set of hypotheses:

1. Outcomes:

- (a) *Candidate Capability Hypothesis*: More capable candidates will win more

votes than less qualified candidates;

- (b) *Candidate Connectedness Hypothesis*: More connected candidates will win more votes than less well connected candidates.

2. Mechanisms:

- (a) *Social Proximity Mechanism*: Voters select candidates who are socially proximate;
- (b) *Social Connectedness Mechanism*: Voters select candidates who are socially connected.

6.3.1 Capability Hypothesis

Voters who care about outcomes are more likely to vote for well qualified candidates than poorly qualified candidates; so, under candidate capacity considerations well-qualified candidates will win more votes than less qualified candidates. This effect operates independent of any social information, though quality information and social information can operate in concert with each other. Indeed, many of the traits that might be associated with a quality politician are also associated with a quality friends – empathy, motivation, capacity.

There are a number of reasons why one candidate may be more qualified than another. One candidate may be more capable of taking on the cognitive load associated with leadership and complex decision making. In this case, increased education which signals a capacity for increased load, will be associated with increased vote share. Additionally, one candidate may be better qualified because he is more familiar with unique constituent issues. In this case, living in the community longer and being older should be associated with increased vote share. Finally, one candidate may be more empathetic and willing to address the problems of constituents and find compromise where possible. In this case, it possible that personal, social lives are a useful signal of the ability of a candidate to take on responsibility. Then, being married may signal positive responsibility while being divorced may signal a lack of responsibility.

6.3.2 Candidate Connectedness Hypothesis

In addition to capability considerations, candidates with social networks that have desirable traits may perform better in an election than candidates who do

not have those same desirable traits. In particular, candidates might receive votes from their local-neighborhood of social contacts – described in the next subsection as the social proximity mechanism – or might receive votes from a more distributed set of electors – described in the next subsection as the connectedness mechanism.

6.3.3 Social Proximity Mechanism

The main competing theories are the social proximity theories of Sinclair and Rolfe. Instead of using social networks to coordinate behavior, actors may just use their social networks to find the candidate with whom they are the closest friends. Choosing the most socially proximate candidate satisfies Sinclair’s conflict avoidance mechanism and Rolfe’s conditional cooperation model of behavior. Figure 6.1 presents one realization of voting patterns under the social distance theory. Under this theory, rather than evaluating how well-connected is a candidate, instead, voters make a simpler calculation about which of the candidates is the most proximate and cast a vote for that candidate. Voters make this assessment with some inaccuracy. Because of this inaccuracy, there is some drift between the conceptualization of the social network and its measurement. These factors lead to some mis-matched voting, where a voter votes for a candidate who is not measured as the most socially proximate, but the social distance pattern is still clearly observable.

6.3.4 Connectedness Mechanism

To elect a candidate, voters must coordinate to vote in a similar way. One way to coordinate is for a voters to evaluate the connectedness of candidates. With this assessment, voters can form mutually consistent beliefs about whom other voters will choose—the most connected. Socially connected actors are in active in their communities and are in frequent contact with other residents. They tend to participate in community, church and work organizations, and in these groups they strengthen relationships with other community members. The strength of these relationships, and the frequent interaction with other members of the community as a result of these relationship make actions of well-connected actors increasingly visible within the community.

Acemoglu and Jackson (2015) demonstrate that under conditions of imperfect observation – an assumption that appropriately matches social systems – the actions

of “prominent” actors can shape future behaviors because all future actors are more likely to have received the signal from the prominent actor than other actors. While Acemoglu and Jackson stipulate that prominent actors’ signals are perfectly received by alter, I relax the perfect observability requirement, and assume only that better connected individuals actions are increasingly likely to be observed in the number of social connections she holds.

This theory of social information does not imply persuasion, obedience, or any other type of social influence as a mechanism. Rather, the social prominence hypothesis is a statement of common beliefs. Each political actor holds beliefs about the probable actions of other actors; and, conditional on the randomly assigned candidate set, the actor updates her beliefs about the actions of others.

6.3.5 No Social Information Comparison

The recent finding that friends share more genetic material than non-friends – about as much as fourth cousins – makes it tempting to argue that coordinating on socially proximate, high-visibility actors might be an evolutionarily beneficial strategy. This line of reasoning would argue that many decisions are made in conditions of scarce resources, and so timely action is of paramount importance. When under siege, it would be beneficial for all members of a group to move in the same direction, and quickly, rather than either (a) not move for lack of consensus, or (b) move in all directions for lack of consensus. In the first case, they hyenas eat everyone who is sitting still, and in the second case the hyenas pick people off one-by-one as they run pell-mell away. Instead, if the group moves together, they are afforded some protection. This argument basically says that there are social leaders whose actions are more heard more readily than others, but that this position of social leadership is granted by the people who surround him or her.

6.4 Research Design

6.4.1 Reprise: Key Network Concepts

In this subsection, I refresh the concepts described in Table 7.1. One of the assets of approaching a political science problem from a social networks framework is that researchers and readers have a strong intuitive understanding of social

network concepts from their daily lives. Rather than learning about concepts like sampling variability, variance and heteroskedasticity that one typically does not encounter outside of formal training, to understand the argument and evidence in this paper, those who do not have previous experience with the method need only learn two relatively straight forward concepts: connectedness and social distance.

In this paper, I conceptualize social connectedness as *how central* an individual is to her social network. Well-connected, or central, actors are actors that are thought to be important to the network. In a high school, these actors might be the students who are elected to be the prom queen and king; in a business, central actors might be the woman who knows how to quickly address technical problems or the man who knows how to navigate a bureaucracy; in a politicalscience sub-discipline this might be the scholar who everyone wants to talk with at the discipline's annual conference. Freeman defines three core concepts of social connectedness: degree, closeness, and betweenness. In this chapter, I utilize only degree and betweenness, and so I only discuss the operationalization of these two measures.¹

Degree Centrality Degree centrality is a count of the social connections an actor holds. If someone has three friends, we would say that this person's friendship social network has a degree of three; if someone has a single sister, that person's sibling network has a degree of one. That is, degree is the number of relationships, frequently called edges, incident on a particular actor. One further distinction in degree centrality lay in the direction of a tie. Frequently, social relationships are measured as beginning with one person directed toward another. I might call someone a friend, in which case we would say that this social relationship originates with me, and is directed toward my friend. My friend, in turn, might also say that I am her friend, a tie beginning with her and directed toward me. In this case, one would say there exist two relationships – one in each direction, between us. In this paper I make one further distinction about degree centrality based on the direction of a tie. I use indegree – a count of the edges incident on an actor – or the number of others that nominate an actor as a particular social relationship. Canonically, (Wasserman and Faust, 1994, p. 164) define indegree (D_I) for a particular actor (e_i) as,

¹Like both degree and betweenness centrality, closeness centrality also measure the distance between a focal actor and other actors in the network. However, unlike degree and betweenness, closeness is ill-defined in networks that are not *fully-connected* – when there is an actor or group of actors that does not hold any social relationships with the rest of the social network.

$$D_I(e_i) = \sum_{j=1}^g x_{ji}, \quad (6.1)$$

where $x_{ji} = 1$ if there is a social connection from node j to node i and zero otherwise. Indegree has several desirable properties as a measurement of social connectedness, but two worth particular mention for this task. First, indegree increases in the number of social connections an actor has, but does not increase in the number of outgoing ties. For networks that are built from personal interviews, this limits mis-measurement that would result from personality traits that might be correlated with social network indicators, particularly gregariousness and intelligence. Second, indegree is a locally consistent centrality measurement; if there are nodes or edges missing from the social network measurement, as might exist in sampled graphs, indegree still predictably ranks actors without any order reversals (Yoon et al., 2007; Illenberger and Flötteröd, 2012). I use degree centrality to operationalize hypothesis two that better socially connected actors will perform better in the election.

Betweenness Centrality Betweenness centrality measures how much a particular actor is a necessary link between two others. Intuitively, “interactions between two nonadjacent actors might depend on other actors... especially [those] who lie on paths between the two,” (Wasserman and Faust, 1994, p. 188). In a university, graduate students would hold high-scoring betweenness position between faculty members and undergraduates because ideally all communication about grade appeals filters through the graduate students. On the Hill, Congressional schedulers hold positions of high betweenness centrality because information about a Member of Congress’ face-to-face meetings between staffers, other Members of Congress and outside interests must all be coordinated by the scheduler.

Described formally, the betweenness centrality (B) of a particular actor (e_i) is,

$$B(e_i) = \sum \frac{\sigma_{uv}(e_i)}{\sigma_{uv}}, \quad (6.2)$$

where $B(e_i)$ is the betweenness of node e_i , $\sigma_{uv}(e_i)$ is the number of shortest paths (using social network connections) between nodes u and v that cross node e_i , and σ_{uv} is the total number of shortest paths between u and v . A shortest path between two nodes is the path that traverses known social connections without

“wandering off.” The theoretical maximum betweenness of any node is $(g - 1)(g - 2)/2$, where g is the total number of nodes in the network. Then, graphs can be scaled by this theoretical maximum, projecting betweenness onto a $[0,1]$ range by dividing the betweenness score of each node by the theoretical maximum, noted $B'(e)$ is,

$$B'(e) = \frac{B(e)}{(g - 1)(g - 2)/2}. \quad (6.3)$$

The rescaling permits the comparison of betweenness centrality in networks of different sizes. I use betweenness centrality to operationalize hypothesis three that actors who bridge factions will perform better in the election.

6.5 Experiment Design

To test how social characteristics effect vote choice, I randomly selected 115 candidates to be elected to 30 positions as an officer to a microfinance organization. These officers were to be initial contacts between the microfinance organization and they people who lived in each city. As such, these candidates had real influence over who the firm would eventually work with, and therefore real influence over the distribution of resources. Practically, the elections took place at town meetings on the weekend, and were well attended. In the rest of this section, I describe how candidates were selected and provide attributes of those candidates that were drawn from the population; I provide the text of the election and further describe the incentives at play; and I describe how votes were recorded.

6.5.1 Selecting Candidates

In each town we selected, at random, between 3 and 6 individuals to serve as “candidates” for an election to be held at a town meeting in the following days. The number of candidates randomly selected was strictly a function of town size. In towns that were smaller than 50 residents, we selected 3 candidates; in towns between 50 and 500 we selected 4 or 5 candidates, and in towns larger than 500 we selected 6 candidates. We did not inform candidates that they would be placed on the ballot, and so there was no opportunity for candidates to campaign, pledge, or promise benefits to constituents. By disallowing candidate statements

and campaigning, we held constant campaign effectiveness and campaign effort; in doing so we are able to isolate the independent effect of social connections without the contaminating effect of selection into campaign effort, which through positive/pro-active selection would likely be positively correlated with candidate skill and/or probability of election. An hour before the election, we informed the candidates of the election that we were holding, the costs and benefits of being elected to the post, and also the reality that any candidate could drop out at any time. Because the costs to serving in this position quite small – a single organization meeting and follow-up to take place in the candidates’ home villages – none of the randomly drawn candidates opted to withdraw their names from consideration.

We held all elections as a part of a town-meeting held on the weekends when the greatest number of town residents were likely to be available to meet. Meetings were scheduled in one of three times: Saturday mid-morning, Saturday mid-afternoon, and Sunday mid-afternoon. Mean turnout at these meetings was just over fifty percent of town residents; most villages fell between twenty-five percent turnout and seventy-five percent turnout. Figure 6.3 displays the distribution of turnout across the villages.

6.5.2 Candidate Attributes

One important characteristic distinguishes candidates from non candidates – candidates for election are uniformly male. After extensive interviews with local leadership in the county seat, the church, leadership at the village level, and male and female residents of the areas it became clear that local tradition required that nominations for a position of this sort required a male candidate. Female residents in the region are well integrated into the society – they own shops, restaurants, and business at commensurate rates as to male residents; they hold positions of leadership in schools and churches at commensurate rates; females hold wealth stores that are likely indistinguishable from the wealth stores of men. However, for this position local custom dictated that in forms of negotiation formally related to dealing with a bank-like entity, that the position be filled by a male candidate.

On characteristics that are not gender, candidates look very similar to non-candidates. Both groups have similar educational histories, are married at similar rates, are employed in similar jobs, and have lived in town for similar amounts of time. Table 6.1 and Table 6.2 show the observables for both candidate and

non-candidate residents.

6.5.3 Election Text

Subjects were presented with the following instructions when they were asked to make a candidate selection:

A microfinance organization will soon be starting to work in this area. Microfinance organizations are groups that make low-interest loans to members of the community so that these members can make investments in capital to start or improve a business. Because we do not know everyone in town, we want to select a representative from the town that can help us with introductions and setting up this firm. We have selected a small number of candidates who live in this town to represent you in these first meetings. We were not able to talk with everyone before selecting candidates, so the person that you would MOST LIKE to be a representative might not be a candidate. If this is the case, then please select the candidate who you think would BEST REPRESENT you.

In this way, the election field experiment fairly recreates a number of dynamics at play in municipal, regional and national elections. First, electoral candidates were selected through an external selection mechanism. In the United States this selection mechanism depends on the state but typically involves the coordination of party elites. In consolidating democracies, this mechanism is frequently local and national party machines. Second, electoral candidates have clear influence over the distribution of resources. In “real” elections, these resources are manifest in terms of appropriations, pork-projects or other constituent services. In this election, electors were informed that the microfinance organization planned to utilize the local elected official as a point of contact for the distribution of capital-improving resources. Third, in “real” elections with a secret ballot, there electors have scant ability to credibly claim they voted for one candidate over another. This inability to clearly demonstrate a vote is present in elections in consolidating democracies as well as consolidated democracies. Fourth, in “real” elections where the issue space is relatively dense – when candidates are not able to run, and win, on a single issue – there are not typically clear policy mandates conferred through the electoral process. In this very local election, the same truth held; a vote for one candidate over another transmitted precious little programmatic information from the electorate to the finally selected candidate.

Table 6.1: Candidate Characteristics

Statistic	N	Mean	St. Dev.	Min	Max
Age	111	43.946	14.265	18	94
Years Education	82	4.000	2.244	1	12
Monthly Income	60	1,770.500	1,482.270	280	8,000
Eigenvector Centrality (All)	113	0.102	0.097	0.000	0.388
Indegree (Friends)	113	3.372	3.709	0	26
Outdegree (Friends)	113	2.354	1.231	0	7
Indegree (All)	113	5.788	4.345	0	31
Outdegree (All)	113	4.885	2.219	0	12
Betweenness (All)	113	922.021	1,156.320	0.000	6,501.134
Scaled Betweenness (All)	113	0.018	0.022	0.000	0.133

Table 6.2: Non Candidate Characteristics

Statistic	N	Mean	St. Dev.	Min	Max
Age	1,701	36.240	16.703	5	145
Years Education	1,331	4.184	2.319	0.000	12.000
Monthly Income	1,014	1,581.764	1,063.349	0	10,000
Eigenvector Centrality (All)	1,706	0.058	0.071	0.000	0.444
Indegree (Friends)	1,706	1.808	2.305	0	19
Outdegree (Friends)	1,706	2.019	1.452	0	14
Indegree (All)	1,706	3.842	2.946	0	23
Outdegree (All)	1,706	4.267	2.400	0	17
Betweenness (All)	1,706	863.884	1,325.787	0.000	11,628.640
Scaled Betweenness (All)	1,706	0.008	0.012	0.000	0.157

6.5.4 Voting Measurement

After presenting the justification for why we were holding an election to the residents of the village, we provided residents with a short “ballot” that included the names and photographs of the men randomly selected as electoral candidates. Each resident then individually recorded their vote choice in a computer system; residents who were unfamiliar with computers were assisted by survey enumerators. Electors were informed that they need not vote if they had no preference over the candidates. The computer system was designed to ensure the correct identification of both the candidates and also the voters; in this way, the data can be keyed and merged back onto other demographic features captured earlier in the study.

6.6 Candidate Results

I analyze this experiment at two levels. First I analyze how social connections shape outcomes at the candidate-centered level. Here, I use models that estimate the relationship between candidate characteristics and voteshare. Then, to draw inference about the mechanism at work in this election, I estimate models at the voter-candidate dyad level. Across the two sets of analysis, I find strong evidence in support of socially coordinated behavior. At the candidate-centered level, candidates who have higher indegree receive significantly, and substantially more votes than candidates with lower indegree, even controlling for candidate skill characteristics. Moreover, controlling for indegree, candidates who bridge groups in the social network also garner more votes, controlling for all other characteristics. I also find evidence that voters prefer candidates who are closer relationships. When I analyze the data at the voter-candidate dyad level, I find more limited evidence in support of the theory. While I find that voters cast votes for better connected candidates, somewhat puzzlingly I find that at the voter-candidate level there is a negative relationship between candidate betweenness and probability of a voter casting a vote for that candidate. I conclude this section by providing my thoughts on what might be driving this correlation.

6.6.1 Unit of Analysis: Candidate

The elections produced considerable variation in results. In some elections, a single candidate ran away with a super-majority of the votes, leaving the other three,

four, or five candidates with less than ten percent of the vote-share between. In other elections, the distribution of votes was plurally distributed and the candidates split the vote total evenly. Figure 6.6 displays the empirical distribution of votes that candidates received. A relatively large number of candidates received zero votes. This is most likely caused by the random selection of candidates from the town population. While every candidate that was drawn lived and worked in each town, and was in principle qualified to hold the position they were standing for election for, it is evident that a large number of the candidates were viewed by the town as low-quality candidates. In contrast, there were several candidates who swept the election returns and garnered a majority, or super-majority of the votes cast. Five candidates received greater than 80% of the votes cast in the election, signaling either that these candidates were exceptionally well-qualified for the position or that the other candidates drawn to run were exceptionally poorly qualified to run.

At the candidate level, two results emerge that are consistent with the capabilities and connectedness hypotheses presented in section 6.3. First, consistent with the *capability hypothesis*, candidates who were better educated, and reported having taken political action in the past (voting) were significantly more likely to receive votes. Second, consistent with the *connectedness hypothesis* candidates who have more social contacts were significantly more likely to receive votes.

6.6.2 Capability Hypothesis

As is reported in Table 6.4, candidates with greater education and who had previously taken political action were significantly more likely to receive votes. These education and political cues are readily available, easily accessible heuristics that voters in the elections could use to identify the capability of the candidates. In addition, candidates who reported being in a married won significantly more votes. One possibility that these candidates won more votes is that they are of a type that is responsible, and this responsibility is known among the community. Another possibility is that these candidates received votes from their spouses. While the test does not settle the issue dispositively, the evidence in Table 6.4 hues toward additional votes coming from the candidates' spouse. The magnitude of the effect is roughly equivalent to a single additional vote for being married and, if it were a signal of quality that was being transmitted through being married, candidates who were widowers would also receive this quality bonus. They do not – the impact

of being a widower is indistinguishable from being single. The results that higher quality candidates – those with higher education and those who have taken past political action – receive more votes is hardly surprising. This result does, however, provide a *prima facie* check that considerations expected to be at play in an election of this type, are indeed being considered by voters.

6.6.3 Connectedness Hypothesis

Before presenting a fully specified model, I present the simple bivariate relationship between the connectedness and voteshare to demonstrate that there is a plausible relationship between the two *on their face*. Indeed, as Figure 6.8 demonstrates, there is a strong positive relationship between the number of times an individual is called a social relationship and the number of votes that he receives.

6.6.4 Vote Share and Degree

To assess the strength of the relationship between indegree and voteshare, I estimate two bivariate models, one a Poisson model to predict the count of votes received by a candidate, and the other an OLS to estimate the percentage of votes received by a candidate. The results of these model are shown in Table 6.3. The OLS model is easier to immediately interpret, and shows that in this model for each additional friendship nomination a candidate received that candidate is predicted to receive an additional 2.1% of the votes in the town. Figure 6.10 shows the predicted effects of a change in the number of social connections.

6.6.5 Vote Share and Betweenness

The second hypothesis derived from a social theory of voting is that voters select candidates that bridge groups of voters in the social network. As I describe earlier, I operationalize this bridging using betweenness centrality – which just measures how many times a particular node is between two other nodes. To do so, I estimate model three in Table 6.4 which includes a measure of candidate betweenness centrality. If voters are selecting candidates for their connectedness and also how well they bring groups of voters together, I would expect that both model terms would be positive and significant. However, these results suggest a negative relationship between betweenness centrality and candidate performance. This is

Table 6.3: Votes and vote share won by candidates as a function of the candidates' extitdegree calculated two ways: among all social contacts, and among only social contacts who are friends.

	<i>Dependent variable:</i>			
	Votes Won		Vote share	
	<i>Poisson</i>		<i>OLS</i>	
	(1)	(2)	(3)	(4)
Indegree (All)	0.084*** (0.004)		0.021*** (0.005)	
Indegree (Friends)		0.090*** (0.004)		0.025*** (0.005)
Constant	1.812*** (0.044)	2.005*** (0.038)	0.126*** (0.033)	0.164*** (0.027)
Observations	113	113	111	111
R ²			0.171	0.175
Adjusted R ²			0.163	0.168
Log Likelihood	-614.464	-633.079		
Akaike Inf. Crit.	1,232.928	1,270.157		
Residual Std. Error (df = 109)			0.207	0.207
F Statistic (df = 1; 109)			22.467***	23.151***

Note:

*p<0.1; **p<0.05; ***p<0.01

Table 6.4: Poisson model estimating factors associated with candidates winning more votes in each election.

	<i>Dependent variable:</i>					
	Votes Won			election		
	<i>OLS</i>			<i>Poisson</i>		
	(1)	(2)	(3)	(4)	(5)	(6)
Indegree	0.021*** (0.005)	0.084*** (0.004)	0.153*** (0.010)		0.077*** (0.005)	0.107*** (0.012)
Betweenness			−12.040*** (2.685)			−7.101** (3.354)
Age				0.0003 (0.001)	0.0002 (0.001)	0.0002 (0.001)
Years Edu				0.046*** (0.014)	0.064*** (0.016)	0.140*** (0.024)
Married				1.524*** (0.246)	1.089*** (0.251)	0.935*** (0.291)
Divorced				0.550* (0.309)	0.343 (0.311)	0.126 (0.341)
Widowed				0.688** (0.301)	0.466 (0.303)	0.161 (0.343)
Voted				−0.023 (0.083)	0.221** (0.091)	−0.028 (0.123)
Constant	0.126*** (0.033)	1.812*** (0.044)	1.761*** (0.173)	0.942*** (0.268)	0.500* (0.275)	0.974*** (0.370)
Town Fixed Effects	No	No	No	No	No	Yes
Observations	111	113	113	82	82	82
R ²	0.171					
Adjusted R ²	0.163					
Log Likelihood		−614.464	−486.205	−544.857	−436.125	−317.798
Akaike Inf. Crit.		1,232.928	1,040.410	1,103.715	888.249	715.596

Note:

*p<0.1; **p<0.05; ***p<0.01
Model 6 includes town-fixed effects.

most likely due to the correlation structure between indegree and betweenness centrality.

To interpret these poisson coefficients, Figure 6.12 displays the predicted vote share a candidate receives as a function of changing network characteristics (IVs) from the mean minus one standard deviation to mean plus one standard deviation. Holding betweenness constant, and changing only the number of incoming social ties a candidate has from low to high leads to a predicted change in the number of votes from 8 to 12, or an increase of about 15% of the voteshare on average. Holding indegree constant and instead manipulating betweenness, a change in the IV from the twentieth-fifth percentile of the variable's distribution to the seventy-fifth earns a candidate about two more votes, or about 8% of the total vote share in the election.

Indeed, Figure 6.12 shows just this relationship. Betweenness is distributed over a much larger range than indegree, as shown in Table 2, candidates indegree ranges only from zero to thirty-one, while betweenness centrality ranges from 0 to over one thousand.

6.6.6 Vote Share Complete Model

Finally, in Table 6.4 Model (5), I include social and demographic characteristics of candidates. Including age and education covariates in this model increases the precision of the estimate of the causal effect, reducing the estimated standard error of the *Indegree* coefficient by a factor of two. Estimate of the effect of network characteristics do not change including these covariates. The age of a candidate has no measurable impact on the number of votes a candidate receives, and is in fact a precisely estimated zero. Education increases the number of votes a candidate receives. Each additional year of education has an impact that is similar in magnitude to the impact of additional social connections. Being married increases the number of votes a candidate receives (presumably because one's partner votes) against a baseline of being widowed, while being single decreases the number of votes a candidate receives. The marginal effects of these estimated parameters in Table 6.4 are presented graphically in Figure 6.14.

6.7 Candidate-Elector Dyad Results

6.7.1 Social Proximity

The main competing theory is that of Sinclair (2012) who argues voters just cast votes for the candidate with whom they are the most socially connected, rather than using social information to coordinate actions. By Sinclair's conflict avoidance mechanism, voters are the least inclined to be in conflict with their closest social connections, and would therefore be most likely to vote for these individuals. A social proximity explanation also fits well with components of Rolfe's conditional cooperation argument because taking action for the more socially proximate candidates generates the small clusters of action around *unconditional* cooperators who are first-movers for larger political action. Then, if social proximity is the leading factor that shapes peoples' votes, previous theories more accurately characterize political behavior, taking back support from the theory of social information presented in this paper.

In fact, there is a complex relationship between social distance and the vote choice. Figure 6.15 shows the empirical distribution of votes as a function of social distance. Surprisingly, only thirty percent (12 of 40) of people who were assigned to be candidates voted for themselves; seventy percent cast a vote for another candidate. This low rate might mean there was some trepidation on the part of candidates who may have been uncertain about the requirements for the position. Alternatively, this low vote-rate might indicate that a relatively large number of candidates were skeptical of their chances to win election, and cast a vote for some other candidate. Of voters who had a direct social alter assigned as a candidate, forty-eight percent (154 of 320) cast a vote for that candidate. As voters and candidates become further separated — friends of friends, friends of friends' friends — voters are decreasingly likely to cast a vote for a candidate.

The top panel of Figure 6.16 presents the distribution of social distances between voters and the candidates for whom they voted. The bottom panel of Figure 6.16 and the candidates for whom they did not vote. On average, in these 30 communities, voters are about 3 social connections separated from the candidates they vote for and about 3.8 social connections separated from the candidates they did not vote for. Because the assumption of independent samples is untenable in this case, I rely on randomization for inference to assess how likely this difference is to be caused by chance.

For this randomization inference regime, I hold the social networks and candidates constant, but repeatedly simulate random votes. This generates a hypothetical distribution of geodesic distances under the null hypothesis that there is no relationship between social distance and voting for a particular candidate. If this simulation of random voting returned an average social distance between candidates and voters that was demonstrably larger than observed in this field experiment—say, the random distribution found that the mean distance was 3.5 geodesics—it would suggest that there is positive vote-selection based on social proximity. This would mean that voters more frequently cast votes for candidates with whom they held stronger relationships. This is not the case.

Under the null-hypothesis simulation, the 90% CI ($\alpha = 0.10$) for the mean geodesic distance between a voter and the candidate she votes for is $3.00 \pm 0.05 = [2.95, 3.10]$. The 95% CI ($\alpha = 0.05$) is $3.00 \pm 0.07 = [2.93, 3.11]$. These estimates of social distance under a random voting rule are precisely the same as the estimate observed in the election, and suggest that a mean social distance of 3 between candidates and voters is *exceedingly* likely to be caused by random chance. Therefore, there is little evidence from this trial that social distance is the mechanism that led some candidates to do better than others in the election.

6.7.2 Social Information

To further examine whether social proximity outweighs the coordination incentive, in Table 6.5 I present a regressions that predict whether a voter cast a vote for a candidate. All models are fit on identical data – a “tall” dataset organized at the voter-candidate dyad level with 4362 dyad observations. While the point estimates for the relationships between these RHS variables and the voters’ choice in the election can be reliably estimated without bias in this framework, the nature of the repeated observations, if simply calculated, would downward bias the standard error estimates and increase the probability of detecting false-positives. To address this concern, following Ling and Zeger (1986) and Yan and Fine (2004), I correct for multiple observations of an individual voter’s decision by clustering the standard errors at the voter level.² The advantage of this method over other ML-type estimators is that the empirical covariance can be estimated rather than specified *ex ante*. Although this comes at an efficiency cost, the results of MCMC trials suggest this

²Practical execution is via the R package *geepack* (Yan and Fine, 2004).

penalty is slight (Schildcrout & Heagerty, 2005).

Model 1 includes only social distance as a predictor of the vote. Here, social distance is negatively associated with casting a vote or a candidate, consistent with the theories of Sinclair and Rolfe. In Models 2, 3, 4, and 5 I include centrality and betweenness characterizations. Including these model features decreases the estimated effect of social proximity by a factor of about 4.5, suggesting that it isn't social distance *per se* that leads to a vote, but rather that candidates who are better connected, tend to be more socially proximate to everyone.

The centrality indicators in Models 2, 3, 4, and 5 support a theory that central actors are more frequently voted on. Regardless of the conceptualization of centrality, and the operationalization of that concept, more central actors earn more votes, while social proximity plays no estimated role in the process.

6.8 Robustness

6.8.1 Connectedness is Not Only a Low-Quality Candidate Screen

The large number of candidates who received zero votes is evidence of how experimental control comes at the tradeoff of real-world mechanisms. By randomly selecting men to serve as candidates, we excluded the important real-world election process of clearing the field of unqualified candidates. In the American case, the party and primary elections clear the field; in contexts like Honduras, it is likely that social sanctioning or other informal mechanisms would have led to the selection of different candidates. Because the differences in this process are real, it limits the ability to make predictions *in the real world* about absolute vote shares as a function of social connections; however, it does not in any way limit the interpretation of the causal mechanism.

Social information, rather than helping voters positively coordinate their votes, may instead flag some candidates as poor quality candidates. That is, it may be that selecting candidates at random from meant that some candidates were presented to the voters who would never have otherwise been under consideration for this leadership role. In this argument, there are two reasons that low-quality candidates would not make it onto the ballot. First, low-quality candidates, knowing their type, would not select to run for the position. Because there is some cost associated with running for office – time, material, or social – if a potential candidate

knew that he had a zero probability chance of winning office, he would never incur that cost. Second, low-quality candidates would not be placed on the ballot because in typical elections, institutional factors clear the field of weak candidates before the balloting begins. While both of these mechanisms are in play during *in situ* elections, if I restrict the sample of candidates to include only those who meet some minimum quality threshold, and the analysis still identifies a role for social information, then one may still draw the inference that social information plays a role in candidate selection, even when field-clearing mechanisms remove deadbeat candidates.

To check this threat to inference, I restrict the sample to include only those candidates who receive more than zero votes and re-run the models presented in Table 6.4, and Table 6.5. This robustness check does not re-run the experiment, and so the results are not fully independent. However, continuing to increase the threshold for inclusion in the re-estimated models provides a series of increasingly stringent tests for an alternative hypothesis that the results are being driven selection of low-quality candidates that would not have otherwise be elected. After fitting these models, included in an appendix available on request, I perform the same prediction task as Figure 6.12. The results of this prediction find a consistent, identical relationship for thresholds up to a minimum of twenty-one votes: even among more restrictive samples, better connected candidates receive more votes.³ Beyond twenty-one votes, there become relatively few observations to fit the model and the standard-errors around the estimates increase.

6.8.2 Social Distance Does Not Confound Connectedness Results

One concern about the models fit in subsection 6.7.2 concerns the difficulty in simultaneously estimating the relationship between social connectedness and social distance. This concern arises because of the structural relationship between the two variables: that one candidate for office had relatively higher connectedness metrics is guaranteed to decrease that candidates' social distance from alters in the network. Indeed, in the case of closeness, the two concepts are fundamentally the inverse of one another – the canonical calculations of closeness centrality is the inverse of the average distance between and ego and her alters. The consequence of this relationship between RHS variables would be those of classic collinearity, and

³Plots demonstrating this relationship are available online at:
<http://polisci2.ucsd.edu/dhughes/robustnessSubset.pdf>.

would mean potentially failing to identify a relationship when in truth one exists.

To evaluate this possibility, in this section I employ a strategy in the style of matching analyses. For every voter in the dataset, I identify instances where the voter was randomly assigned candidates who were equally socially distant. In the simplest case, this would take the form of a mother (who is a voter) who lives in a town where we randomly assigned two of her sons to stand for election.⁴ Another example is a voter in a town where we randomly assigned two of her friends to stand for election. Importantly, although both of these examples are framed in terms of a social distance of one between ego and alter, the analysis is conducted across all matching pairs of candidate – extending up to 6 degrees of separation between the voter and nominated candidates.

The empirical prediction from a theory of social information and coordination is that among these matched pairs of candidates, the more highly socially connected candidate should be voted for more frequently. Because this method effectively eliminates social proximity as a consideration, the chief social consideration on voters minds can only be one of social connectedness.⁵ As a final benefit of this analysis, it is still possible to estimate the relationships between covariates and vote choice at the individual level.

Figure 6.17 plots the relationship between social distance, mean indegree of candidates and the vote among matched pairs of candidates. Because this is a complicated conditional statement, some care is due to explain the plot. At a social distance of zero, the voter is the candidate; there is no evidence that these voters' choice to vote for themselves has a social component. At every social distance from one to 6 the average indegree of candidates who received a vote is higher than the average indegree of candidates who did not receive a vote. The decrease in mean indegree as social distance increases is sensible – candidates who are socially distant from the voters must be relatively peripheral in the social network.

⁴In practice, this *particular* instance did not occur.

⁵This method has the added benefit of controlling for the possibility of homophily. In this framework, because all candidates are equally socially distant there is no possibility for selection on RHS variables to cause the observed relationship between vote choice and connectedness.

Votes in Social Distance Hypothesis

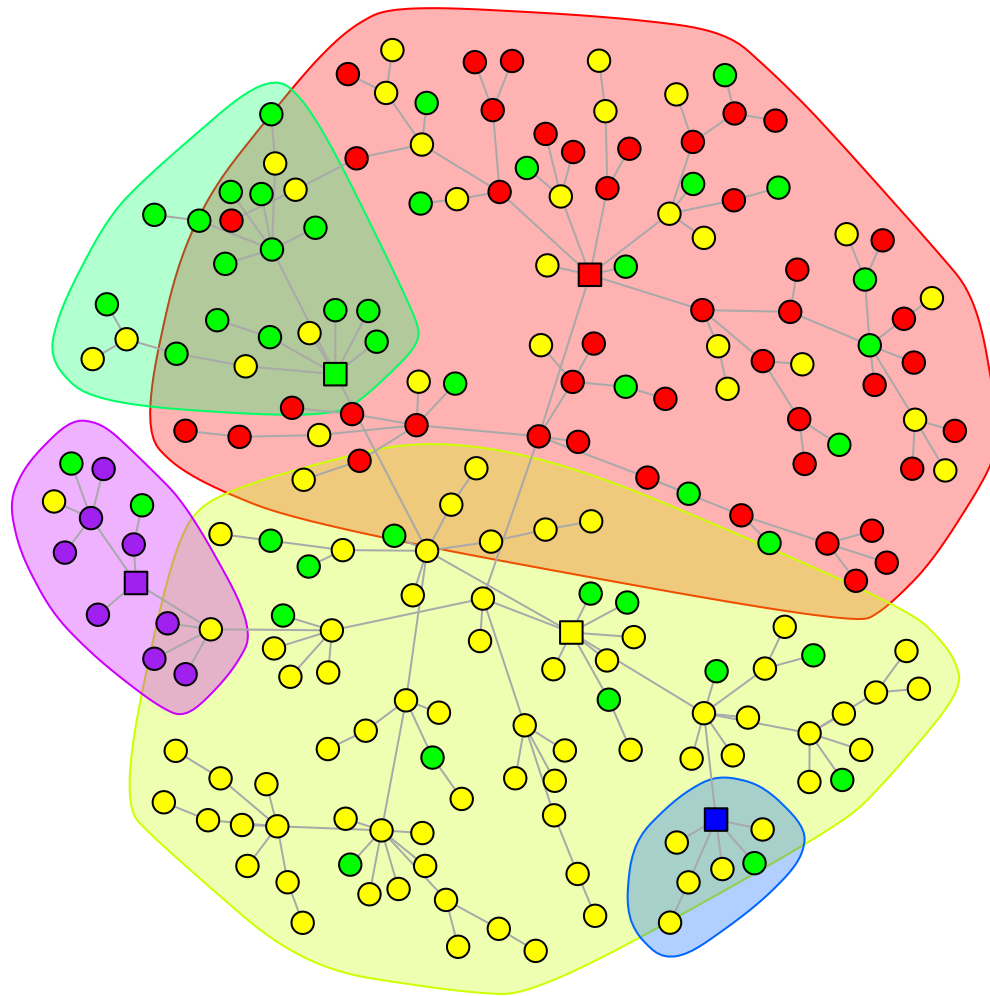


Figure 6.1: Possible voting pattern under social distance theory of voting. Square nodes represent randomly selected candidates, each assigned a unique color. Circular nodes represent voters, and grey edges between nodes represent social ties. Overlaid polygons are the *closest candidate* catchment areas, and the colors correspond to the candidates' assigned colors. Colors of circular nodes are votes for a candidate, and are a function of social distance and stochastic noise.

Connectedness Hypothesis

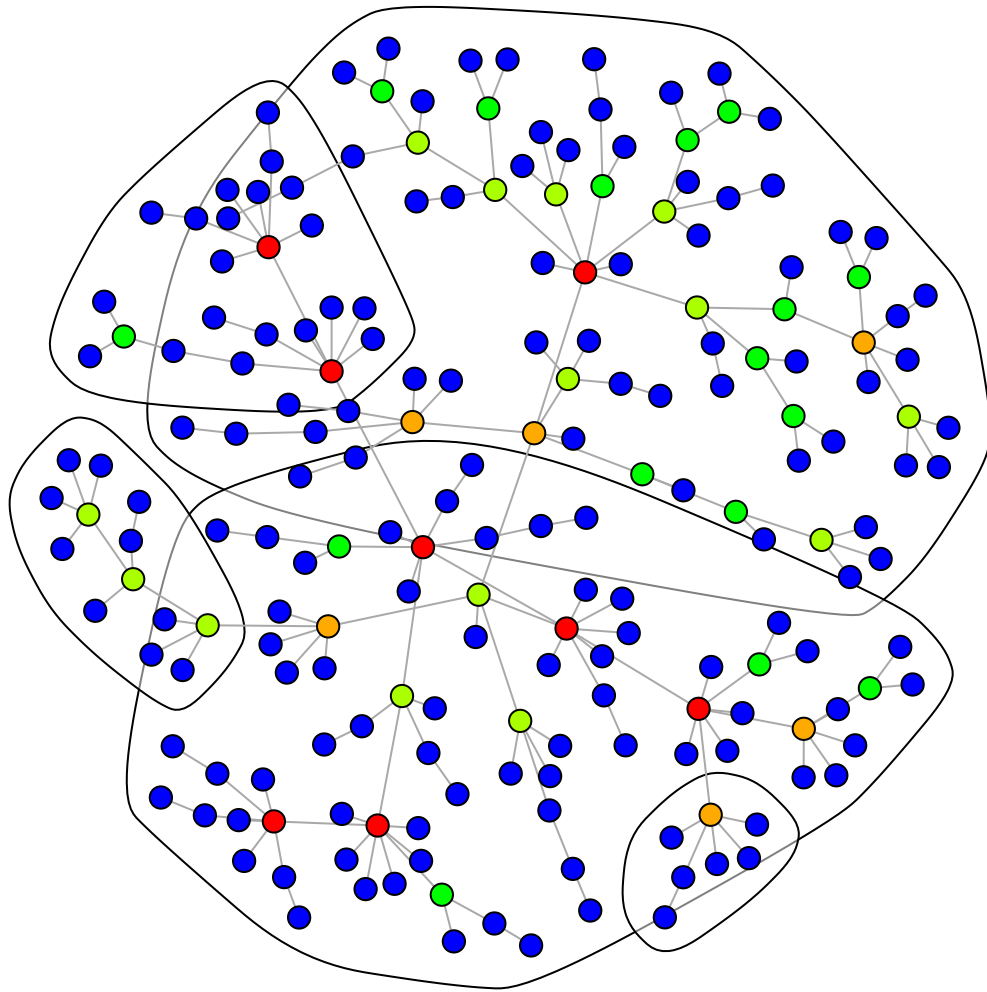


Figure 6.2: Possible voting pattern under social information theory of voting. Nodes represent individuals in the community and grey edges between nodes represent social ties. Overlaid loops are the same *closest candidate* catchment areas as Figure 6.1. Colors represent the probability of winning an election if voters vote based how well connected is a candidate, and range from red–orange–yellow–green–blue from most likely to win to least likely to win.

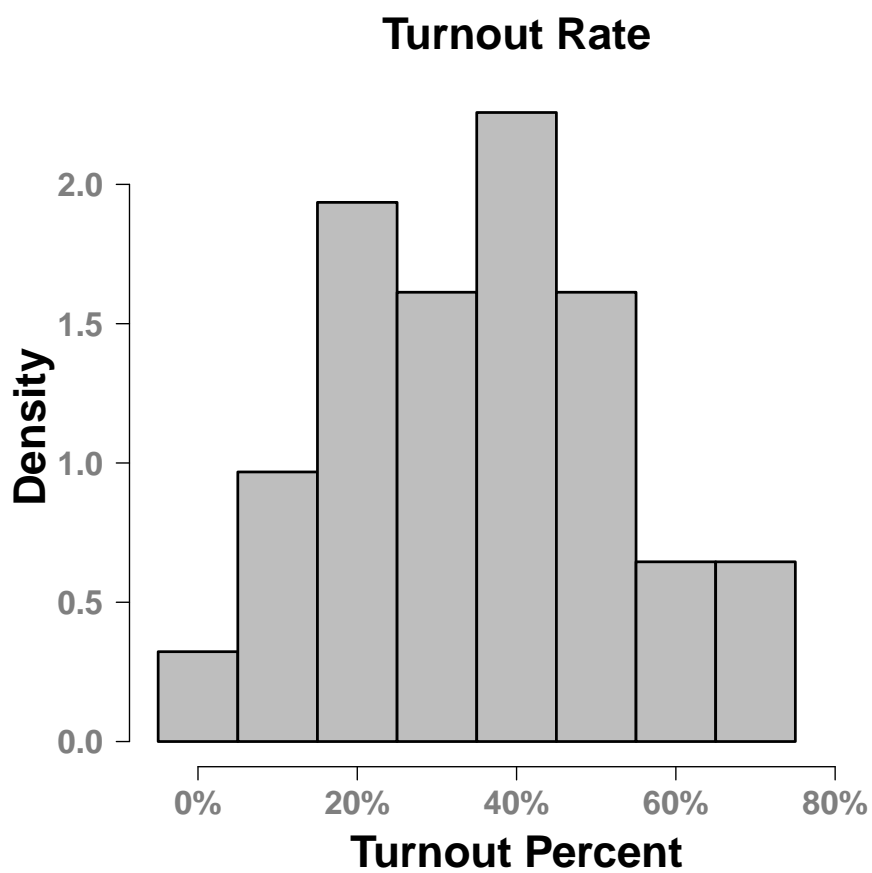


Figure 6.3: Turnout at village meeting, reported for every town. Turnout percentage is along x-axis.

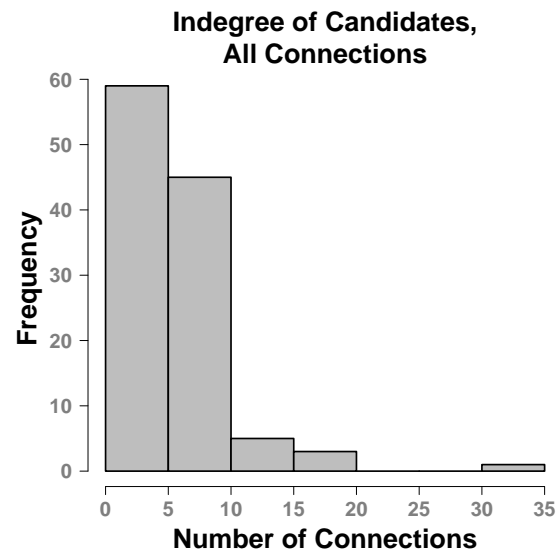


Figure 6.4: Histogram of indegree. The total number of times an individual was named as a social alter.

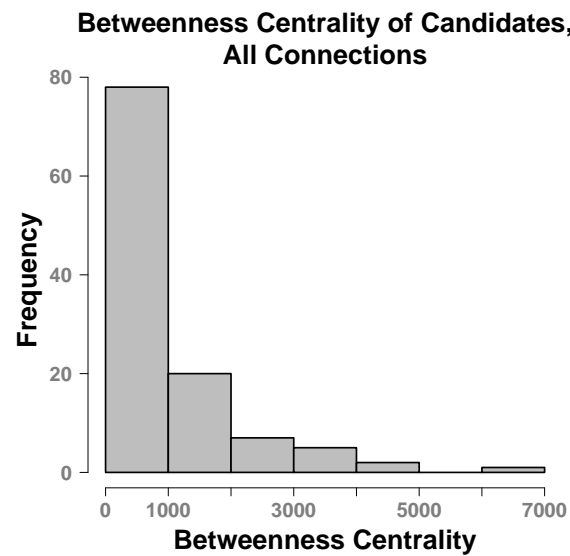


Figure 6.5: Histogram of Mobilizers' betweenness centrality, measured as non-standardized centrality. Histogram of betweenness.

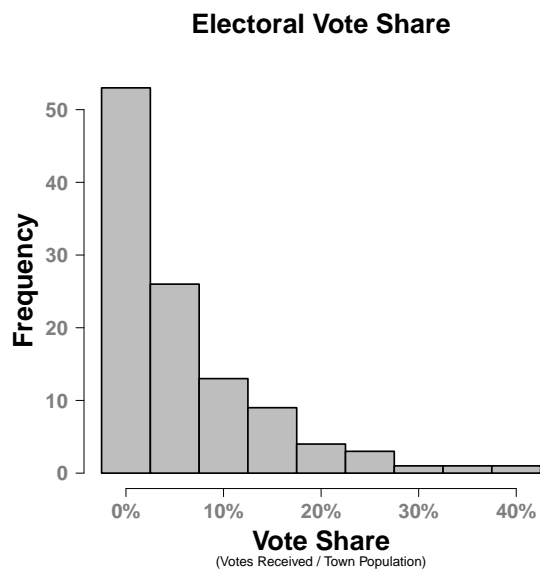


Figure 6.6: Vote Share Distribution. Candidates' vote share calculated by dividing votes for each candidate by total town population.

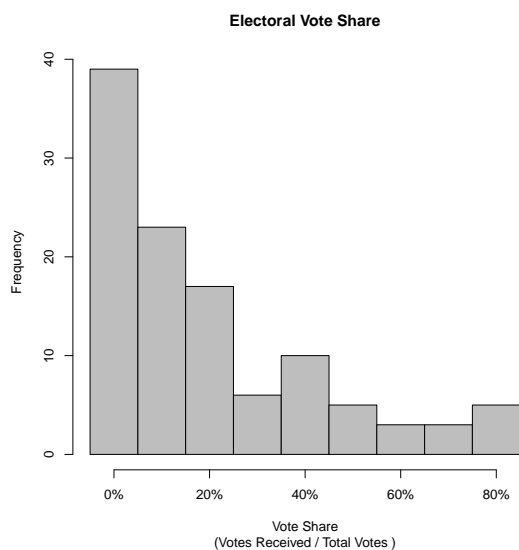


Figure 6.7: Vote Share Distribution. Candidates' vote share calculated by dividing votes for each candidate by total number of votes.

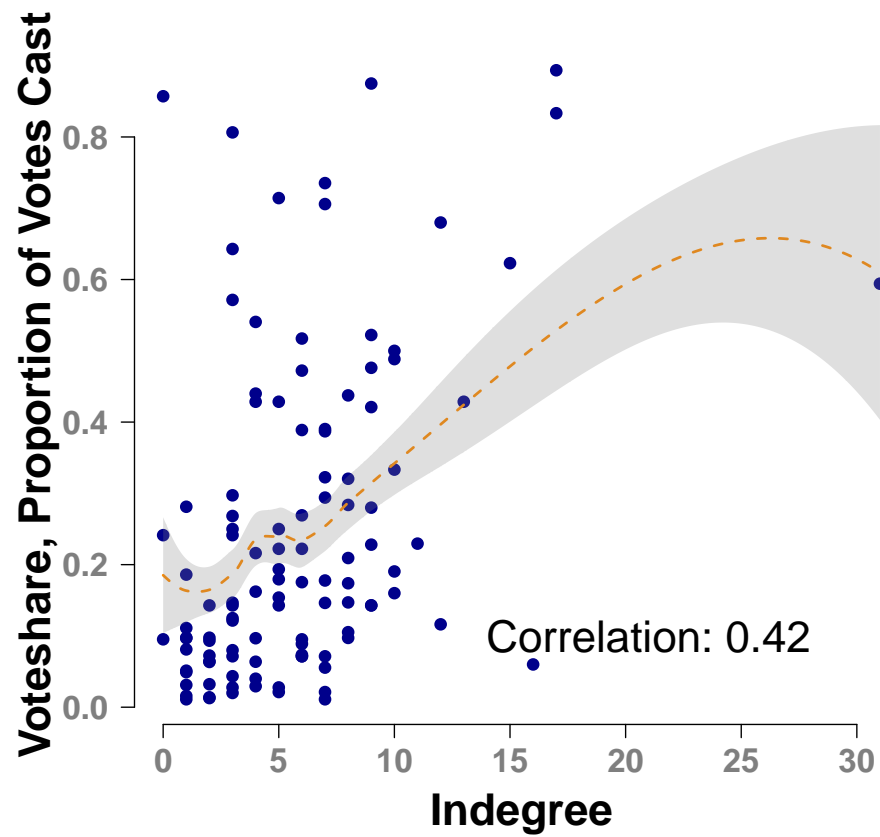


Figure 6.8: Proportion of votes won by a candidate on the y-axis and count of nominations by friends, siblings, and spouses on the x-axis. Correlation reported is Spearman rank order correlation.

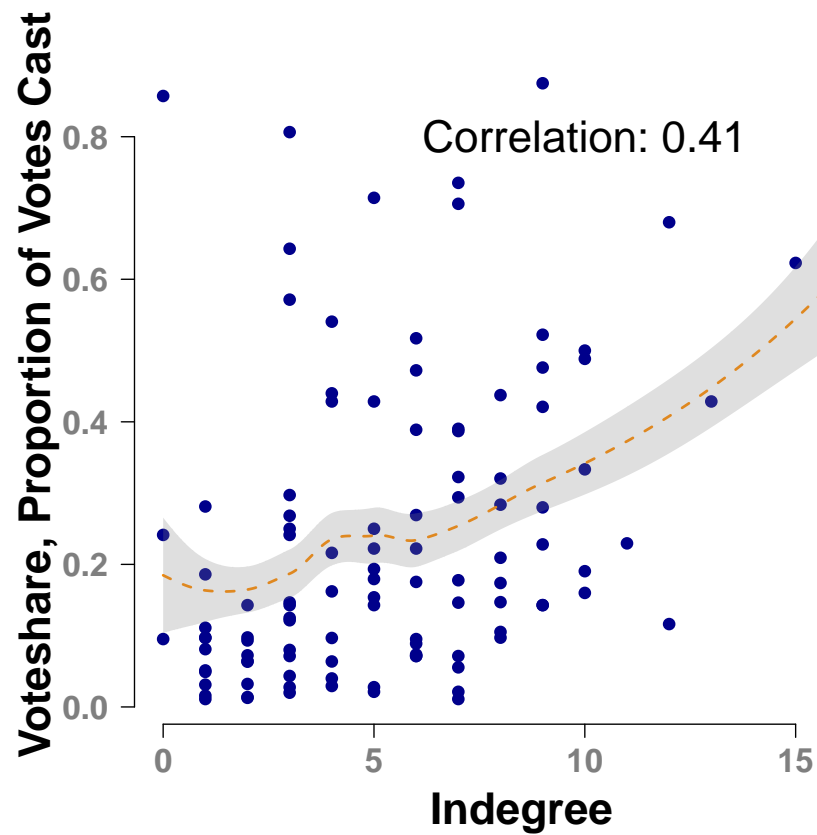


Figure 6.9: Proportion of votes won by a candidate on the y-axis and count of nominations by friends, siblings, and spouses on the x-axis, omitting the outlier candidate with 31 social connections. Correlation reported is Spearman rank order correlation.

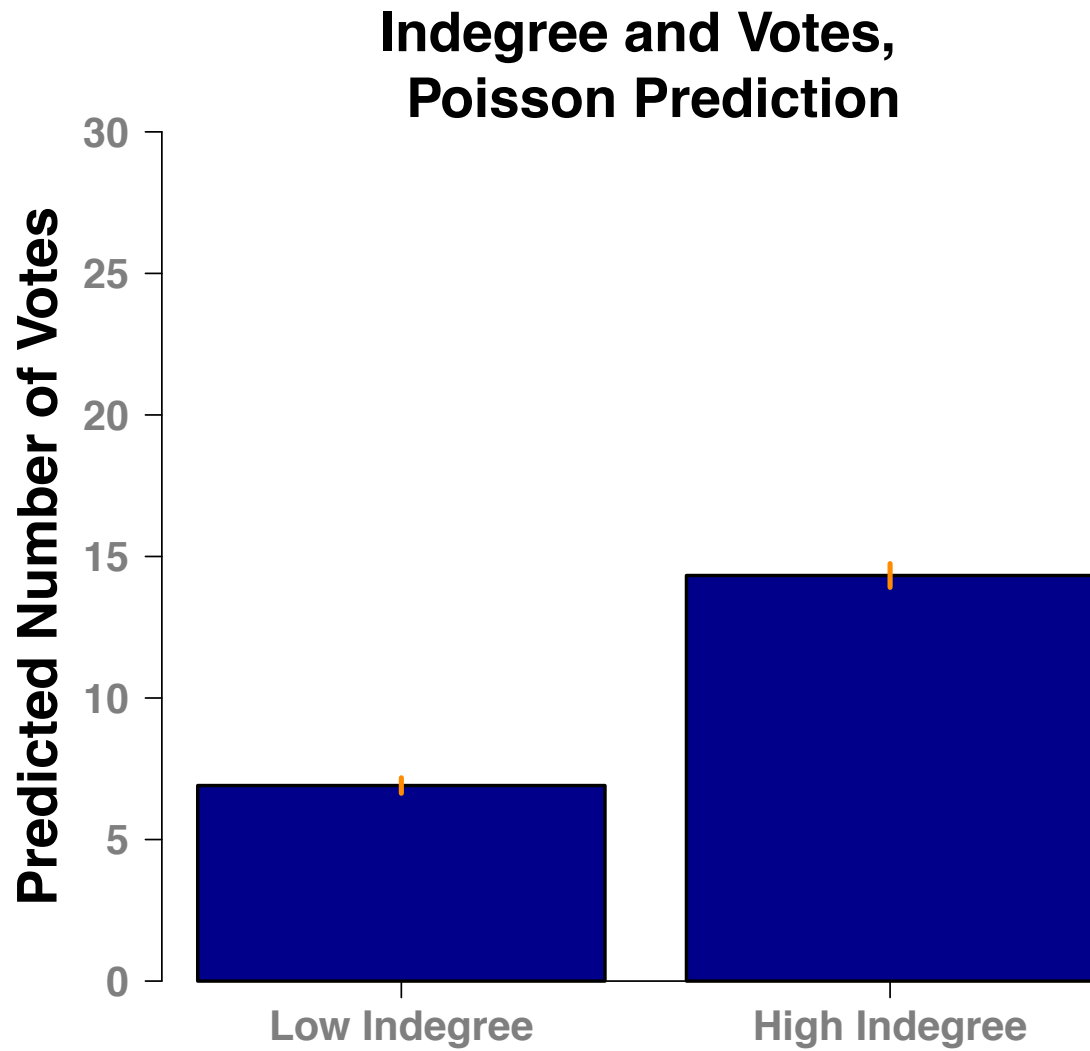


Figure 6.10: Predicted first differences in number of votes predicted from a Poisson model. The change in explanatory variable is $\mu \pm \sigma$. Error bars are standard errors of the prediction.

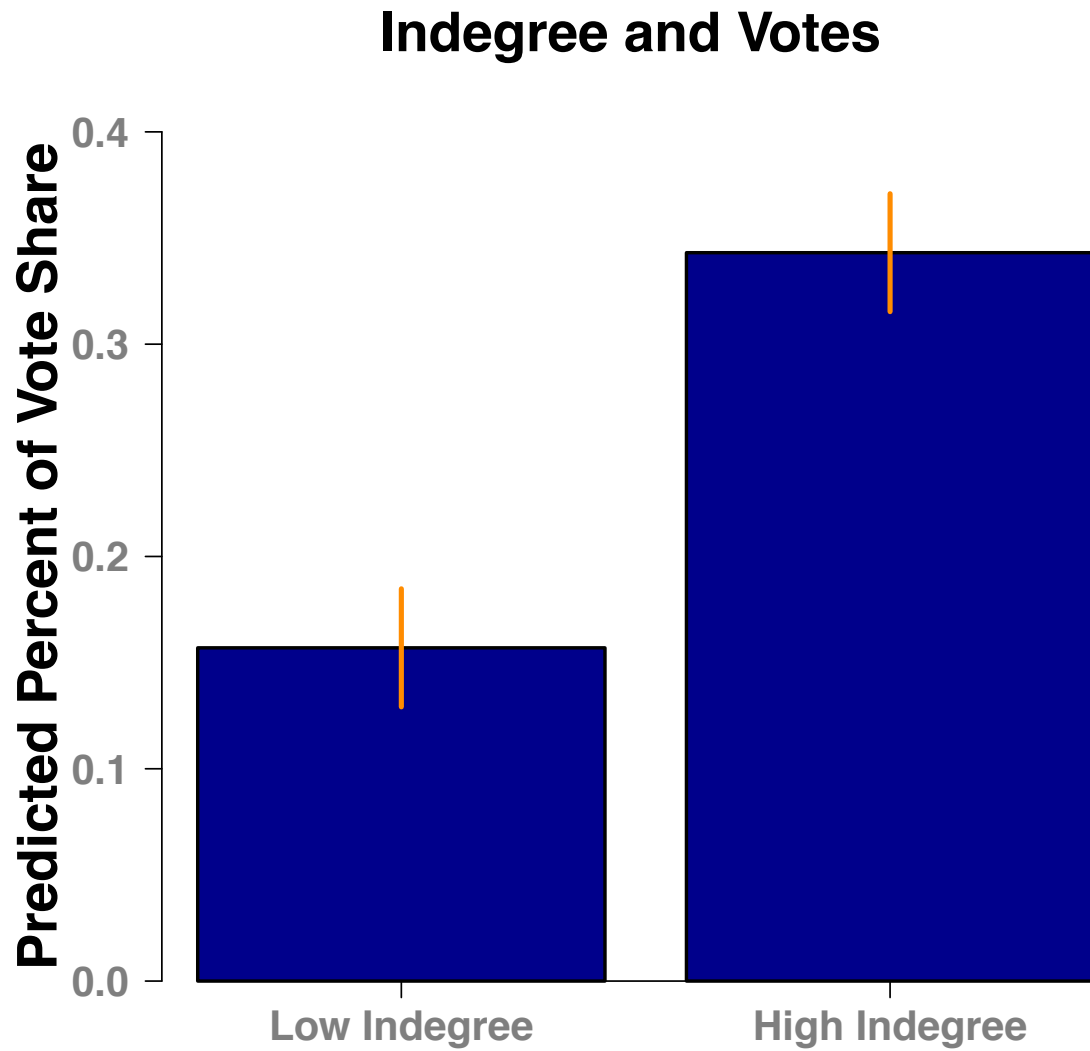


Figure 6.11: Predicted first difference in number of votes predicted from an OLS model. The change in explanatory variable is $\mu \pm \sigma$. Error bars are standard errors of the prediction.

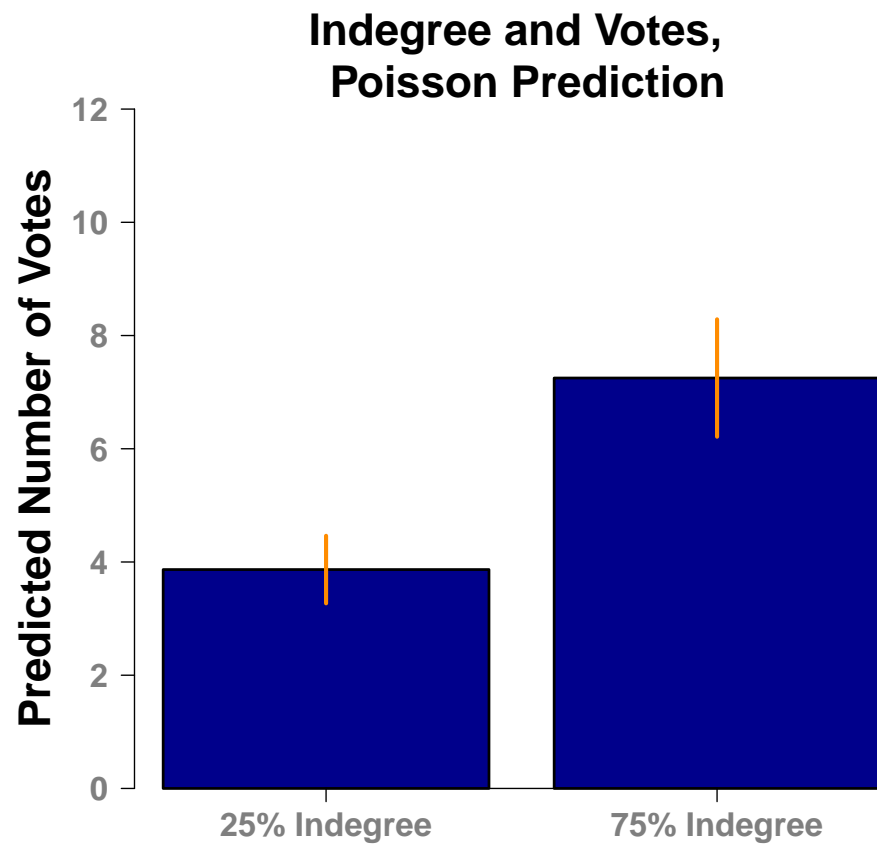


Figure 6.12: Predicted first difference as a result of a two standard deviation change in candidates' indegree centrality. The change in the explanatory variable is $\mu \pm \sigma$. Error bars are standard errors of the prediction.

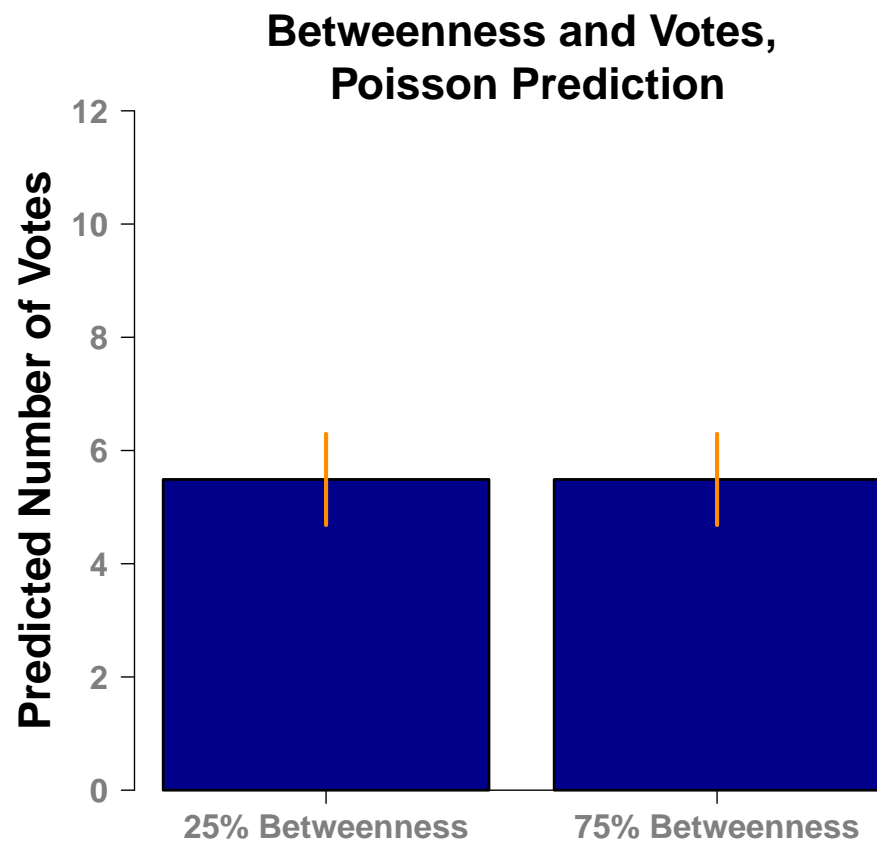


Figure 6.13: Predicted first difference as a result of a two standard deviation change in candidates' betweenness centrality. The change in the explanatory variable is $\mu \pm \sigma$. Error bars are standard errors of the prediction.

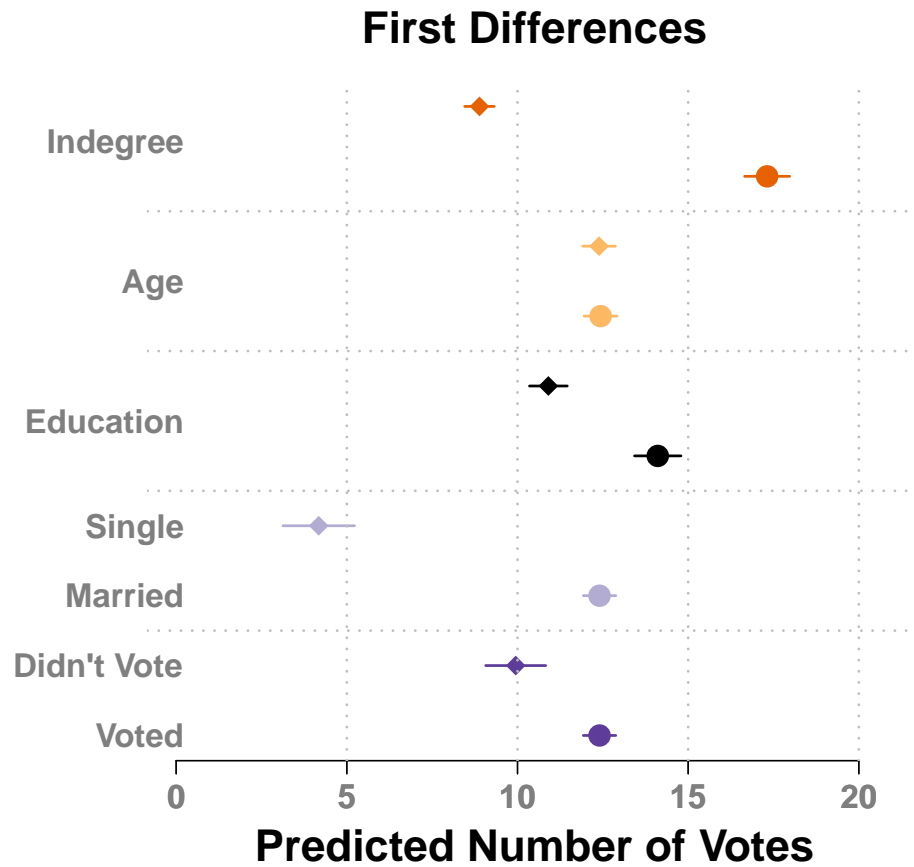


Figure 6.14: Colors are matched for independent variables. For each independent variable, the top row sets the IV at the 25th percentile of the variable's distribution and the bottom row sets the IV at the 75th percentile of the variables' distribution. The prediction scale is on predicted number of votes a candidate will receive. Changing degree causes a candidate to earn five more votes. Changing Betweenness also causes a candidate to garner five more votes. Age has no effect. Additional years of education garner three votes per year, and changing from being married to being divorced causes a candidate to lose ten more votes. Bars are standard errors.

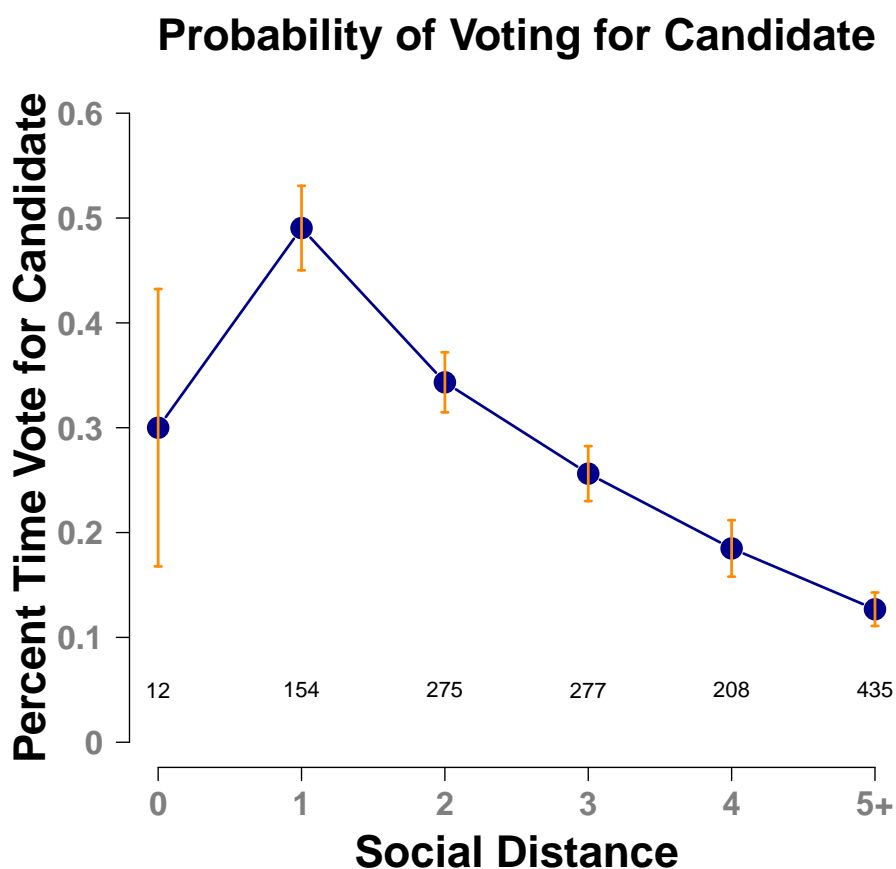
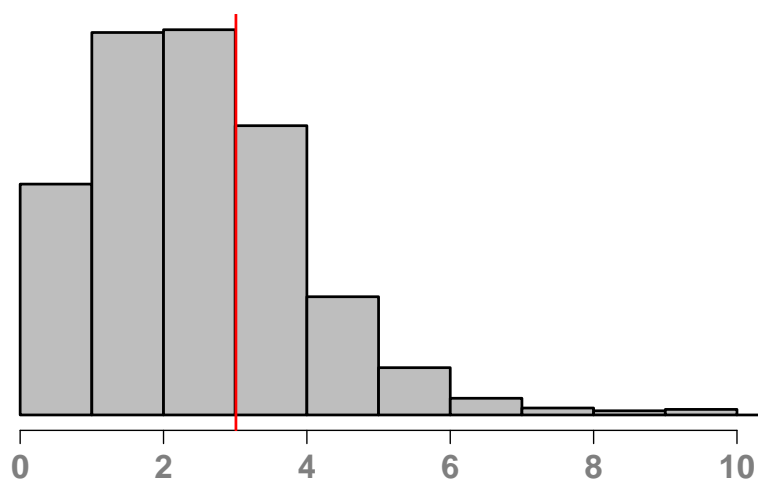


Figure 6.15: Probability of voting for a candidate, conditional on the social distance between the candidate and elector. Social distance of zero is a voter who was himself a candidate; social distance of one is an immediate friend, or family member, social distance of two is a friend of a friend, and so on. The small numbers at $y=0.5$ represent the number of votes cast at each social distance. For example, there were 154 votes cast for candidates at a social distance of one degree.

Distance to Chosen Candidate



Distance to Non-Chosen Candidates

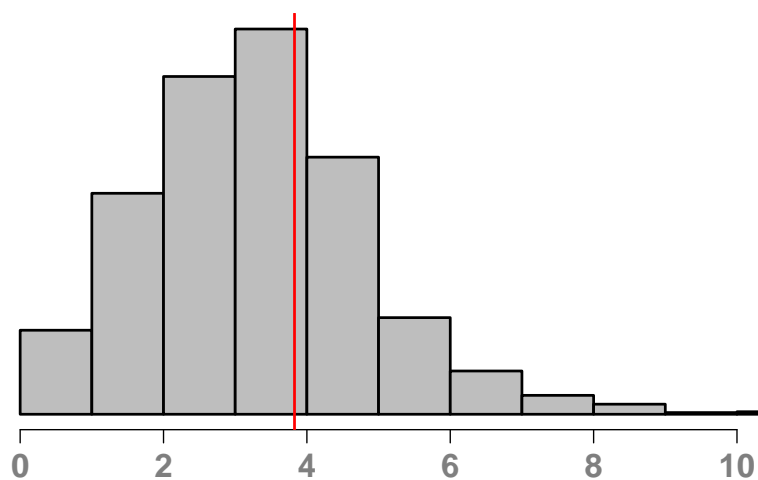


Figure 6.16: Histogram of distance between each voter and the candidate for whom she voted. *Bottom Panel:* Histogram of distance between each voter and the candidates for whom she did not vote. The vertical red line in each is the mean geodesic distance for that subset.

Table 6.5: Voting for a Candidate as a Function of Social Characteristics

	Vote for a particular Candidate				
Geodesic Distance	−0.018*** (0.007)	−0.004 (0.007)	0.002 (0.007)	0.009 (0.007)	0.011 (0.007)
Candidate Indegree		0.013*** (0.001)	0.013*** (0.001)	0.008*** (0.002)	0.009*** (0.002)
Candidate Eigenvector			0.566*** (0.106)		0.225* (0.117)
Candidate Betweenness				5.387*** (0.692)	4.913*** (0.773)
Constant	−1.154*** (0.023)	−1.313*** (0.030)	−1.390*** (0.036)	−1.409*** (0.032)	−1.432*** (0.036)

*p<0.1; **p<0.05; ***p<0.01

Note: Logit model of vote choice at the voter-candidate dyad level; Huber-White standard errors clustered at the voter level. Betweenness is $B'(e)$, scaled betweenness centrality. All models include demographic and quality controls (not reported).

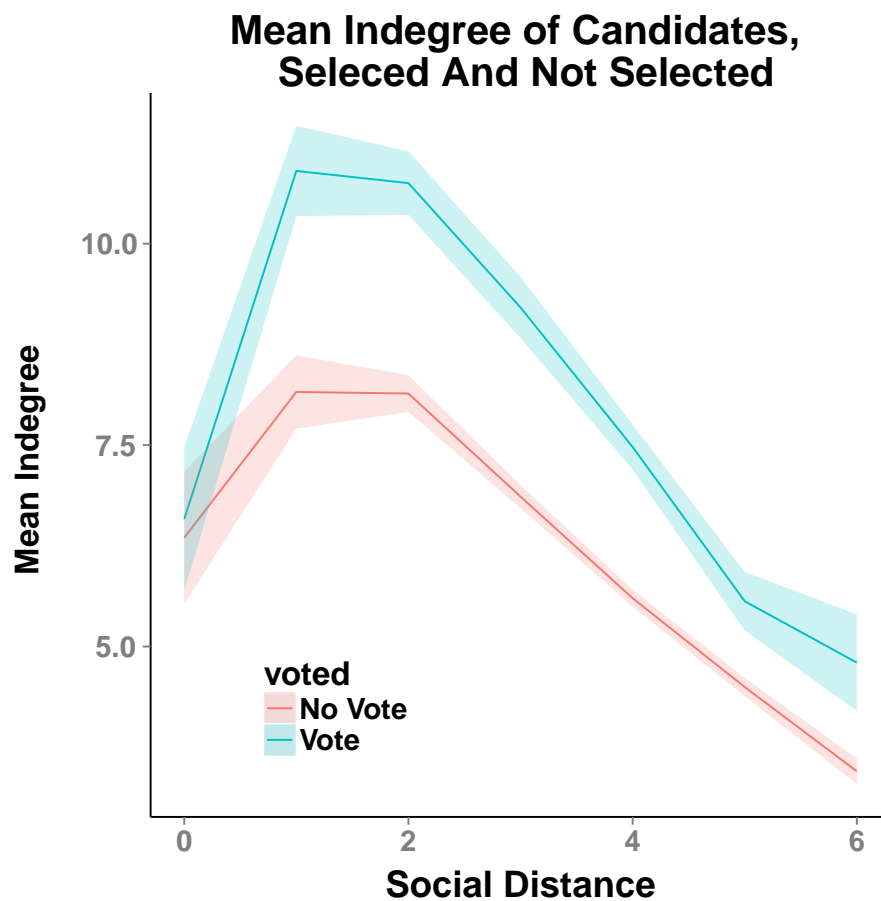


Figure 6.17: The mean indegree of matched pairs of candidates who received voted (in blue), and did not receive voted (in red). Candidates who receive votes uniformly have higher indegree. The polygon shape is a 95 percent CI for the mean.

6.9 Discussion

In this chapter, I examine how social networks shape political coordination in elections. I begin by arguing that for nearly fifty years, much political science has ignored the importance of connections between actors. The first generation of political scientists to seriously reexamine correlates of political behavior and social network (Fowler, 2005; Sinclair, 2012; Rolfe, 2012) utilized technological advances unavailable to early scholars (Berelson, Lazarsfeld and McPhee, 1954) and generated considerable new knowledge about how our friends' political behaviors shape our own (Bond et al., 2012). However, some concerns exist surrounding observational studies in this complex social-information environment (Noel and Nyhan, 2011).

By measuring social networks and randomly assigning actors to stand for candidate in an election, I avoid these criticisms, and for the first time, show how social networks cause some candidates to earn more votes in an election. Furthermore, I present clear evidence that performance in these elections is not caused by voters voting for candidates who are socially proximate; instead voters prefer candidates who are well connected within the social network. This has important implications for our understanding of how actors function in their pursuit of politics. In the American context, this provides more evidence for how Congressional co-sponsorship may be influenced by social connections. It might also explain how groups of actors come to hold norms of behavior—like who they will support among a broad slate of primary candidates. In the Comparative context, this finding might mean that delegated governance to local powers, under some circumstances, may not lead to increase provision of personalistic goods.

6.9.1 Limitations

Despite the researcher's efforts to carefully design this experiment, there remain limitations worth considering. First, to avoid the myriad set of unobserved factors that a pre-election campaign begets, the experiment was expressly designed so that campaigning could not occur. Voters were fully informed of the stakes at play and the identities of the candidates, but they were not familiar with the particular policy positions *per se* the candidates might perform. This is an acknowledged limitation, and one which well cuts against the generalizability of this finding into a context where a vibrant campaign plays a critical role in the eventual slate of candidates brought to the vote.

However, a well-established line of research has previously undertaken limited or artificial campaign environments to evaluate features of voters' preferences (e.g. Eldersveld, 1956). Other research projects in this area have constructed information boards, mailed synthetic election information, shown either organic or synthetic campaign ads (e.g. Brader, 2005; Gerber et al., 2011; Bartels, 2006; Freedman, Franz and Goldstein, 2004; Hillygus and Jackman, 2003; Thielmann and Wilhite, 1998; Finkel, 1993).

Chapter 7

Knowledge and Use of Social Information

7.1 Problem and Proposed Solutions

In chapter 5 and chapter 6, I provided evidence of how social information shapes political outcomes. The theory of social information is predicated on agents understanding the mid-level structural features of their social networks. To date, scholarship has been divided about whether this is a reasonable expectation; early scholars concluded that individuals did not understand their broader networks (Friedkin, 1983; Krackhardt, 1987), while recent scholarship has challenged these assertions (esp. Banerjee et al., 2014). In this chapter I report the results from two tasks. First, I report that among a sample of individuals in a small town in Ghana, individuals are highly accurate when asked to ruminate on their own social network. This is a difficult task, but one that is necessary in order for a theory of social information to hold water. Second, I provide evidence based on theory, simulations, and empirically gathered information that this ability to of subjects to conceptualize and verbalize about their social networks can be used by political actors and interventionists to rapidly and inexpensively target individuals with high scores on specific social dimensions.

Although measuring actors' connections to other actors may be feasible in some cases (Bond et al., 2012), in many more, even those cases where social mechanisms are directly identified as leading political behavior, it is not possible to measure how actors in reference to each other (Putnam, 1995; Walsh, 2004; Lim

and Putnam, 2010; Sinclair, 2012; Cruz, Labonne and Querubin, 2015; Banerjee et al., 2014). Although in the study and the analyses conducted herein I have measured existing real-world social relationships, if there is not a readily-applicable way to transpose these findings into a way that is replicable by other scholars and practitioners (political campaigns, international organizations, non-governmental organizations), then the development of new theory is of little practical or real-world importance. This chapter presents one possibility for how to readily and inexpensively target the high-impact individuals that have been implicated in the previous chapters as instrumental for political activity.

The contributions of this chapter are not the first contributions toward low-cost and robust identification of focal nodes in a networked system.

Others use membership in community groups as an easily measurable, or heuristic means of identifying likely social interactions (e.g. Putnam, 1995; Sinclair, 2012).¹ While this strategy provides some capacity to say that political actors are more likely to share a social tie, this ability it is unlikely that this method if subjected to testing would be either highly specific or have high recall.

7.1.1 Existing Research

Targeting specific individuals to receive treatment holds considerable promise empirical intervention. Presently, three research groups are actively developing techniques for social-network based interventions: Duflo, Banerjee & Jackson; Beaman et al.; and, the group this researcher is affiliated with at UCSD and Yale University. While the groups differ in specific goals and implementation strategies, at a fundamental level, they share the same goal: understanding how to design spillover in a way that maximizes the number of people who are affected by a treatment regime.

In a leading example of this work, Beaman et al. (2014) examines success rates of network based targeting in the adoption of novel farming techniques in Malawi. In particular, Beaman et al. (2014) encourage farmers in distinct positions within the farmers' social network to take-up a *pit-planting* technique that increases water and

¹Briefly, this form of tie-identification is typically referred to in the literature as a two-mode graph. That is, there is one mode of node (the people) who are connected to each other through another type of node (the social group). Thus, there are two-nodes present in the population and nodes between the type-one nodes (people) are inferred through their joint identification with the type-two nodes.

nutrient retention, and thereby maize yield. Across four treatment arms the authors randomly assign *pairs* of farmers to receive treatment in 200 villages. Treatments select the pairs of farmers according to targeting by (1) *benchmark assignment* that relies on local knowledge to select two individuals²; (2) a *simple contagion* model which seeks to identify individuals that hold first-degree social connections with the maximum number of alters; (3) a *complex contagion* model which seeks to identify individuals that have the highest overlap with first-degree alters; and (4) a geographic targeting mechanism as a low-cost means of targeting the same set of individuals.

The extent and overlap of social connections distinguishes targeting based on theories of simple and complex contagion. A theory of simple contagion predicts that every alter who comes into contact with the treatment materials will take up the treatment behavior. In this way, simple contagion is directly analogous to simple epidemiological spreading models like the spread of highly-infectious viruses. In contrast, a theory of complex contagion predicts that multiple, or reinforcing exposures to an intervention stimulus are necessary before an alter changes her behavior. Since the predictions of these theories identify different mechanisms for spillover (simple or reinforcing-complex), when one holds the full social network information about a target population, it should be possible to adjudicate which of the two models is more appropriately characterizing the transitions of alters. Beaman et al. (2014) do just this and find in their work of pit-planting, that compared to baseline, social targeting of any type perform better. They find more limited evidence, however, for the geographic measure of targeting, and identify the important need for better implementations in future iterations.

The other major scholarship group actively addressing this question concerning heuristic identification of socially-connected individuals is Banerjee et al. (2014). In this work, Banerjee et al. utilize their considerable social network measurement effort to develop and test a model of "gossip-centrality." While the purpose of the heuristic measurement is similar to the heuristic geographic measurement in Beaman et al. (2014), the work of Banerjee et al. is better performing.

The basis of the heuristic measurement in the model of Banerjee et al. is the notion of *gossip*, which to the authors is an evaluative token that all members of a

²"Typically, this involves asking village leaders to nominate a pair of extension partners and is similar to what many extension workers normally do outside of our study context," (p. 2 Beaman et al., 2014).

community hold about all others. Conceptually, this gossip functions as a simple counter for actors in the social network, and any news about a person³ is included in this counter. Individuals who are frequently mentioned in daily life are highly rated whereas individuals who are not frequently mentioned are not highly rated. To test this model, the authors measure the social networks of 35 village and then as a followup question of each village member about who the "best individual in the village" would be to start the spread of information. The results are striking: on average the individuals who were nominated were in the seventy first percentile of the eigenvector centrality measure in the village; and, perhaps more strikingly, the individuals identified in this form of elicitation were only infrequently individuals identified as holding formal political, religious, or labor roles in the village.⁴

The final step in Banerjee et al. (2014) is to test whether seeding information with those who are nominated as likely to spread the information leads to increased diffusion of the information between individuals. For this test the authors partner with a cellular telephony company and offer anyone who has the information the possibility to win either a cash prize or a cell-phone. The authors do not take a clear stance on whether they model this spreading of information as either a simple or complex contagion event; however, because the cost of spreading information is so low (sharing a telephone exchange), and the cost of uptake is equally low (calling the shared number), it would seem to be straightforward to classify this as a simple contagion event. Once more, the results that Banerjee et al. (2014) find are striking: in villages where gossip targeting took place response to the stimulus is nearly four times as large as baseline villages.

Despite well-crafted design, these two previous groups of studies leave considerable room for theoretic and empirical improvement. First, despite the careful design to assess whether pit-planting is a simple or complex contagion event, Beaman et al. (2014) spend very little time distinguishing specifically what would make the transmission of planting knowledge behave as one, rather than the other, contagion model. Are the costs of pit-digging high enough that multiple encouragements are required to overcome the increased startup costs? Are the

³This might be an "at mention" in current digital representations of this phenomena. For example, @ThadKousser (doen't exist!)

⁴This result is consistent with the work of Obradovich and Hughes (N.D) who found, similarly, that in a social network of fishermen and fisherwomen in the Western Region of Ghana, that the most highly socially connected individuals were the lieutenant fishing chief and his wife. In stark contrast, those individuals who might classically be targeted by intervention groups – the village mayor and English-speaking school teacher – were relatively poorly socially connected.

benefits so large that any farmer who hears of their possibility would undertake such a project? Is the norm around planting of trees so strongly injunctive that no new tree planting regimes could possibly be tolerated? None of these issues are addressed in this working paper.

Second, although the authors have undertaken considerable cost to collect the social network data, they have only a very limited snap-shot of the social network. As such, the authors implicit assumption is that there are no other pathways of communication and transmission that are relevant to the outcome. Indeed, somewhat troublingly, the authors have acknowledged that social institutions shape behavior, but subsequently assume that they have identified the single social institution that matters to the question at hand.

Third, Beaman et al. (2014) find only limited evidence in support of their heuristic measure of geographic proximity; likely do to the difficulties in measurement and uncontrolled spillover. Furthermore, the research team this author is affiliated with is currently developing and analyzing the relationships between physical geography and social network connections. It is our expectation that geography will play some role in the formation and maintenance of ties; however, much as addressed earlier in this chapter, geography as a heuristic means of identifying well-connected individuals suffers from both low precision and low recall.

Fourth, whereas Beaman et al. (2014) is cognizant of simple the subtleness of any contagion framework, in contrast, Banerjee et al. (2014) develop an intervention that is evidently structured toward a simple contagion event. Despite the usefulness of this measures to understand very simple contagion and diffusion dynamics, and the nicely developed model for behavior, the empirical test in Banerjee et al. (2014) does not speak clearly to the performance of even moderately complex events like political or health decisions. Within the context of the authors' empirical research, anyone who receives the information about either free money or a free phone, and believes that information is credible is very likely to accept that information and adopt the behavior. Indeed, because the cost of adoption of the behavior is so low, there is very little evaluation that subjects must undertake; and so the test very nicely assesses how highly virulent pathogens might spread, but does not assess how political attitudes might shift or norms be held.

Apart from the work of Banerjee et al. and Beaman et al., scholars working in other contexts with connected data propose additional alternative methods for identifying well connected nodes. Of note are cases of internet- and mobile-based

connections, for example Facebook, Instagram, and especially Twitter. In these contexts, while there exists large amounts of relatively high-quality data about the connections between individuals, the large amounts of data actually comes to serve as a limitation, rather than a feature of the data. Many higher-order measures of connectedness scale poorly in time as data complexity increases; even more, when data become very large, researchers face limitations in accessing the entire corpus of data, and so are limited in the form of solutions present for calculating connectedness.

For example, Steinert-Threlkeld (N.D.) works to identify well-connected actors in a network of Twitter users. After identifying that the most frequent operationalization is indegree, Steinert-Threlkeld identifies limitations in this measure. Because indegree utilizes only local information, two individuals with similar indegree may exist in largely divergent positions in the overall structure of the network. Like Banerjee et al. (2014) the author proposes expanding the horizon for evaluation, proposing using the cumulative indegree of the specific targets' alters. That is, Steinert-Threlkeld proposes summing the degree of all two-degree alters.

An important similarity between this cumulative indegree measurement and the proposed modification I propose in this chapter is that both methods propose using additional information beyond what is included in the set of one-degree social alters. In both internet- and mobile-based data collection, as well as in democratizing contexts, it may be difficult or costly to obtain full network data. This difficulty, together with the lack of measurement validity of relatively easier to obtain centrality measures, provide reason for including additional data in the characterization of well-connected individuals.

7.1.2 Unique Contribution

The key difference between the work that I present in this chapter and existing work is that I propose to utilize individuals' own knowledge of the network, rather than relying on information provided by that network. This choice is motivated primarily by pragmatic considerations. In the case of connected internet data the marginal cost of contacting and computing scores for alters is relatively low. In contrast, in the case of the social data in rural locale, the cost of contacting alters is relatively high. In a connected computer system, the data exists about not only the identify of other nodes in the system, but also other useful information: a

permanent address of the alters (perhaps an IP address, a record in a database, or a location on a hardware rack). In the case of interpersonal connections between humans, especially democratizing contexts, frequently such information is difficult to come by.

For concreteness, consider the example of calculating how well connected is the same individual, but using different technology. In one case, imagine the scholar seeks to determine the structural position of an individual within the network of Twitter users. Two lines of code can identify a user, and spider to each of her friends, and then her friends' friends, generating enough information to compute neighbors' cumulative connectedness. In contrast, imagine the scholar seeks to determine the structural position of an individual with the real-world social network of Capital Hill staffers. Rather than issuing two lines of code, the scholar would need to first contact the target staffer, solicit from this staffer an exhaustive list of her social contacts, and then track down these social contacts. Then, the scholar would need perform this same task for each of the social contacts' social contacts.

7.1.3 Friendship Paradox

With a theory that requires the identification of highly-connected individual to test its propositions, and in front of a backdrop of the problems with which the determination of well-connected individuals in a sampled network, in this section I examine a way forward using iterative, *friendship based* targeting. This targeting utilizes a property of social networks known as the *friendship paradox* whereby every persons' friends have more friends. I develop this concept in through the rest of this section.

The friendship paradox is the observation that, in the real world, most people have fewer friends than do their friends (Feld, 1991; Newman, 2003). The paradox has been applied to sexual partners. Stated differently, one's friends typically have more friends than does she. Recent research has leveraged the friendship paradox to identify well-connected individuals in the social network of Harvard undergraduates. This work show that when monitoring the outbreak of influenza friends nominated by randomly selected individuals are able to serve as early indicators of influenza outbreak (Christakis and Fowler, 2010). Christakis and Fowler authors argue this increased lead-time is driven by the fact that more central individuals are more likely to be in social contact with others within the social

network, thereby contracting the virus earlier and also spreading to a broader catchment of friends.

Christakis and Fowler (2010) is the leading past application of the friendship paradox to circumvent the measurement of a whole network (c.f., e.g. Schneider et al., 2011).⁵ Here, though, the authors note that the beneficial properties of friends' higher degree is driven by the mechanism that people tend to be friends' with the highest connected members of the network. As the authors identify, if a random sample of the population has some mean degree (μ), then the performance of friendship targeting increases in proportion to the variance of the degree distribution. Thus, although the general observation that "the mean number of contacts for friends will be greater (and potentially much greater) than the mean for a random sample" (Christakis and Fowler, 2010) is true, in there is no such guarantee that this phenomenon holds in a small sample.

This use-case from Christakis and Fowler (2010) highlights a more general point to be made about the usefulness of the friendship paradox, especially as a method to identify well-connected actors: The friendship paradox is a population-level phenomenon expected when many individuals have their friends sampled many times. As such, neither a particular sampling of individuals is guaranteed to provide alters with a larger number of friends, nor is a particular sampling of friends of an ego guaranteed to produce a larger number of friends. And so, in the context of using the friendship paradox to raise the connectedness and centrality of a subject receiving some program intervention, any particular sampling realization might (or might not) yield more highly connected subjects. Indeed, after the initial application in Christakis and Fowler (2010), later cases have also utilized the friendship paradox to identify leading actors in spreading hashtag memes on twitter (Garcia-Herranz et al., 2014), but these studies too relied on a very large sample of users to ensure the applicability of the friendship paradox.

7.2 Proposed Modification

I build on the standard method in the following way: rather than sampling a large number of individuals and identifying the universe of these individuals' friends, instead, I propose nominating a small number of individuals and narrowing

⁵Although, also see the later use-case by Garcia-Herranz et al. (2014). Also see, c.f. Schneider et al. (2011).

the nomination set of their alters to highly connected alters. That is, within a network, I propose sampling individuals, and to each, querying, "Who are your friends with the most friends?"

The best-connected friends methods holds several potential benefits over current methods. First, like the other friendship targeting methods, this method stands to identify better connected individuals than those selected at random. Second, to the extent that subjects are able to reliably name high connected social alters, this method stands to reduce the likelihood of a "bad draw" of poorly connected social alters. Third, as identified by Christakis and Fowler (2010) the friendship paradox is the most beneficial when the degree-distribution of connections is highly skewed; a method which nominates well-connected individuals may perform well in networks where the degree distribution is not so highly skewed, presenting the potential for use-cases in networks that were previously untenable.

Effectively, by asking for "friends with the most friends" the researchers is able to limit the response set from which subjects are choosing. Consider (the admittedly toy) example of a subject who has only two social alters. For the sake of concreteness, suppose this individual is Blake O'Neill the punter for the Michigan Wolverines who, on October 17, 2015, with less than :10 seconds left in the game, fumbled a snap which was recovered by the opposing team and returned for a touchdown, leading to a Wolverine loss. In the moments following that play, O'Neill may have had only two social alters: the long-snapper who poorly snapped the ball (and has few social connections) and the quarterback (who has many social connections). In this scenario, if O'Neill were randomly selected as an ego for a friendship-nomination task, he would nominate one poorly connected individual and one well connected individual, and an algorithm that could intervene only with one would stand a 50-50 chance of selecting the poorly connected individual. Limiting the nomination set to the "friend with the most friends" stands to remove the possibility of nominating the long-snapper, and instead increases the probability of nominating the quarterback.

7.2.1 Requirement for Proper Functioning

Simulation-based studies suggest the improvements that may be realized through best-connected friend based, targeting. But, there is considerable cause for concern that individuals who are provided this targeting prompt may not function

as those agents in the simulation. In this subsection, I make explicit the two tasks necessary for this form of targeting to produce reliable results. Following the introduction of these criteria, I present results from data collected in rural Ghana to assess to what degree these criteria might be expected to be met by subjects.

For this proposed alteration to function properly the following conditions must be met. First, egos must reliably nominate alters with whom they hold social connections. Second, egos must reliably nominate alters who are the best connected of their set of social connections.

1. Egos must be able to reliably nominate "best connected alters" who are, in fact, *their alters*. A concern with this form of targeting is that rather than nominating individuals who are their social alters, this targeting scheme may just converge to the few individuals who are very well connected in the network. That is, it is possible that in small-scale groups where respondents know most others in the group, that the "best connected" criteria holds greater primacy than the "who you are connected to" criteria. To the extent that convergence were to occur, the targeting would yield high levels of overlap. In the limiting case, all randomly selected alters may nominate the same individual, in more plausible cases, this convergence issue may limit network coverage.
2. *Egos must be able to reliably nominate well-connected individuals as the "best connected alters."* Because this is a behavioral measure, it is conceivable that randomly selected egos may lack the capacity to reliably nominate well-connected individuals in response to this prompt (Friedkin, 1983; Krackhardt, 1987). Rather than nominating some friend, formulating the question in this way places demands a slightly higher cognitive load of the respondent to search her mental rolodex to find a well-connected friend. While this response is likely to be subject to the same cognitive limitations faced by all survey respondents (see e.g. Hafner-Burton, Hughes and Victor (2013)), minimally, this small modification limits the likelihood that an individual nominates a friend who *in actuality* has fewer friends. To the extent that the nomination procedure breaks down, the performance of the modified algorithm will suffer.

Provided these two criteria hold, the best-friend nomination technique can perform no worse than the baseline nomination technique and may potentially avoid randomly selecting individuals who are the very poorly connected friends

of seeds. And, if the best-connected friend nomination avoids nominating poorly connected friends, this proposed algorithm stands to potentially improve the performance of the friendship targeting algorithm. I test the performance of these two points later in the chapter, but the data suggest subjects are well equipped to handle this task.

7.2.2 Evidence Suggests Proper Functioning

The results from the Ghana nominations provide clear evidence that both of these criteria are met by subjects. First, subjects' response to the "Friends with the most Friends" name generator were in line with the requirement of in fact being socially connected. Fully eighty-three percent of these nominations were social alters. To provide context, simulations suggest that if subjects were nominating alters without regard to being socially connected, in a network of this size and structure, 8.9% of nominations would be of a social alter by chance ($SD = 3.7\%$). If, instead each person were to simply nominate the single best connected individual in the network then 61% of the network would be social alters. Limiting the nomination set to only those who are Friends (specifically eliminating siblings, spouses, and boat mates) has little effect on the success of the best connected friend nomination being a social alter. Under this more strict criteria, the success rate is little changed at 75%, while only 36% of the network is socially connected to the best connected individual through a friendship channel.

Second, the evidence from this data also suggests that individuals are capable of nominating the best connected of their social alters. To evaluate this second criteria, it is possible to compare the incoming and outgoing connections for those nominated as the "best connected friend" and simply score whether this individual was the best connected of the set of friends. As Figure 7.1 demonstrates, more than half of the nominations for best connected friend correctly identified a social alter that was the best connected. One concern in this comparison is that the researchers measured only a sub-set of the set of possible social connections; in addition to those measured, there may exist other social dimensions such as colleagues from school, sharing, borrowing, or health advice. To the extent that these alternative social dimensions exist and are salient, subjects' may tally the connectedness of their social alters on these other dimensions. To evaluate how this conceptual difference might translate into performance of the heuristics, one might incrementally increase the

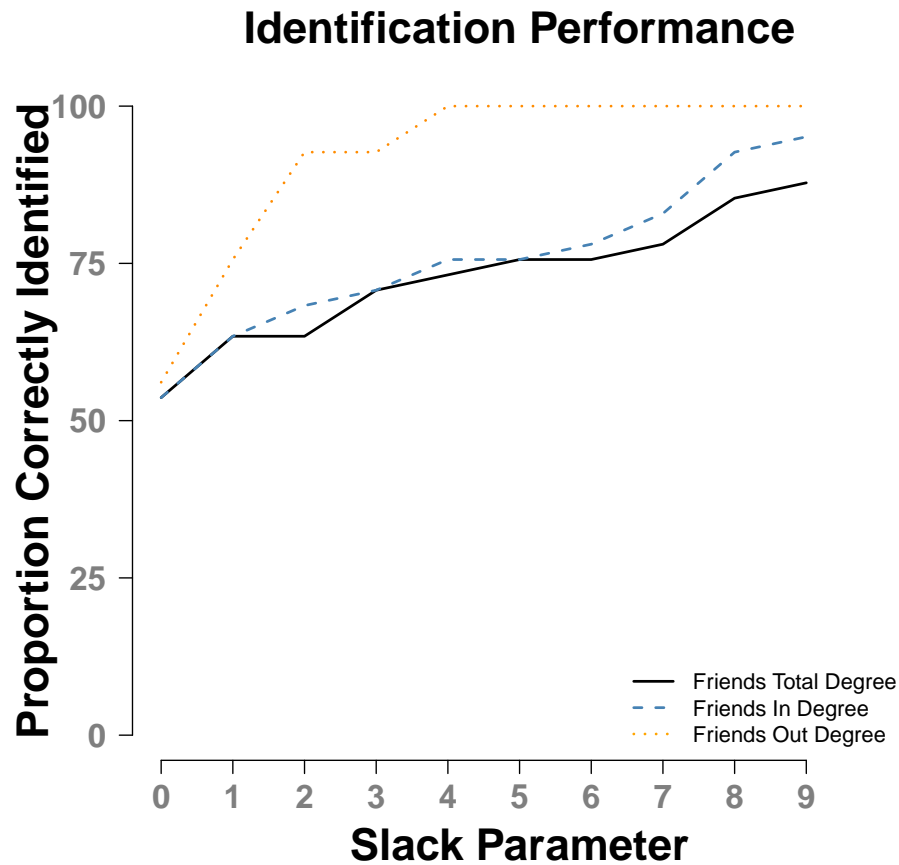


Figure 7.1: Best connected friend identification performance. On the y-axis is the percent of nominations of best connected friends that were, indeed, the best connected. Along the x-axis is a "slack" parameter to allow for potentially unmeasured social connections to exist within subjects' conceptualization.

number of connections held by those nominated as being best connected, holding constant the connections held by others. The results of this mapping are reported in Figure 7.1. Slackening the comparison by a single degree increase the performance of the Friends' outdegree metric from 56% to 75%.

7.3 Criteria for Success and Evaluation

I propose three criteria to evaluate the performance of the best connected friend proposed nomination regime. Because the aim of this technique is to identify subjects who fill important roles in their social networks, each test addresses a

separate component of this identification. The leading test for the performance of this strategy is whether the strategy is, in fact, leading to the identification and nomination of well connected individuals. The second and third evaluative criteria, network coverage and non-overlap, are secondary performance checks to ensure that best friend targeting is not over-identifying only central individuals.

7.3.1 Node Centrality

The principle performance task for this algorithm is identifying highly *central* individuals in the social network. To assess this performance, I use a suite of first-order and higher-order centrality measures. Each measure, founded against a particular theoretical background, aims to measure a slightly difference conceptualization of being an important actor.

Degree, the first-order centrality measure identified earlier in this chapter is the first of three operationalizations. Degree simply counts the number of social connections a particular individual holds, without any transformation or weighting of these connection (hence the first-order designation). To further distinguish the components of behavior leading to having a high degree score, I further distinguish between friendship nominations that are made *by* the reference ego⁶ and friendship nominations made *of* the reference ego.⁷ Friendship nominations made by ego are designated "out-degree" for the outgoing nature of the social connection while friendship nominations made of the reference ego are designated "in-degree". This is a common distinction.

Beyond degree, I test two well-established higher-order centrality measures. The first, *betweenness* captures how many times a particular focal node is between two other nodes. "Betweenness, as one might guess, is a measure of the extent to which a vertex lies on the paths between others," (p. 2 Newman, 2005). The most frequently identified betweenness measure, attributed to Freeman (1977, 1978), typically called just betweenness, is defined by Wasserman and Faust (1994) in the following way: "Let g_{jk} be the number of geodesics [paths] connecting two actors [j and k]. Then let $g_{jk}(n_i)$ be the number of geodesics linking the two actors that contain the actor i ," (Wasserman and Faust, 1994, p. 190). Then, the actor betweenness centrality is the count of these for every pairing jk that does not

⁶Ego saying, "I am ego. Alter 1, Alter 2, Alter 3 are my friends."

⁷Alter 3 saying, "I am friends with ego." Alter 4 saying, "I too am friends with ego." Alter 5 saying, "As well, I am friends with ego."

include i ; or

$$C_B(n_i) = \sum_{j \neq k \neq i} \frac{g_{jk}(n_i)}{g_{jk}} \quad (7.1)$$

Because the number of geodesics scales with the size of the network, it is necessary to standardize this measure by the total number of theoretical ties $(g-1)(g-2)/2$ where g is the total number of nodes in a graph. This arises because there are $(g-1)$ nodes for a tie to originate from⁸ and, naturally, $(g-2)$ nodes for a tie to end. In an undirected graph, a tie from jk is scored the same as a tie from kj , and so the max is bounded by one-half the total of $(g-1)(g-2)$. Then, the standardized formulation is

$$C'_B(n_i) = \frac{C_B(n_i)}{(g-1)(g-2)/2} \quad (7.2)$$

$$= \frac{g_{jk}(n_i)/g_{jk}}{(g-1)(g-2)/2} \quad (7.3)$$

After standardizing the measure, C'_B falls on the range $[0, 1]$ and, “can easily be compared to the other actor indices, as well as across networks and relations,” (Wasserman and Faust, 1994, p. 190).⁹

The second well-established higher-order network centrality measure I use to evaluate heuristic performance is eigenvector centrality. This measure scales the centrality score of each node in proportion to the centrality scores of the other nodes to which it is connected; the measure is, “based on the idea that an actor is more central if it is in relation with actors that are themselves central,” (Ruhnau, 2000, p. 360). The centrality score for any individual, i , then is just the i^{th} value of the principle eigenvector of the adjacency matrix.

In addition to the well-established higher-order node centrality metrics, I examine diffusion centrality (Banerjee et al., 2013, 2014). Diffusion centrality takes as some prior q the likelihood of passing information between nodes, which the authors suggest might be modeled as the inverse of the first eigenvalue of the adjacency matrix, g , and some set number of iterations, or time periods T wherein

⁸ $(g-1)$ not including reference node i .

⁹Newman (2005) develops a related measure of random-walk betweenness that does not rely on shortest paths between actors but instead stochastic walks between nodes.

Table 7.1: Implementation of centrality scores

Name	Citation	Formula	Implementation
Degree	W&F, p. 178	$\sum_j x_{ij}$	igraph, degree
Betweenness	W&F, p. 190	$\frac{g_{jk}(n_i)/g_{jk}}{(g-1)(g-2)/2}$	igraph, betweenness
Eigenvector	Ruhnau, p. 360	$Ac = \lambda c;$	igraph, centr_eigen
Diffusion	Banerjee 2013; 2015	$[\sum (qg)^t]$	keyplayer, diffusion

information can potentially spread. Diffusion centrality, DC , is then

$$DC(\mathbf{g}, q, T) = \left[\sum_{\forall t \in T} (q\mathbf{g})^t \right] \quad (7.4)$$

When T is equal to zero, then DC is proportional to degree centrality.¹⁰ As T grows large, DC approaches eigenvector centrality.¹¹ This diffusion centrality is implemented using the keyplayer package with default arguments in R. The implementation of each centralization metric, along with citations to scholars who proposed these methods, and a simple formula for the method is included in Table 7.1.

7.3.2 Network Coverage

Theory suggests that social changes in behavior are most likely to occur when subjects hold close social connections. Bond et al. (2012) examines immediate friends' effect on voter turnout; Aral and Walker (2012) examine the Facebook social network to find influential alters. In fact, looking more broadly at the field experimental literature, many treatments cluster treatment assignment at the household level as a means of minimizing spillover of treatment which contaminates the two-group comparison as an estimate of the average treatment effect.

A fully-informed intervention, one which is appraised of the entire treatment groups' social network ahead of assignment of treatment and control roles, might recursively block to build maximal network coverage. Such a regime might take the following form: After measuring the whole social network, a researcher might

¹⁰degree centrality multiplied by q

¹¹For some values of q . In particular, if q is greater than the first principle eigenvalue from the adjacency matrix, as $T \rightarrow \infty$, DC approaches eigenvector centrality. If q is less than the first principle eigenvalue, then DC approaches Katz-Bonacich centrality, which is not addressed in this dissertation chapter.

sample a single individual without replacement. However, in addition to removing the sampled individual from the next random assignment draw, if the goal is to ensure maximum one-step network coverage, the researcher might also remove all of the social alters who are one degree separated from the first draw. Upon making a second draw from the pool of possible treatment assignments, the researcher would then also remove all the one-degree separated individuals from future assignments. In this way, while the social connections between actors are non-random, a researcher can build a fully random sample from the population using this subset sampling. The goal of this targeting is to build a technique for when the intervention is not fully informed of the social network.

To evaluate network coverage, I propose to measure the proportion of the network within one degree of the randomly assigned treatment receiver. This one-degree metric, while somewhat more simple than other possible choices has several desirable properties. First, and foremost, because social connections are not weighted or scaled this measure is easily understood by non-expert practitioners. Second, this metric scales well across networks of different sizes; or, minimally, it scales better than higher-order metrics. Third, and finally, this measure is more consonant with theory about social influence than other techniques that provide higher order weightings to social connections.

In particular, I define *coverage*, c_r as the sum of the nodes within a single degree of a randomly selected treatment nodes, divided by the sum of the nodes in the network that are potentially within one degree.

$$c_r = \frac{\sum_i^N n_i}{\sum_i^N}, \quad (7.5)$$

where c_r is the coverage given a randomization of selected treatment nodes; n_i is an indicator function that evaluates as 1 when node i is within a single degree of a treatment node, and N , the summation maximum is the total number of nodes who are a part of the connected graph.

To test the reach performance of the best-connected friend nomination strategy I utilize randomization inference.¹² Nodes are, at a fundamental level, not independent of one another, and so statistical test that rely on assumptions of independence to compute p-values will generate downwardly biased estimates of the

¹²Randomization Inference the *term d'art* in political science and field experiments applications is alternatively called a permutation test in the network inference literature.

standard errors and inappropriately small standard error leading to Type-I error that is more frequent than desired. Randomization inference remedies this problem by producing a large number of hypothetical treatment assignments and assessing the proportion of those hypothetical assignments where the quantity of interest is greater than calculated in the single realized case (Gerber and Green, 2012).

It is important to note that the amount of coverage and one-degree overlap that produce the greatest rates of intervention adoption are empirical quantities of interest that are dependent upon the incentive set before subjects. For example, the experiments of Banerjee et al. (2013) and Banerjee et al. (2014) with very low cost and very high likelihood of spillover might do best with a very broad one-degree network coverage and very low redundant connections; in contrast, if transmission of intervention is more consistent with a complex contagion model, finding individuals to seed that both have overlapping ties, and many ties may be the most beneficial set of intervention seeds. In this section, I acknowledge this interest, but largely hold it aside to examine the performance of this identification measure.

7.3.3 Overlap and Non-Overlap

The final potential concern in the nomination of best-connected friends is that all randomly sampled nodes nominate the same individual or small set of individuals. In a small network, especially, it might be the case that there is a single very-highly connected individual with whom everyone shares a social connection.¹³ Such a process would lead to randomizations that effectively fail to reach the number of individuals desired to directly receive treatment.

Consider, for example, the left panel of Figure 7.2. In this graph with six vertices, if we were to sample any two vertices from the set {B,C,D,E,F} the response to the query "Who is your friend with the most friends?" would be A. In this stylized case, no matter who is sampled (that is not A) the scholar would identify the same best-connected friend. In addition, if the transmission of intervention is a sufficiently simple-contagion event, then seeding the intervention with A would reach all members of the graph in one-geodesic.

Continuing this example makes clear the benefits of friendship-targeting

¹³In small networks, while it is likely that every actor is familiar with all others, the type of relationship I am measuring is stronger than only familiarity.

as a means of identifying well-connected individuals. First, if it were the case that a researcher had measured the social connections of all actors **A–F**, then that researcher would intervene with node **A** if she were targeting a well-connected seed actor. Second, consider the case where a researcher does not know the social connections of these actors, but still wanted to intervene with highly connected individuals. The naive method for this intervention would be to select one individual at random; and, under this strategy, she has a one-in-six chance of randomly sampling node **A**. A more informed strategy might examine covariates of the nodes and use some prior knowledge about covariates that are predictive of being well-connected. Indeed, this was the strategy of Beaman et al. (2014) who used geographic targeting. Third, consider using the knowledge of the individuals who are a part of the social network. In this case, asking subjects to identify their best-connected friends will identify **A** in five-of-six cases.

Finally, consider the right panel of Figure 7.2 which depicts simulated social network under common tuning parameters. This network has one-hundred nodes and is broadly representative of the social networks that exist throughout the developing world; these are precisely the types of social networks that policy seeks to identify well-connected individuals to for intervention. In a network of this size, most individuals can maintain a strong social relationship with all others (Dunbar, 1992). As a consequence, it is possible that a heuristic elicitation technique that asks for the single most connected individual might converge on all members of a town identifying the same individual, in this example identified as orange.

What is the probability that two or more nominations nominate the same individual? Three features describe this probability: the size of the network, the number of individuals making nominations within that graph, and the centralization of the graph. Consider the comparative statics: the larger the graph, the smaller the probability that two ego nominate the same alter as their best connected alter; the larger the number of individuals making nominations, the larger the probability of two or more egos nominating the same alter; finally, the more centralized is the graph the more likely is a nomination from two or more egos to nominate the same alter. Rather than developing closed form descriptions of these features, later in this chapter I present simulation results.

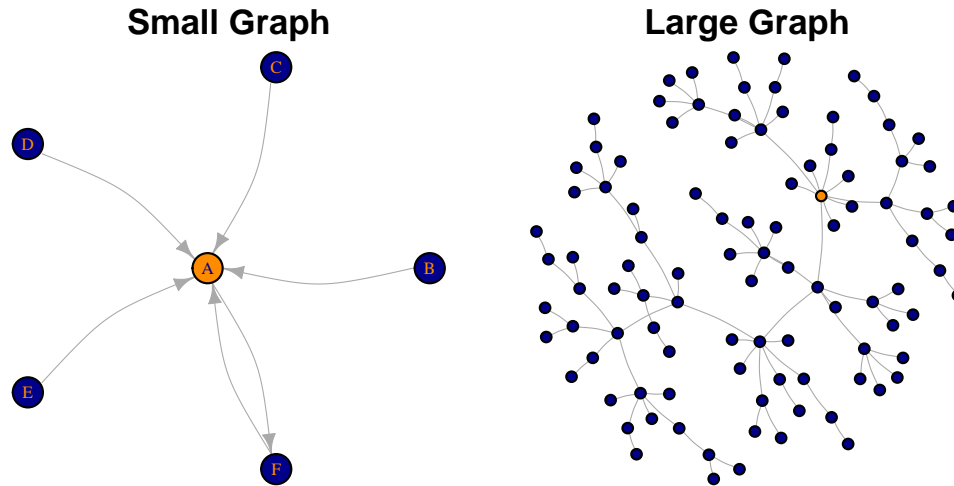


Figure 7.2: A small social network with six nodes and six connections between nodes. In this case, node A is very well connected within the network. A naive sampling strategy would identify this node as a seed in only one-in-six trials, whereas a network-informed strategy would identify node A with probability five-in-six.

7.4 Datasets

I utilize two datasets, from Honduras and Ghana, described in detail in chapter 4. The data collected in Ghana serves as the primary data source for this chapter. Key in this data collection are question about the connections between political actors. As a reminder, the research team asked individuals living in a coastal fishing town of 51 residents name generator questions consisting of:

- Are you married or living as married? If so, then who is your spouse?
- Who are your brothers and sisters you are friends with?
- Who are your best friends in the village?

As in the data generated from Honduras, subjects' answers to these questions define a *strong-tie* network of the social connections held by individuals in the social network. In addition to these name generators, subjects were asked one additional question intended to measure individuals' understanding of the structure of the social network (Friedkin, 1983; Krackhardt, 1987). This question was formed as,

- Who is your friend with the most friends?

For subjects, responding to *Who is your friend with the most friends?* requires considerably greater cognitive load than other name generator questions. Subjects are required to access knowledge not only of the set of individuals who are their own social networks, but also to access knowledge about the social network connections of these individuals. From the standpoint of individuals as limited-cognition actors, there may be concern that such high-load cognition may be either unreliable or inaccessible at the time of query. The evidence suggests that these forms of limitations are largely not-present, and the leading theories of how individuals process this information is through the use of *gossip* tokens (Banerjee et al., 2014). These gossip tokens are mnemonics stored by individuals that count the relative frequency that one subject (ego) hears another subject's name (alter). This collapses the observational task from a network-wide observation to a more local, strong-tie network observation (Banerjee et al., 2014).

7.5 Simulation Results

Figure 7.3 demonstrates through simulation how the friendship paradox works, as well as how the proposed targeting strategy can dramatically increase the performance over previous targeting strategies. For this simulation, networks of size n are created via a random network generation procedure.¹⁴ After creating networks, targets of number k are chosen at random from within these networks, with the targeting criteria varied across simulations. The baseline model is random targeting, and selects at random k nodes from the network; the *friendship paradox* targeting, or *friends of friends* targeting selects at random k nodes as initial seeds, and then draws at random one alter from this set of seeds; finally, the *best connected friend* selects at random k nodes as initial seeds, and then draws the from each of those k nodes' alters the single alter who is the best connected. Finally, after identifying target nodes, these nodes are evaluated on an objective criteria; in the case presented here this objective criteria is the total number of social connections.¹⁵ The code for these simulations is produced in the supporting material to this chapter.

The results of this simulation strongly support the benefits that might be

¹⁴The specific generative procedure is through a *preferential attachment* mechanism, detailed as a *Barabasi-Albert* model. Alternative models can (and have been) easily checked; there are very few substantive differences in this simulation.

¹⁵Alternative objectives include indegree, outdegree, eigenvector centrality, betweenness centrality, or k-core centrality, among many, many others.

realized by some form of targeting. A simple first comparison, shown in Figure 7.3, plots as an outcome variable the total degree of targeted individuals under the three targeting strategies.¹⁶ Randomly selected nodes in a network with one-hundred nodes hold slightly fewer than 6 social connections, on average; friends of the randomly selected nodes average approximately eight social connections; the best connected of these friends average more than 15 social connections. Thus, starting from a random sample of seeds and locating those seeds' best connected friends produces a set of social alters nearly three times as well-connected as could have been achieved through random selection. The second row in Figure 7.3 reports the results of the same simulation, but with a smaller social network.¹⁷ Once again, the results support the realized benefits that exist to targeting. The friendship targeting method yields 1.4 times the number of social connections and the best connected friend yields 2.3 time the number of social connections. Indeed, additional simulation results (not presented) suggest that, at least when scoring on node-level characteristics like degree, the benefits to targeting become increasingly acute as the size of the network grows. Figure 7.4 plots the same simulations of targeting, on the same network type with the same size of targets, but with the objective criteria set to eigenvector centrality. Targeting in this case produces even greater improvements, in objective criteria, with the best friend targeting generating greater than a 4.5 times increase in targets' eigenvector centrality.

7.6 Targeting Algorithm Results

To assess the performance of the best-connected friends targeting algorithm, a sensible test need be established.

In this subsection, I describe a randomization inference test that well-captures the distribution of possible outcomes to a targeting event, and compares the outcomes from the particular realization of the nomination event against the null, random, hypothesis. This procedure is common not only in the network literature where it is termed a permutation test,¹⁸ but also in the political science and applied statistics disciplines more broadly where it is termed randomization inference (e.g.,

¹⁶P-values are not reported in this section because all results are *strongly* parametric, and partially determined by the number of simulation iterations.

¹⁷This social network size is chosen because it comports with the size of the social network utilized in the empirical section of this paper.

¹⁸see, e.g. Quadratic Assignment Procedure tests

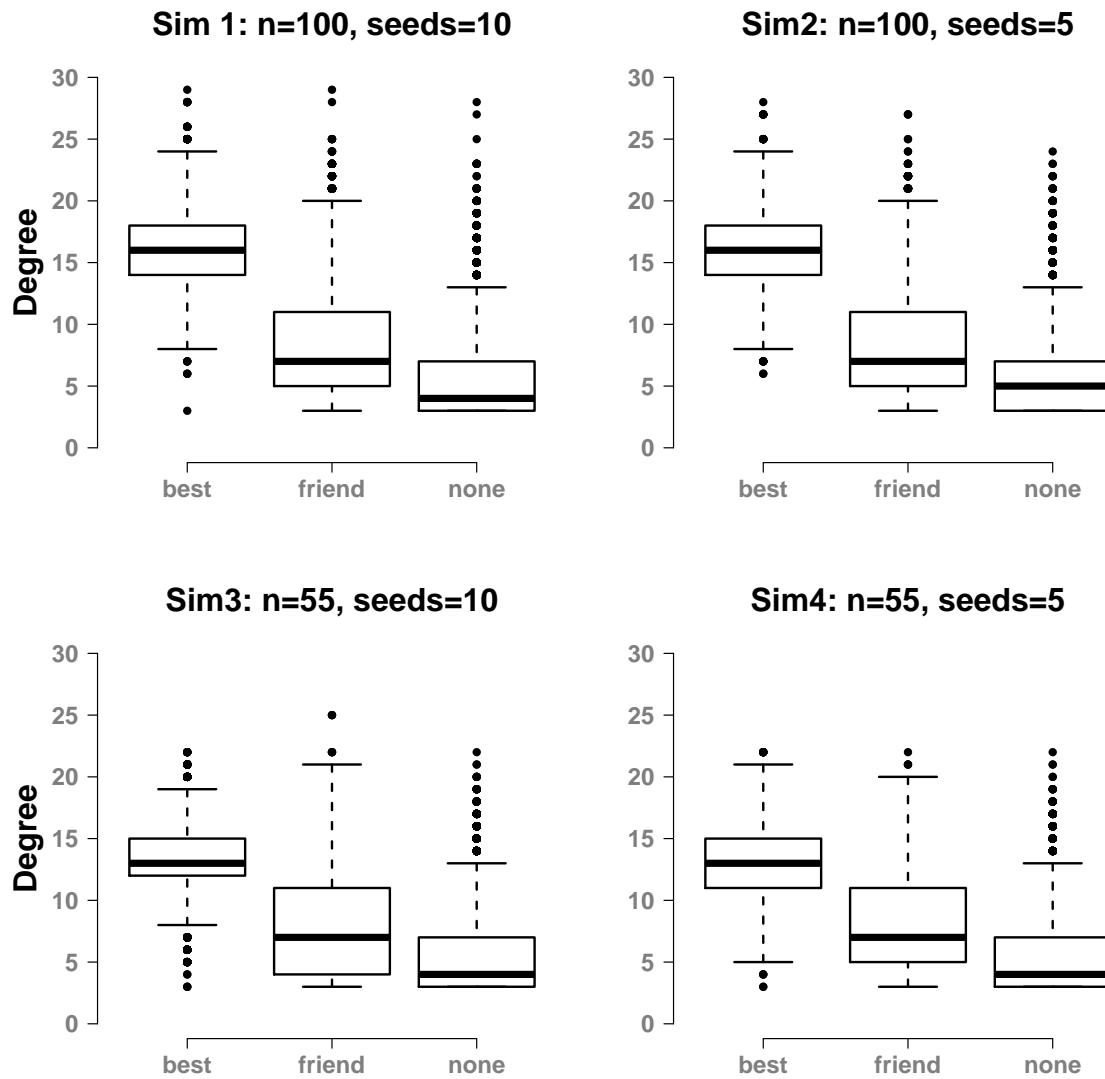


Figure 7.3: Simulations of targeting algorithm in networks of different scale and targeting densities. In each plot window, the total number of social connections of targeted individuals is reported on the y-axis, by targeting method on the x-axis. In the first row are simulated networks of size 100, and in the second row are simulated networks of size 55, the size of the network in the Ghana data.

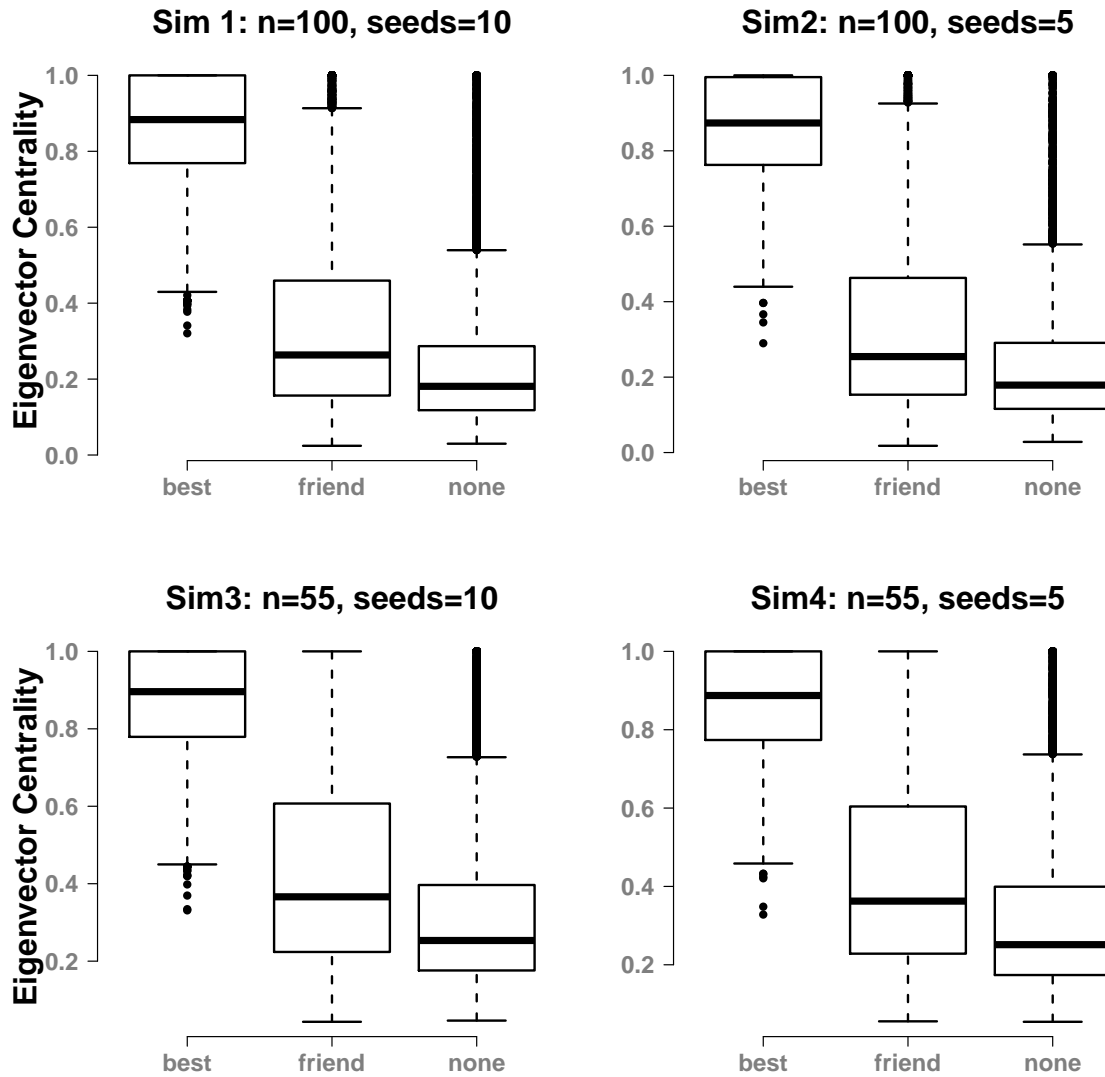


Figure 7.4: Simulations of targeting algorithm in networks of difference scale and targeting densities. In each plot window, the eigenvector centrality of targeted individual is reported on the y-axis, by targeting method on the x-axis. In the first row are simulated networks of size 100, and in the second row are simulated networks of size 55, the size of the network in the Ghana data.

Gerber and Green, 2012).

The randomization procedure that I utilize in this section holds fixed the number of nodes, n , the number of "best-connected friend" nominations, c_b , made by nodes in the graph. The randomization inference proceeds in two steps. In the first step, a sample of size c_b is drawn without replacement from set of nodes to serve as the origination of a best-connected friend nomination. In the second step, a sample of size c_b is drawn with replacement from the set of nodes. The only stipulation in the second step is that sampled nodes cannot have been sampled in the first step. This stipulation removes the possibility that an individual nominates herself as the best-connected friend. The code which implements this procedure is included in the appendix to this chapter.

7.6.1 Node Centrality

Table 7.2 presents the results of a comparison of the node-level characteristics of those nodes who were nominated as a best-connected friend compared against those who were not. Consistent with expectations and simulation data, best-friend targeting increases node-level measures of centrality. Compared to friends who were not nominated as "best-connected" friends, those who were nominated as best connected friends hold twice the number of total social connections and three times the number in indegree connections.

Figure 7.5 presents the result of an expanded comparison of the node-level characteristics of those identified by the "best-connected" targeting against a simulated set of randomly chosen individuals drawn from within the same network. In this simulation, I draw 10,000 sets of random targets, drawn from the population and calculate node-level characteristics of these sets. This comparison presents a more difficult task for the targeting mechanism, as different between this comparison and the comparison reported in Table 7.2 is that individuals who were nominated as best connected friends may also be included in the randomly selected comparison set. Consequentially, the mean degree of the comparison set increases from five when the comparison set is only those without a best-connected friend nomination to seven. Still, this evidence strongly suggests that best-connected friend targeting is out performing the random targeting.

A still more difficult task is to demonstrate that best-connected alter targeting is outperforming the simpler friendship-paradox targeting. Indeed, as I report in

Table 7.2: Means and T-test of Centrality Scores. The first two rows report mean values, and the second two rows report t-test associated scores.

Type	All Degree	Indegree	Outdegree	Eigenvector
Nominated	10.20	6.12	4.07	0.11
Non-nominated	5.21	2.12	3.09	0.09
P-value	0.00	0.00	0.03	0.24
Bonferroni P-value	0.00	0.00	0.12	0.96

Note: Rows one and two report the mean value of the targeted set scoring on the objective noted in the column heading. Rows three and four report p-values from t-tests for differences of these means.

Figure 7.6 the evidence suggests that median responses (plotted in Figure 7.6) and mean responses (not plotted in Figure 7.6) are demonstrably higher for several of the characterizations of node-level centrality. In particular, best-connected friend sets have higher total degree, driven by differences in the indegree of those nominated; there is no measurable difference in outdegree between the two methods. As well, eigenvector centrality, a measure of social connectedness that spans beyond the individual, focal, node is also measurably higher among those targeted by a best-connected friend mechanism rather than a random friend mechanism. Although the step-wise nature of the k-core centrality makes presentation of differences difficult in Figure 7.6, tests for difference in the k-core centrality of those targeted by the best-connected and random alter targeting also find greater embeddedness for those targeted by the best-connected friend strategy.

7.6.2 Network Coverage

The data suggests that those individuals who are targeted by a best-connected friend targeting are more *close* to alters in the social network than individuals who would be chosen at random. In Figure 7.7 presents the results of a comparison between the calculated closeness of individuals identified by a best-connected friend targeting, compared against a run of 10,000 samples of the same size drawn at random from the social network. The realized set of connections was closer than 92 percent of the data generated by the simulations, suggesting that the best friend targeting is out performing random selection.

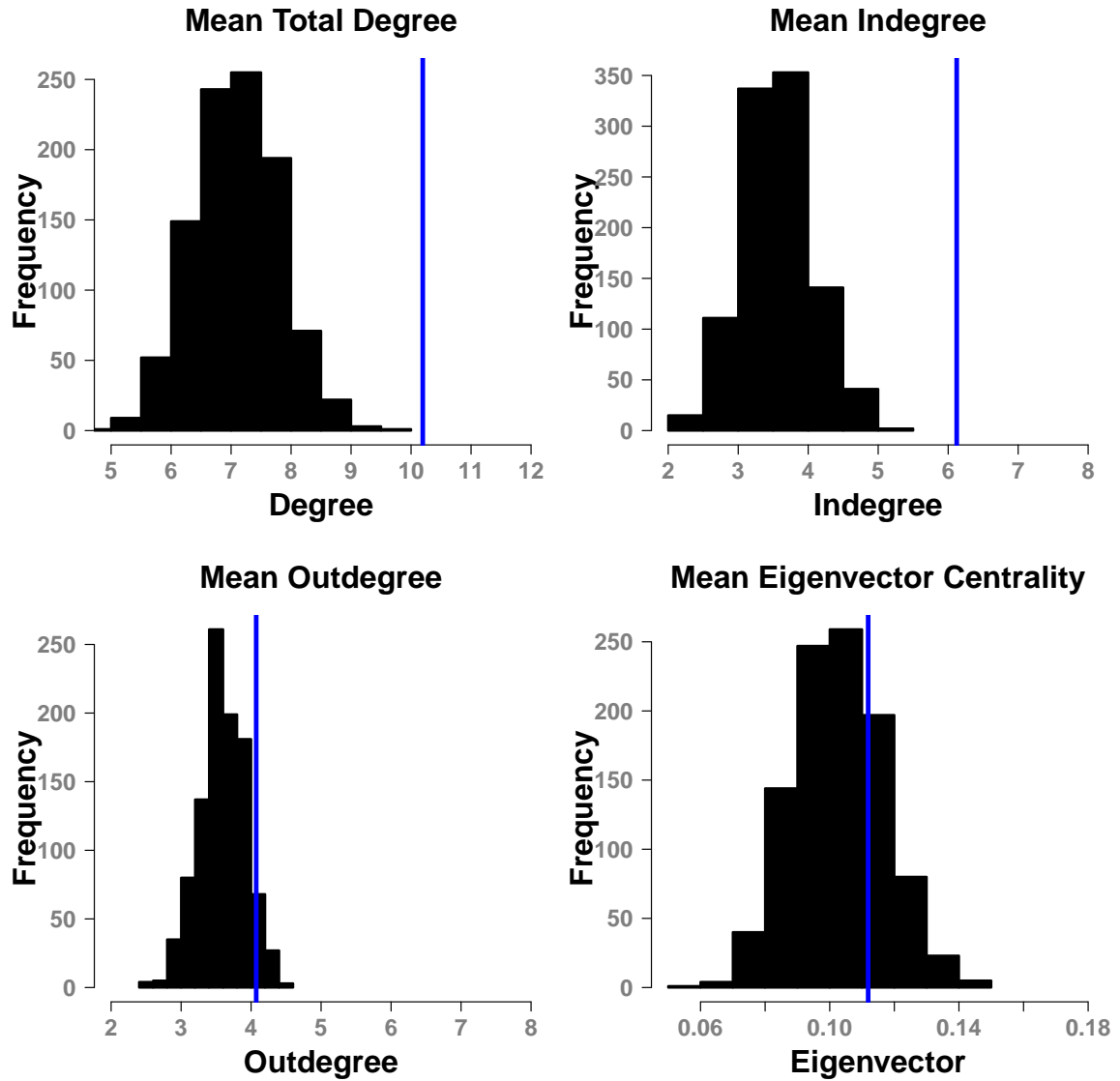


Figure 7.5: Histograms plot node-level characteristics of randomly selected individuals in the Ghana social network. Samples are run 10,000 times to produce a null distribution. Vertical blue lines plot realized "Best Connected Friend" characteristics.

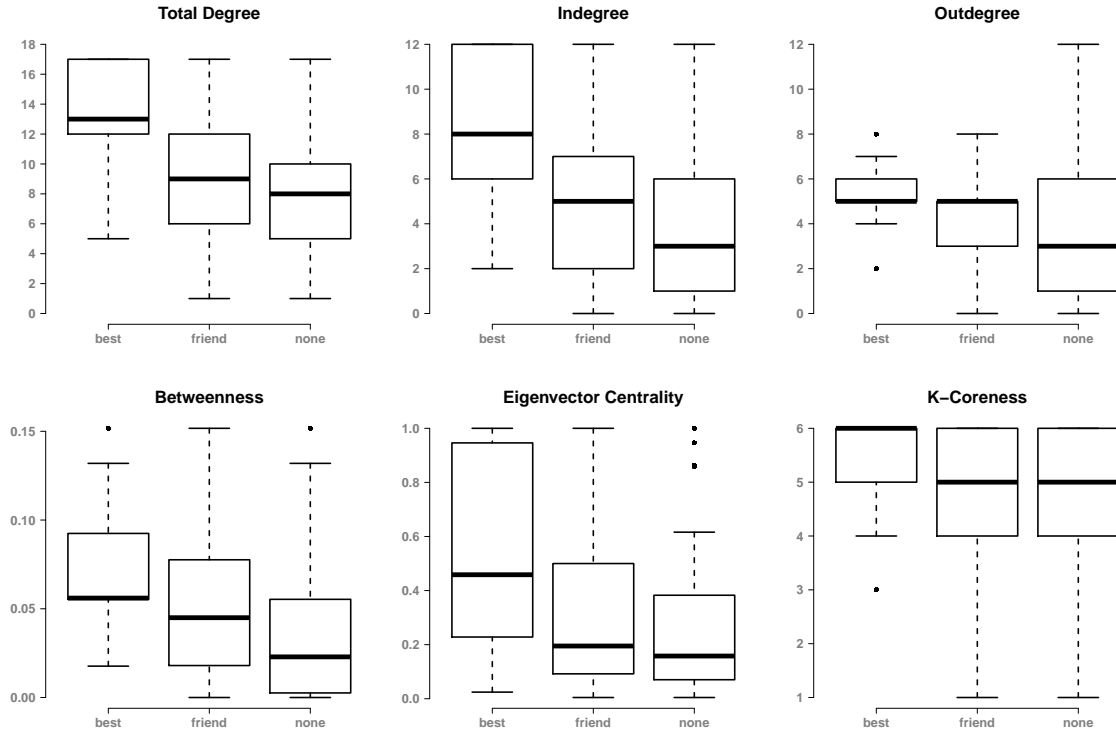


Figure 7.6: Boxplots plotting node-level connectedness measures. Each measure is noted in the window title, and drawn from bootstrapped samples of the Ghana data. In each window, "best" is best-connected friend targeting, "friend" is simple friendship-paradox targeting; and, "none" is a randomly drawn alter.

Further examining this relationship, Figure 7.8 and Figure 7.9 plot the median reach, c_r , number of social alters reached by randomly selected alters (Figure 7.8) and targeted alters (Figure 7.9). Noteworthy in these plots is that the median number of individuals reached in this network is quite similar across targeting strategies, though across strategies best-connected friend targeting identifies nodes whose median distance is one degree closer than the random-friend targeting. Additionally, consistent with similar gains made in node-level characteristics, because relatively fewer very poorly connected individuals are targeted under the friendship targeting mechanism, the variance, especially on the low-side of the distribution is relatively smaller under best-connected friend targeting than random-friend targeting. Therefore, despite concerns that best-connected friendship targeting might lead to decreased network coverage through the redundant shell covering of central nodes, these results do not produce evidence of this effect.

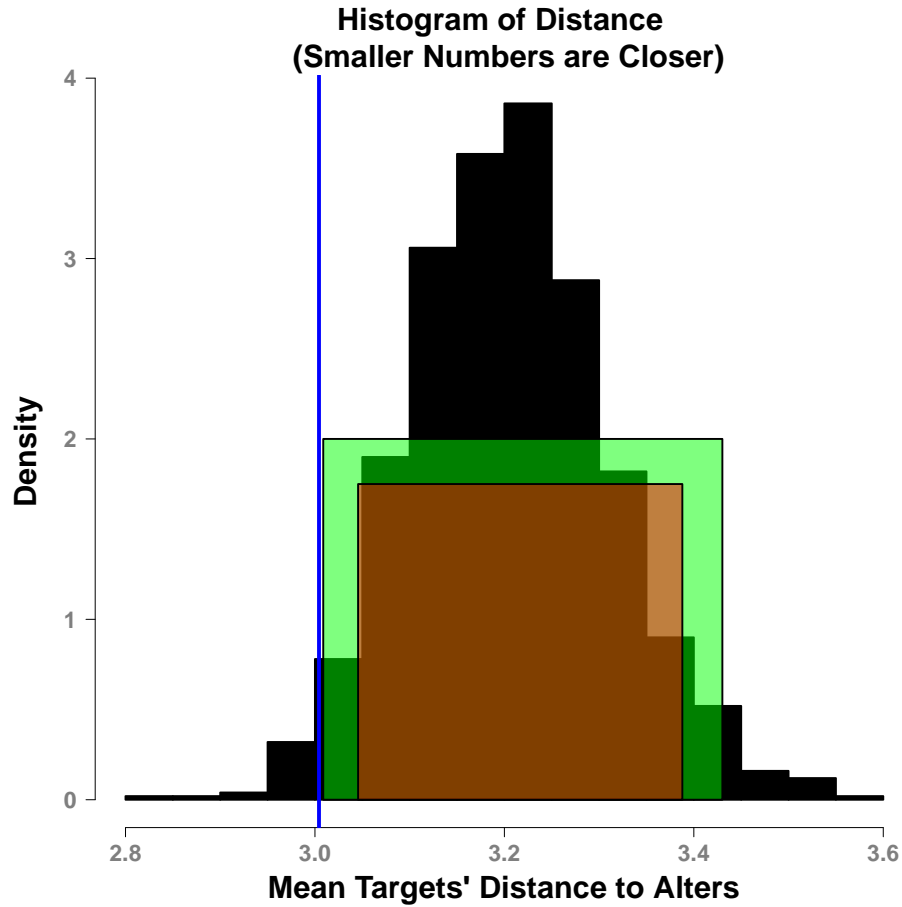


Figure 7.7: The simulated mean distances of a target set of size 22 is plotted in the black histogram. The red shaded area is 1.64 standard errors of this distribution; the green area is 1.96 standard errors of the distribution. The realized nomination set in the data is closer than 92 percent of the ties in the simulated data.

7.6.3 Variance Reduction

In addition to the improved performance of the best-connected friend targeting on first-moment characteristics, the best-connected friend targeting performs measurably better on second moments.

Consider Figure 7.10 which plots the distribution of characteristics across outcome variables and targeting methods. For every outcome, while there is no measurable difference in the mean outcome between the random friend nomination and best-connected friend nomination, there is a considerable reduction in the variance of the targeted set. In these plots, visible pure black histogram mass indicates more data inside that bin is being generated by the friendship targeting

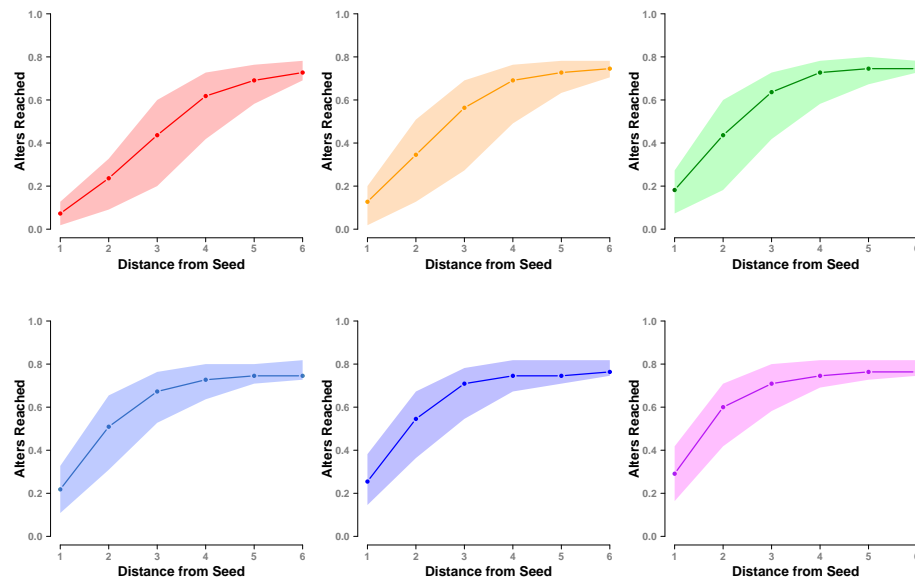


Figure 7.8: Number of alters reached by sets randomly selected seeds. Number of seeds are increased across plot cells. Along the x-axis is the social distance from a seed, and along the y-axis are the number of alters reached at that social distance. Lines and points are plotted at the median number of alters reached, and envelope is plotted at 0.025 and 0.975 of the distribution.

method than by the best-connected friend targeting method. Note that in each test, the dispersion about the first moment is greater under the friendship targeting method than by the best-connected friend targeting method. Tests for ratios of variance between the two targeting methods find that in all cases the best-connected targeting reduces variance.¹⁹ To characterize this relationship, consider the ratio of variance for Total Degree (95% CI, ratio of variance = [0.38, .49]) and indegree (95% CI, ratio of variance = [0.51, 0.65]).

7.7 Recommendations and Conclusions

In this chapter I have presented simulation evidence in support of increased performance of network-based targeting algorithms that relies on knowledge of the social network held by those who live their daily lives within this social network. I present evidence that individuals are able to meet the criteria that would be necessary for this targeting to work when actually applied. Namely, subjects are able to reasonably nominated social alters who meet the criteria that they be real social alters and also nominate alters who are, indeed, among the best socially connected of their set of alters.

Finally, I demonstrate that compared to simulation data, this best-connected friend targeting method produces two important performance improvements over either random targeting, or even friendship-paradox based targeting. Best connected friend targeting consistently identifies targets who hold node-level characteristics that are signal better social connectedness than the other targeting strategies. Furthermore, this best-connected friend targeting improves graph-level closeness of targeted nodes. And, finally, on both node-level and graph-level characterizations, the best-connected targeting technique avoids uniquely poor targets.

While these results should be replicated in a separate study-population to confirm that the effects measured in this chapter are not *sui generis* to the particular locale in Ghana, these results suggest a useful technique that might improve the targeting of intervention and monitoring in social settings.

¹⁹In all cases, the F-test p-value approaches zero.

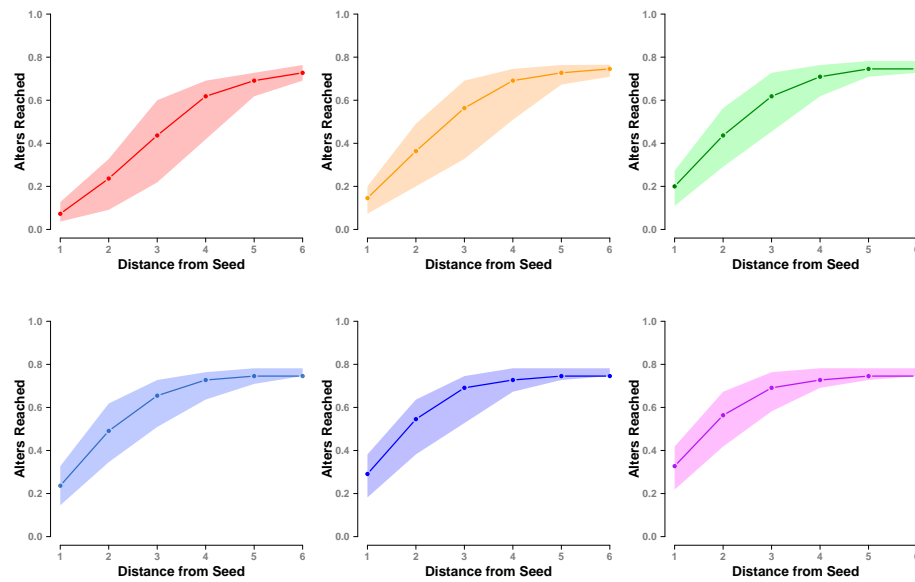


Figure 7.9: Number of alters reached by sets of targeted seeds. Number of seeds are increased across plot cells. Along the x-axis is the social distance from a seed, and along the y-axis are the number of alters reached at that social distance. Lines and points are plotted at the median number of alters reached, and envelope is plotted at 0.025 and 0.975 of the distribution.

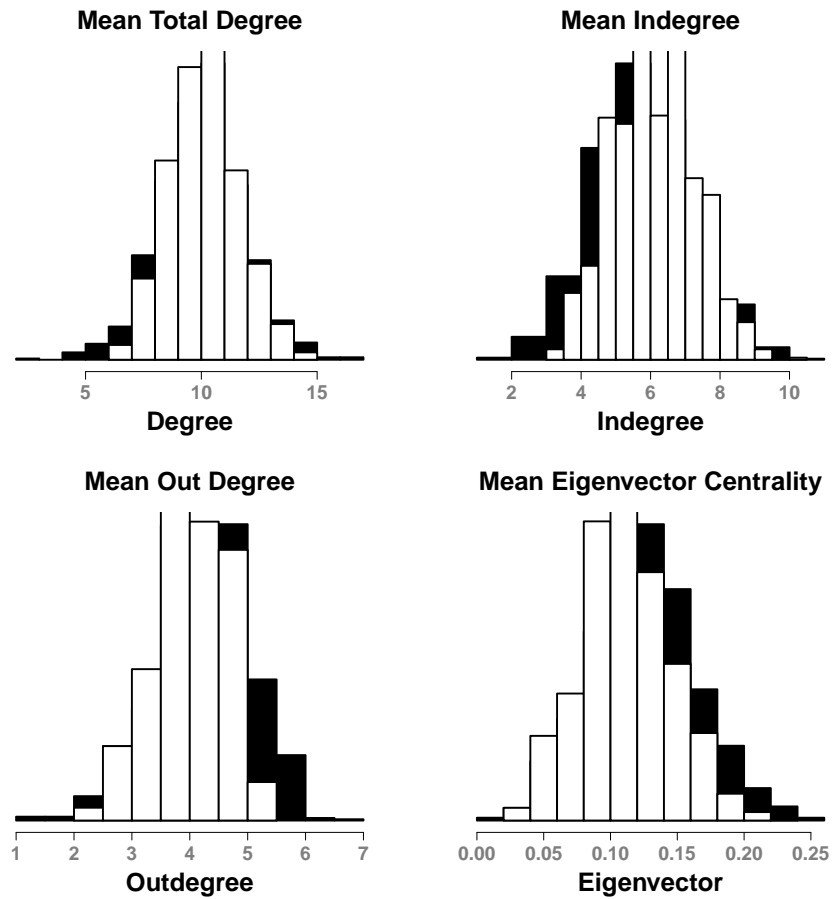


Figure 7.10: Two histogram are overlaid in each of these plots. In black are histograms of randomly selected *alters* of randomly selected individuals in the Ghana social network. Five targets are drawn for each of 10,000 simulated draws. Overlaid in white are histograms of randomly selected "best-connected" alters in the Ghana social network. Five targets are drawn for each of 10,000 simulated draws. In all cases, despite relatively small differences in central tendency, best-connected friendship targeting demonstrably reduces the variance in outcomes.

Chapter 8

Conclusion

8.1 Overview of Findings

In this dissertation I have made a general argument that when individuals take political action, they do so not only with knowledge about their own preferences, but also with some knowledge about others. I have argued that this knowledge, which I have termed *social information*, is gathered through repeated interactions between political agents as these agents move through their daily lives.

In chapter 7 I presented evidence from a set of rural fishermen and fisherwomen in Ghana. When asked to ruminate on features of their social networks these individuals were able, with high reliability, to think in terms consistent with a theory of *social information*. The results of this chapter are consonant with modern literature about individual ability for social introspection.

In chapter 5 and chapter 6 I presented evidence about two *core* political activities, mobilizing individuals to take costly action, and building a coalition of support. In both chapters, the evidence suggests a role for social information to shape outcomes. The results in these chapters make several contributions to the standing literature. They are the first to demonstrate these effects using face-to-face, individual-level social network information. They are the first to show specific paths of spillover of treatment on social networks in physical world that extend beyond

8.2 Limitations

The greatest challenge that lay between the results presented in this work and a convincing argument for the general applicability of this theory of *social information* is demonstrating that the effects observed in this data are also present when actors live in larger, more diverse social networks that are shaped by stronger formal political institutions. I have argued, in chapter 3 that evidence exists in a city council in Michigan; I have also argued that the spectre of a brokered Republican nomination for presidential candidate provides a salient example of how this *social information* theory might come to shape political outcomes in a highly-structured, formal political environment. Ultimately, however, these arguments about plausibility need to be adjudicated and tested with data.

8.3 Are People Social? Are they Political?

Recently, it is often been repeated by scholars of politics and social networks that “Man is, by nature, a ‘social animal,’ ” with the purported authority for this quote given to Aristotle. In fact, though, those who attribute this quote are inappropriately decontextualizing Aristotles’ argument, and in doing so, invert Aristotles’ conclusion.

In full, the quote reads,

“Man is by nature a social animal; an individual who is unsocial naturally and not accidentally is either beneath our notice or more than human. Society is something that precedes the individual. Anyone who either cannot lead the common life or is so self-sufficient as not to need to, and therefore does not partake of society, is either a beast or a god.”

– Aristotle, *Politics*.

Aristotle here is arguing that individuals are reactive elements of the social system; at least, the individuals that Aristotle believes are of political concern. To Aristotle, the distinction is not that everyone is necessarily aware of, and reactive to a social environment. Instead, he makes a more careful statement. First, and seemingly uncontroversially, Aristotle claims that that society logically precedes the individual. At the time of inception a person becomes a part of her society and her group. This claim serves as context for the development of the person into the group, and also implicitly makes a statement of equality concerning people at birth. All,

regardless of to whom they are born, are born as a part of society. To Aristotle, the social environment and the political environment are closely enmeshed; the political environment is one particular expression of a more general social environment. And so, just as everyone is born into a social environment on an equal footing; so too is everyone born into the political environment on an equal footing.¹

That everyone is born into a political world leads to Aristotles' second, more frequently cited and more controversial claim: those who choose to exist outside the social environment are choosing to remove themselves from the political environment as well. To be clear, Aristotle is not claiming that being a person implies that the person is purpose-built for political action in the same way that a gazelle is purpose built for running. Instead, Aristotle's claim is that people who choose to remove themselves from the society are also choosing to remove themselves from consideration for political activity. This distinction is important, because under this reading it is not the people are by their nature political; but instead that they cannot be political without also being social.

A final point on this matter. If, as Aristotle argues, citizens who fail to engage as a part of society are poorly equipped to participate in and undertand politics—that they are either beasts or gods—then as well, the scholar who fails to fully acknowledge not only the role of the individual but also the social is poorly equipped to understand political behavior.

¹One's eventual stature in this political environment is certainly not ensured in Aristotles' conception. Indeed, this conception, while placing everyone in an equal position to take political action, does not ensure that all individuals will have an equal representation within the political system.

Chapter 9

Resources Utilized in this Work

In this chapter are included resources such as experiment language and questionnaire wording.

9.1 Mobilization Script in English

You have been selected for a special role in our investigation this week. On [–] day at [–] time there will be a town meeting at [–] building. In this meeting we will be discussing with your community the future of microfinance in this region. It is important that we have as many people as possible at this meeting so that your community is well-represented. We don't have the time to convince everybody in the town to go to this meeting, so we are selecting certain individuals to inform the community. You have been selected as one of these people for your community. This means that we want you to get as many people as you can to come to the meeting. Because we want as many people as possible at this meeting, each person selected for this role has the chance to win a cash prize. Everybody who comes to the meeting will be asked who is responsible for them being there. For each person that names you as the person who got them to come, we will put one ticket with your name on it into a drawing for the cash prize. Therefore, the greater the amount of people you get to come to the meeting and identify you as the person who influenced them to come, the greater chance you have at winning the cash prize. Your goal should then be to get as many people you can to come to the meeting. It is important not only that they come to the meeting, but also that they must identify you as the reason that they came so that we can put your name on the raffle ticket. You may convince them to come by explaining that at the meeting we will be discussing the future of microfinance in your region, or however else you would like.

9.2 Mobilization Script in Spanish

Usted ha sido seleccionado/seleccionada para tener un rol especial en nuestro estudio esta semana. El día [] de [] a las [] vamos a tener una reunión en esta comunidad en este [] edificio. En esta reunión vamos a platicar con su comunidad sobre el futuro de microfinanza en la región. Es muy importante para nosotros que haya el máximo de personas posibles en esta reunión para que se le de una justa representación a su comunidad. Nosotros no tenemos el tiempo de poder convencer a todos los de la comunidad a que vayan a esta reunión, es por eso que estamos seleccionando a ciertos individuos para que le informen a la comunidad. Esto significa que queremos que usted convenza a todas las personas que pueda para que atiendan esta reunión. Por el hecho de que queremos que todas las personas posibles vayan a esta reunión, cada persona seleccionada para este rol especial va a tener la oportunidad de ganarse un teléfono celular o 200 Lempiras de crédito para su teléfono. A cada persona que venga a la reunión se le va a preguntar quien fue la que le dijo que atienda la reunión. Por cada persona que lo/la nombre como la persona que le recomendó atender, vamos a poner un boleto con su nombre en un sorteo para ganarse el celular o los minutos de tiempo aire para su celular. Entonces, entre más personas vengan a la reunión y lo/la identifiquen como la persona que les dijo que atiendan, más oportunidades va a tener usted de poder ganar estos premios. Su meta entonces debería de ser la de juntar a todas las personas posibles para la reunión. No solo es importante que vengan a la reunión, sino que también es de gran importancia que ellos sean capaces de identificarlo/ identificarla como la razón por la cual atendieron a la reunión para poder poner su nombre en el sorteo de los premios. Usted podría convencerlos a que atiendan a la reunión si les explica que vamos a estar platicando acerca del futuro de microfinanza en la región o de la manera que usted guste.

9.3 Election Script in English

Based on some of the initial discussions, we wanted to see who might make a good representative to a microcredit company, please choose who you would want this to be based upon the people who have been chosen.

9.4 Election Script in Spanish

Basándonos en algunas de las discusiones iniciales, nosotros queríamos ver quien podría ser un buen representante de una compañía de microcrédito. Por favor escojan al candidato que gusten basándose en las personas que han sido escogidas.

Bibliography

- Acemoglu, Daron and Matthew O Jackson. 2015. "History, Expectations, and Leadership in the Evolution of Social Norms." *Review of Economic Studies* 82(2):423–456.
- Achen, C H. 1975. "Mass political attitudes and the survey response." *The American Political Science Review* 69(4):1218–1231.
- Alexander, C, M Piazza, D Mekos and T W Valente. 2001. "Peer networks and adolescent cigarette smoking: An analysis of the national longitudinal study of adolescent health." *Journal of Adolescent Health* 29:22–30.
- Almond, Gabriel A and Sidney Verba. 1965. *The Civic Culture*. Boston: Little, Brown and Company.
- Andreoni, James and Ragan Petrie. 2004. "Public goods experiments without confidentiality: a glimpse into fund-raising." *Journal of Public Economics* 88(7-8):1605–1623.
URL: <http://linkinghub.elsevier.com/retrieve/pii/S0047272703000409>
- Apicella, C L, F W Marlowe, J H Fowler and N A Christakis. 2012. "Social networks and cooperation in hunter-gatherers." *Nature* 481(7382):497–501.
- Aral, S and D Walker. 2012. "Identifying influential and susceptible members of social networks." *Science* 337(6092):337–341.
- Baird, Sarah, Aislinn Bohren, Craig McIntosh and Berk Ozler. 2014. "Designing Experiments to Measure Spillover Effects." *World Bank Policy Research Working Paper* 6824(March).

- Balkundi, Prasad and Martin Kilduff. 2006. "The ties that lead: A social network approach to leadership (DOI:10.1016/j.leaqua.2005.09.004)." *Leadership Quarterly* 17(4):419–439.
- Banerjee, Abhijit, Arun G Chandrasekhar, Esther Duflo and Matthew O Jackson. 2013. "The Diffusion of Microfinance." *Science* 341(6144):363–370.
- Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo and Matthew O. Jackson. 2014. "Gossip: Identifying Central Individuals in a Social Network." *National Bureau of Economic Research* p. 29.
URL: <http://arxiv.org/abs/1406.2293>
- Bartels, L M. 2008. "The Study of Electoral Behavior.".
URL: <https://www.princeton.edu/~bartels/electoralbehavior.pdf>
- Bartels, Larry M. 1993. "Messages received: The political impact of media exposure." *American Political Science Review* pp. 267–285.
- Bartels, Larry M. 2005. "Homer Gets a Tax Cut: Inequality and Public polity in the American Mind." *Perspective on Politics* 3:15–32.
- Bartels, Larry M. 2006. Priming and Persuasion in Presidential Campaigns. In *Capturing campaign effects*, ed. Henry E Brady and Richard Johnston. Ann Arbor: University of Michigan Press p. 78.
- Baumeister, Roy F and Mark R Leary. 1995. "Baumeister & Leary (1995). The Need to Belong–Desire for Interpersonal Attachments as a fundamental human emotion.pdf." *Psychological Bulletin* 117(3):497–529.
- Beaman, Lori, Ariel Benyishay, Coll William and Ahmed Mushfiq Mobarak. 2014. "Can Network Theory based Targeting Increase Technology Adoption ?".
- Bearman, P S, J Jones and J R Udry. 2000. "The National Longitudinal Study of Adolescent Health: Research Design.".
URL: <http://www.cpcp.unc.edu/projects/addhealth/design.html>
- Beck, Paul Allen, Russell J Dalton, Steven Greene and Robert Huckfeldt. 2002. "The social calculus of voting: Interpersonal, media, and organizational influences on presidential choices." *American Political Science Review* 96(01):57–73.

- Berelson, Bernard R, Paul F Lazarsfeld and William N McPhee. 1954. *Voting: A Study of Opinion Formation in a Presidential Campaign*. The University of Chicago Press.
- Boksem, M A S, E Kostermans, B Milivojevic and D De Cremer. 2011. "Social status determines how we monitor and evaluate our performance." *Social Cognitive and Affective Neuroscience* .
- Bond, Robert M, Christopher J Fariss, Jason J Jones, Adam D I Kramer, Cmeron Marlow, Jaime E Settle and James H Fowler. 2012. "A 61-Million-Person Experiment in Social Influence and Political Mobilization." *Nature* 489:295–298.
- Bowers, Jake, Mark M. Fredrickson and Costas Panagopoulos. 2013. "Reasoning about interference between units: A general framework." *Political Analysis* 21(1):97–124.
- Boyd, Robert and Peter J Richerson. 1985. "Culture and the Evolutionary Process." *The Journal of Nervous and Mental Disease* 175(2):331.
URL: <http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0226069338>
- Brader, Ted. 2005. "Striking a Responsive Chord: How Political Ads Motivate and Persuade Voters by Appealing to Emotions." *American Journal of Political Science* 49(2):388–405.
- Brady, H E, S Verba and K L Schlozman. 1995. "Beyond SES: A resource model of political participation." *American Political Science Review* 89(2):271–294.
- Brewer, Devon D. 2000. "Forgetting in the recall-based elicitation of personal and social networks." *Social Networks* 22(1):29–43.
- Brewer, Devon D. and Cynthia M. Webster. 2000. "Forgetting of friends and its effects on measuring friendship networks." *Social Networks* 21(4):361–373.
- Buckholtz, J W, C L Asplund, P E Dux, D H Zald, J C Gore, O D Jones and R Marois. 2008. "The neural correlates of third-party punishment." *Neuron* 60(5):930–940.
- Campbell, Angus, Philip E Converse, Warren E Miller and Donald E Stokes. 1960. *The American Voter*. Survey Research Center, University of Michigan.

- Cardenas, Juan-Camilo. 2000. "How Do Groups Solve Local Commons Dilemmas? Lessons from Experimental Economics in the Field." *Environment, Development and Sustainability* 2:305–322.
- Christakis, N. a. and J. H. Fowler. 2014. "Friendship and natural selection." *Proceedings of the National Academy of Sciences* 2014:1–6.
URL: <http://www.pnas.org/cgi/doi/10.1073/pnas.1400825111>
- Christakis, Nicholas A and James H Fowler. 2010. "Social network sensors for early detection of contagious outbreaks." *PloS one* 5(9):e12948.
- Christakis, Nicholas a. and James H. Fowler. 2013. "Social contagion theory: examining dynamic social networks and human behavior." *Statistics in Medicine* 32(4):556–577.
- Christakis, Nicholas and James H Fowler. 2007. "The spread of obesity in a large social network over 32 years." *New England Journal of Medicine* 357(4):370–379.
- Christakis, Nicholas and James H Fowler. 2008. "The collective dynamics of smoking in a large social network." *New England journal of medicine* 358(21):2249–2258.
- Converse, Philip E. 2006. "The Nature of Belief Systems in Mass Publics (1964)." *Critical Review* 18:1–74.
- Cox, Gary W. 1997. *Making votes count*. Cambridge University Press.
- Cox, Gary W., Frances M. Rosenbluth and Michael F. Thies. 1998. "Mobilization, Social Networks, and Turnout: Evidence from Japan." *World Politics* 50(3):447–474.
- Cruz, Cesi, Julien Labonne and Pablo Querubin. 2015. "Politician Family Networks and Electoral Outcomes: Evidence from the Philippines *." *American Economic Review (Revise and Resubmit)* .
- Csányi, Gábor and Balázs Szendroi. 2004. "Structure of a large social network." *Physical Review E* 69(3):36131.
- Dasgupta, Partha and Ismail Serageldin. 2001. *Social capital: A Multifaceted Perspective*. World Bank Publications.
- Downs, Anthony. 1957. "An economic theory of political action in a democracy." *The Journal of Political Economy* 65(2):135–150.

- Dunbar, R.I.M. 1992. "Neocortex size as a constraint on group size in primates." *Journal of Human Evolution* 22(6):469–493.
- Edgar, E, G Martin, Tamar Kugler, Edgar E Kausel, C E S Ifo and W Orking P Aper. 2012. "Review of Interactive Decision Making in Groups.".
- Eldersveld, Samuel J. 1956. "Experimental Propaganda Techniques and Voting Behavior." *The American Political Science Review* 50(1):154–165.
- Enemark, Daniel, Mathew D McCubbins and Nicholas Weller. 2012. "Knowledge and networks: An experimental test of how network knowledge affects coordination." *Social Networks* .
- Fafchamps, Marcel, Ana Vaz and Pedro C Vicente. 2013. "Voting and Peer Effects : Experimental Evidence from Mozambique." (0).
- Feld, Scott L. 1991. "Why your friends have more friends than you do." *American Journal of Sociology* pp. 1464–1477.
- Finkel, Steven E. 1993. "Reexamining the 'minimal effects' model in recent presidential campaigns." *Journal of Politics* 55(1):1–21.
- Fowler, J H and N A Christakis. 2010. "Cooperative behavior cascades in human social networks." *Proceedings of the National Academy of Sciences* 107(12):5334.
- Fowler, James. 2005. "Turnout in a small world." *Temple University Press, Philadelphia* pp. 269–287.
- Fowler, James H and Darren Schreiber. 2008. "Biology, politics, and the emerging science of human nature." *Science* 322(5903):912.
- Fowler, James H and Nicholas a Christakis. 2008. "Estimating Peer Effects on Health in Social Networks." *Journal of Health Economics* 27(5):1400–1405.
URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1162213
- Freedman, Paul, Michael Franz and Kenneth Goldstein. 2004. "Campaign advertising and democratic citizenship." *American Journal of Political Science* 48(4):723–741.
- Freeman, Linton. 1977. "A set of measures of centrality based on Betweenness." *Sociometry* 40(1):35–41.

- Freeman, Linton C. 1978. "Centrality in Social Networks Conceptual Clarification." *Social Networks* 1:215–239.
- Friedkin, Noah E. 1983. "Horizons of Observability and Limits of Informal Control in Organizations." *Social Forces* 62(1):54–77.
- Friedkin, Noah E. 1998. *A structural theory of social influence. Structural analysis in the social sciences*. Cambridge: Cambridge University Press.
- Garcia-Herranz, Manuel, Esteban Moro, Manuel Cebrian, Nicholas a Christakis and James H Fowler. 2014. "Using friends as sensors to detect global-scale contagious outbreaks." *PloS one* 9(4):e92413.
URL: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0092413>
- Gee, Laura K and Jason Jones. 2016. "Social Networks and Labor Markets: How Strong Ties Relate to Job Transmission On Facebook's Social Network." *Journal of Labor Economics* 0:1–55.
- Gerber, A S and D P Green. 2000. "The effects of canvassing, telephone calls, and direct mail on voter turnout: A field experiment." *American Political Science Review* 94(3):653–663.
- Gerber, A S, D P Green and C W Larimer. 2008. "Social pressure and voter turnout: Evidence from a large-scale field experiment." *American Political Science Review* 102(01):33–48.
- Gerber, Alan S and Donald P Green. 2012. *Field Experiments: Design, Analysis, and Interpretation*. WW Norton.
- Gerber, Alan S., Gregory a. Huber, David Doherty, Conor M. Dowling and Seth J. Hill. 2013. "Do Perceptions of Ballot Secrecy Influence Turnout? Results from a Field Experiment." *American Journal of Political Science* 57(3):537–551.
URL: <http://doi.wiley.com/10.1111/ajps.12019>
- Gerber, Alan S, James G Gimpel, Donald P Green and Daron R Shaw. 2011. "How Large and Long-lasting Are the Persuasive Effects of Televised Campaign Ads? Results from a Randomized Field Experiment." *American Political Science Review* 105(1):135–150.

- Goette, L, D Huffman and S Meier. 2006. "The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups." *The American economic review* 96(2):212–216.
- Gonçalves, Bruno, Nicola Perra and Alessandro Vespignani. 2011. "Modeling users' activity on twitter networks: Validation of Dunbar's number." *PLoS ONE* 6(8).
- Gosnell, Harold F. 1927. *Getting Out the Vote: An Experiment in the Stimulation of Voting*. Chicago: University of Chicago Press.
- Hafner-Burton, Emilie M., D. Alex Hughes and David G. Victor. 2013. "The Cognitive Revolution and the Political Psychology of Elite Decision Making." *Perspectives on Politics* 11(02):368–386.
URL: http://www.journals.cambridge.org/abstract_S1537592713001084
- Henrich, Natalie and Joseph Henrich. 2007. *Why Humans Cooperate: A cultural and Evolutionary Explanation*. New York: Oxford University Press.
- Hillygus, D Sunshine and Simon Jackman. 2003. "Voter Decision Making in Election 2000: Campaign Effects, Partisan Activation, and the Clinton Legacy." *American Journal of Political Science* 47(4):583–596.
- Huckfeldt, R and J Sprague. 1995. *Citizens, politics, and social communication: Information and influence in an election campaign*. Cambridge University Press New York.
- Huckfeldt, R R, P E Johnson and J D Sprague. 2004. *Political disagreement: The survival of diverse opinions within communication networks*. Cambridge University Press.
- Huckfeldt, Robert. 1979a. "Political Participation and the Neighborhood Social Context." *American Journal of Political Science* 23:579–592.
- Huckfeldt, Robert. 1979b. "Political Participation and the Neighborhood Social Context." *American Journal of Political Science* 23:579–592.
- Huckfeldt, Robert and John Sprague. 1987. "Networks in Political Context: The Social Flow of Political Information." *American Political Science Review* 81:1198–1216.

- Hudgens, Michael G and M Elizabeth Halloran. 2008. "Toward Causal Inference With Interference." *Journal of the American Statistical Association* 103(482):832–842.
- Illenberger, Johannes and Gunnar Flötteröd. 2012. "Estimating network properties from snowball sampled data." *Social Networks* 34(4):701–711.
URL: <http://linkinghub.elsevier.com/retrieve/pii/S0378873312000548>
- Izuma, K, D N Saito and N Sadato. 2008. "Processing of social and monetary rewards in the human striatum." *Neuron* 58(2):284–294.
- Johnson, Garrett A, Randall A Lewis and Elmar I Nubbemeyer. 2015. "Ghost Ads: A Revolution in Measuring Ad Effectiveness." pp. 1–36.
- Jones, Jason J, Jaime E Settle, Robert M Bond, Christopher J Fariss, Cameron Marlow and James H Fowler. 2013. "Inferring Tie Strength from Online Directed Behavior." *PLoS One* 8(1):e52168.
- Kandel, Denise B. 1978. "Homophily , Selection , and Socialization in Adolescent Friendships Author (s): Denise B . Kandel Source : American Journal of Sociology , Vol . 84 , No . 2 (Sep . , 1978) , pp . 427-436 Published by : The University of Chicago Press Stable URL : <http://>." *American Journal of Sociology* 84(2):427–436.
- Katz, Elihu. 1957. "The two-step flow of communication: An up-to-date report on an hypothesis." *Public Opinion Quarterly* 21(1):61–78.
- Katz, Elihu and Paul F Lazarsfeld. 1955. *Personal Influence; the Part Played by People in the Flow of Mass Communications*. Glencoe, IL: Free Press.
- Kenny, Christopher B. 1992. "Political participation and effects from the social environment." *American Journal of Political Science* pp. 259–267.
- Klofstad, Casey A, Anand Edward Sokhey and Scott D McClurg. 2013. "Disagreeing about disagreement: How conflict in social networks affects political behavior." *American Journal of Political Science* 57(1):120–134.
- Kousser, Thad. 2005. *Term limits and the dismantling of state legislative professionalism*. Cambridge University Press.
- Krackhardt, David. 1987. "Cognitive social structures." *Soical Networks* 9:109–134.

- Krauss, Robert. 1981. Impression Formation, Impression Management, and Non-Verbal Behaviors. In *Social Cognition*, ed. E Tory Higgins, C Peter Herman and Mark P Zanna. Vol. 1 Erlbaum: Hillsdale.
- Kuklinski, J H, R C Luskin and J Bolland. 1991. "Where is the schema? Going beyond the "s" word in political psychology." *The American Political Science Review* 85(4):1341–1356.
- LaCour, Michael J and Donald P Green. 2014. "When contact changes minds: An experiment on transmission of support for gay equality (RETRACTED)." *Science* p. 1366.
- Lazarsfeld, Paul F., Bernard Berelson and Hazel Gaudet. 1944. *The people's choice: How the voter makes up his mind in a presidential election*. New York: Duell, Sloan and Pearce.
- Lazarsfeld, Paul F, Bernard Berelson and Hazel Gaudet. 1968. *The People's Choice: How the Voter Makes up his Mind in a Presidential Campaign*. New York: Columbia University Press.
- LeVeck, Brad L. 2013. "Coordination, learning, and partisan models of the economy." *Unpublished Manuscript* .
- Lewens, Tim. 2013. *Cultural Evolution*. Stanford, CA: Center for the Study of Language and Information.
- Lieberman, M D. 2007. "Social cognitive neuroscience: a review of core processes." *Annu. Rev. Psychol.* 58:259–289.
- Lieberman, MA. 2013. *Social*. New York: Crown Publishing.
- Lieberman, Matthew D. 2010. "Social cognitive neuroscience." *Handbook of social psychology* pp. 143–193.
URL: <http://dx.doi.org/10.1002/9780470561119.socpsy001005>
- Lieberman, Matthew D, Darren Schreiber and Kevin N Ochsner. 2003. "Is political cognition like riding a bicycle? How cognitive neuroscience can inform research on political thinking." *Political Psychology* 24(4):681–704.
- Lijphart, Arend. 2012. *Patterns of Democracy*. New Haven, CT: Yale University Press.

- Lim, C and R D Putnam. 2010. "Religion, Social Networks, and Life Satisfaction." *American Sociological Review* 75(6):914.
- Lin, Nan. 2001. "Social Capital:: A Theory of Social Structure and Action."
- List, John A. 2011. "Does Market Experience Eliminate Market Anomalies ? Author (s): John A . List Source : The Quarterly Journal of Economics , Vol . 118 , No . 1 (Feb ., 2003), pp . 41-71 Published by : The MIT Press Stable URL : <http://www.jstor.org/stable/25053898> . Y." 118(1):41–71.
- Lodge, Milton and Charles S Taber. 2005. "The Automaticity of Affect for Political Leaders, Groups, and Issues: An Experimental Test of the Hot Cognition Hypothesis." *Political Psychology* 26(3):455–482.
URL: <http://dx.doi.org/10.1111/j.1467-9221.2005.00426.x>
- Lucking-Reiley, David. 1999. "Using field experiments to test equivalence between auction formats: Magic on the internet." *American Economic Review* 89(5):1063–1080.
- Mackie, Gerry. 1996. "Ending Footbinding and Infibulation: A Convention Account." *American Sociological Review* 61:999–1017.
- Mailath, George and Larry Samuelson. N.d. *Repeated Games and Reputations*. Oxford University Press.
- Marsden, P V and J Podolny. 1990. Dynamic Analysis of Netowrk Diffusion PROcesses. In *Social Networks Through Time*, ed. J Weesie and H Flap. The Netherlands: ISOR: Utrecht.
- Maslow, Abraham Har. 1943. "A theory of Human Motivation." *Psychological Review* 50(4):470.
- McClurg, Scott D. 2004. "Indirect Mobilization: The Social Consequences of Party Contacts in an Election Campaign." *American Politics Research* 32(4):406–443.
- McClurg, Scott D. 2006. "The electoral relevance of political talk: Examining disagreement and expertise effects in social networks on political participation." *American Journal of Political Science* 50(3):737–754.

- McCubbins, Mathew D and Thomas Schwartz. 1984. "Congressional Oversight Overlooked: Police Patrols versus Fire Alarms." *American Journal of Political Science* 28(1):165–179.
- McDermott, Rose. 2011. "New Directions for Experimental Work in International Relations." *International Studies Quarterly* 55:503–520.
- Mckiernan, Kristen A, Jacqueline N Kaufman, Jane Kucera-thompson and Jeffrey R Binder. 2003. "A Parametric Manipulation of Factors Affecting Task-induced Deactivation in Functional Neuroimaging." *Journal of Cognitive Neuroscience* 15(3):394–408.
- Mcpherson, Miller, Lynn Smith-lovin and James M Cook. 2016. "Birds of a Feather : Homophily in Social Networks Author (s): Miller McPherson , Lynn Smith-Lovin and James M . Cook Published by : Annual Reviews Stable URL : <http://www.jstor.org/stable/2678628> Your use of the JSTOR archive indicates your acceptance o." 27(2001):415–444.
- Mehra, Ajay, Andrea L. Dixon, Daniel J. Brass and Bruce Robertson. 2006. "The Social Network Ties of Group Leaders: Implications for Group Performance and Leader Reputation." *Organization Science* 17(1):64–79.
- Milgram, S. 1974. *Obedience to authority: An experimental view*. Taylor & Francis.
- Mutz, Diana C. 2002. "The Consequences of Cross-Cutting Networks for Political Participation." *American Journal of Political Science* 46(4):838–855.
- Neisser, Ulric. 1980. "On" Social Knowing"." *Personality and Social Psychology Bulletin* 6(4):601.
- Newman, M. E J. 2005. "A measure of betweenness centrality based on random walks." *Social Networks* 27(1):39–54.
- Newman, Mark E J. 2003. "2 Random graphs as models of networks." *Handbook of graphs and networks* p. 35.
- Nickerson, David W. 2008. "Is voting contagious? Evidence from two field experiments." *American Political Science Review* 102(1):49.

- Noel, Hans and Brendan Nyhan. 2011. "The "Unfriending" Problem: The Consequences of Homophily in Friendship Retention for Causal Estimates of Social Influence.".
- Nowak, Martin A and Karl Sigmund. 2005. "Evolution of indirect reciprocity." *Nature* 437:1291–1298.
- Olson, M. 1965. *The logic of collective action*. Cambridge.
- Popkin, Samuel. 1994. *The Reasoning Voter*. University of Chicago Press.
- Pruitt, Dean G. 1971. "Conclusions: Toward an understanding of choice shifts in group discussion." *Journal of Personality and Social Psychology* 20(3):495–510.
- Putnam, R D. 1995. "Bowling alone: America's declining social capital." *Journal of democracy* 6:65–78.
- Putnam, Rober. 1993. *Making Democracy Work: Civic Traditions in Modern Italy*. University of Michigan Press.
- Richard F. Fenno, Jr. 1977. "U.S. House Members in Their Constituencies: An Exploration." *The American Political Science Review* 71(3):883–917.
- Riker, W H and P C Ordeshook. 1968. "A Theory of the Calculus of Voting." *The American Political Science Review* 62(1):25–42.
- Robinson, John P. 1976. "Interpersonal Influence in Election Campaigns: Two Step-Flow Hypotheses." *The Public Opinion Quarterly* 40(3):304–319.
URL: <http://www.jstor.org/stable/2748249>
- Rolfe, Meredith. 2012. *Voter Turnout*. New York: Cambridge University Press.
- Rosenstone, S J and J M Hansen. 1993. *Mobilization, participation, and democracy in America*. New York: Macmillan Puublishing Company.
- Rubin, Donald B. 1974. "Estimating Causal Effects of Treatment In Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66(5):688–701.
- Ruhnau, Britta. 2000. "Eigenvector-centrality âĖ a node-centrality?" *Social Networks* 22:357–365.
URL: <http://www.sciencedirect.com/science/article/pii/S0378873300000319#>

- Sacerdote, Bruce. 2001. "Peer effects with random assignment: results for dartmouth roommates." *The Quarterly Journal of Economics* 116(2):681–704.
- Schelling, Thomas C. 1960. *The Strategy of Conflict*. Harvard University Press.
- Schneider, Christian M., Tamara Mihaljev, Shlomo Havlin and Hans J. Herrmann. 2011. "Suppressing epidemics with a limited amount of immunization units." *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* 84(6):2–7.
- Shulman, G L, M Corbetta, R L Buckner, J a Fiez, F M Miezin, M E Raichle and S E Petersen. 1997. "Common Blood Flow Changes across Visual Tasks: I. Increases in Subcortical Structures and Cerebellum but Not in Nonvisual Cortex." *Journal of cognitive neuroscience* 9(5):624–47.
URL: <http://www.ncbi.nlm.nih.gov/pubmed/23965121>
- Sidanius, Jim and Felicia Pratto. 2001. *Social dominance: An intergroup theory of social hierarchy and oppression*. New York: Cambridge University Press.
- Siegel, David a. 2013. "Social Networks and the Mass Media." *American Political Science Review* 107(04):786–805.
URL: http://www.journals.cambridge.org/abstract_S0003055413000452
- Simon, Herbert A. 1955. "A Behavioral Model of Rational Choice." *The Quarterly Journal of Economics* 69(1):99–118.
- Simon, Herbert A. 1956. "Rational choice and the structure of the environment." *Psychological review* 63(2):129.
- Sinclair, Betsey. 2012. *The Social Citizen: Peer Networks and Political Behavior*. Chicago: The University of Chicago Press.
- Sinclair, Betsy, Margaret Mcconnell, Donald P Green, Betsy Sinclair and Donald P Green. 2014. "All use subject to JSTOR Terms and Conditions Detecting Spillover Effects : Design Multilevel Experiments and Analysis of of Chicago." 56(4):1055–1069.
- Skyrms, Brian. 2009. "Groups and Networks: Their Role in the Evolution of Cooperation." *Games, Groups, and the Global Good* pp. 105–114.
URL: http://dx.doi.org/10.1007/978-3-540-85436-4_6

- Smith, Eric Alden and Rebecca L. Bliege Bird. 2000. "Turtle hunting and tombstone opening." *Evolution and Human Behavior* 21(4):245–261.
URL: <http://www.sciencedirect.com/science/article/pii/S1090513800000313>
- Smith, S S. 2005. "'Don't put my name on it': Social Capital Activation and Job-Finding Assistance among the Black Urban Poor." *ajs* 111(1):1–57.
- Sokhey, Anand E. and Paul A. Djupe. 2014. "Name generation in interpersonal political network data: Results from a series of experiments." *Social Networks* .
- Spitzer, M, U Fischbacher, B Herrnberger, G Gröthmann and E Fehr. 2007. "The neural signature of social norm compliance." *Neuron* 56(1):185–196.
- Thielmann, John and Allen Wilhite. 1998. "Campaign Tactics and the Decision to Attack." *Journal of Politics* 60(4):1050–1062.
- Tomz, Michael and Paul M Sniderman. 2005. "Brand Names and the Organization of Mass Belief Systems."
- Truman, David B. N.d. "The Governmental Process Groups and Government : Difficulties in a Group Interpretation of Politics." . Forthcoming.
- Uddin, Lucina Q, Marco Iacoboni, Claudia Lange and Julian Paul Keenan. 2007. "The self and social cognition: the role of cortical midline structures and mirror neurons." *Trends in cognitive sciences* 11(4):153–157.
URL: <http://www.ncbi.nlm.nih.gov/pubmed/17300981>
- Valente, T W, S Watkins, M N Jato, A Van Der Straten and L Tsitol. 1997. "Social Network Association with Contraceptive Use among Cameroonian Women in Voluntary Associations." *Social Science and Medicine* 45:677–687.
- Valente, Thomas W. 2005. Network Models and Methods for Studying the Diffusion of Innovations. In *Models and Methods in Social Network Analysis*, ed. Perer J Carrington, John Scott and Stanley Wasserman. Cambridge University Press chapter 6, pp. 98–116.
- Verba, S and N H Nie. 1987. *Participation in America: Political democracy and social equality*. University of Chicago Press.
- Verba, Sidney, Kay Schlozman and Henry Brady. 1995. *Civic Participation*. Cambridge, MA: Harvard University Press.

- Walsh, Katherine Cramer. 2004. *Talking about politics: Informal groups and social identity in American life*. University of Chicago Press.
- Wang, Dan J., Xiaolin Shi, Daniel a. McFarland and Jure Leskovec. 2012. "Measurement error in network data: A re-classification." *Social Networks* 34(4):396–409.
URL: <http://dx.doi.org/10.1016/j.socnet.2012.01.003>
- Wasserman, Stanley and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge University Press.
- Williamson, Oliver. 1975. *Markets and hierarchies: analysis and antitrust implications: a study in the economics of internal organization*. Vol. 46 Free Press New York.
- Wolfinger, R E and S J Rosenstone. 1980. *Who votes?* Yale Univ Pr.
- Yoon, Sooyeon, Sungmin Lee, Soon-Hyung Yook and Yup Kim. 2007. "Statistical properties of sampled networks by random walks." *Physical Review E* 75(4):046114.
URL: <http://link.aps.org/doi/10.1103/PhysRevE.75.046114>
- Zaller, J. 1990. "Political awareness, elite opinion leadership, and the mass survey response." *Social Cognition* 8(1):125–153.
- Zaller, John and Stanley Feldman. 1992. "A simple theory of the survey response: Answering questions versus revealing preferences." *American Journal of Political Science* 36(3):579–616.