Online Appendices

From Respondents to Networks: Bridging Between Individuals, Discussants, and the Network in the Study of Political Discussion

Political Behavior, 2018
https://doi.org/10.1007/s11109-017-9419-3

Matthew T. Pietryka, Florida State University, mpietryka@fsu.edu
Jack Lyons Reilly, New College of Florida, jreilly@ncf.edu
Daniel M. Maliniak, College of William & Mary, dxmali@wm.edu
Patrick R. Miller, University of Kansas, patrick.miller@ku.edu
Robert Huckfeldt, University of California, Davis, rhuckfeldt@ucdavis.edu
Ronald B. Rapoport, College of William & Mary, rbrapo@wm.edu

Appendix A: Summary Statistics

Table A1. Demographic Characteristics of the Student Body, the Full Sample, and Network Respondents

		N			Proportion			p-value of difference	
Variable	**Value**	**Student Body**	**Full Sample**	**Network Respondents**	**Student Body**	**Full Sample**	**Network Respondents**	**Student Body vs. Full Sample**	**Full Sample vs. Network Respondents**
Race/Ethnicity	African American	407	108	76	0.09	0.04	0.04	0	0.97
	Asian	443	158	96	0.09	0.07	0.06	0	0.21
	Caucasian, Non-Hispanic	3,201	1,904	1,392	0.68	0.79	0.81	0	0.19
	Hispanic	397	105	69	0.08	0.04	0.04	0	0.63
	Other	270	134	91	0.06	0.06	0.05	0.82	0.74
	Missing	1,118	310	109					
Gender	Female	3,175	1,626	1,083	0.54	0.6	0.61	0	0.4
	Male	2,661	1,081	682	0.46	0.4	0.39	0	0.4
	Missing		12	68					
Academic Year	Freshman	1,431	655	411	0.25	0.24	0.24	0.82	0.51
	Sophomore	1,285	682	457	0.22	0.25	0.26	0	0.62
	Junior	1,485	649	415	0.26	0.24	0.24	0.17	0.75
	Senior	1,589	692	463	0.27	0.26	0.27	0.13	0.64
	Missing	46	41	87					
Validated Voting in 2010 Williamsburg Election	Did not vote	4,984	2,220	1,422	0.87	0.82	0.78	0	0
	Voted	728	499	411	0.13	0.18	0.22	0	0
	Missing	124							
Total		5,836	2,719	1,833					

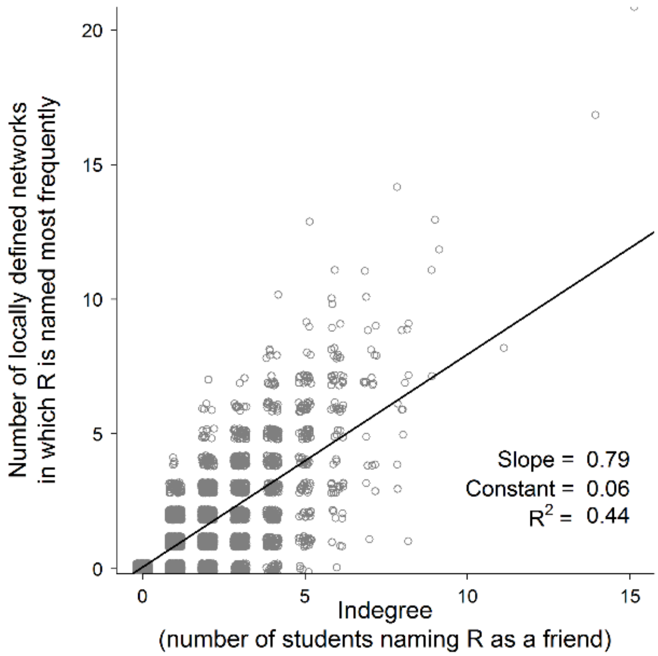
Note: The university provided us with a demographic breakdown of all students attending in Fall 2009. Those data are tabulated in the Student Body columns. The Full Sample and Network Respondents columns represent analogous data from all survey respondents, and the subset of respondents who provided at least one name in the name generator battery. The p-value columns display p-values from continuity-corrected two-sample proportion-equality tests.

Table A2. Summary Statistics for Data used in Analysis.

Variable	N	Mean	Std. Dev.	Min.	Max.
R's Validated Turnout in the 2010 Municipal Election	1575	0.24	0.42	0	1
(0 = DID NOT VOTE; 1 = VOTED)					
R's Noise-Ordinance Awareness	1519	1.66	0.82	0	3
(0 = NOT AWARE; 1 = NOT VERY; 2 = SOMEWHAT; 3 = VERY)					
R Attended Party Cited by Police	1509	0.22	0.42	0	1
(0 = NO; 1 = YES)					
R's Academic Year	1508	1.55	1.11	0	3
(0 = FRESHMAN; 1 = SOPHOMORE; 2 = JUNIOR; 3 = SENIOR)					
R's Interest in National Politics	1495	2.20	0.75	0	3
(0 = NOT AT ALL; 1 = NOT VERY; 2 = SOMEWHAT; 3 = VERY)					
R's Validated Turnout in the 2009 Gubernatorial election	1575	0.15	0.36	0	1
(0 = DID NOT VOTE; 1 = VOTED)					
R's Family Economic Status	1484	2.52	0.82	0	4
(0 = WORKING CLASS; 1 = LOWER MIDDLE; 2 = MIDDLE; 3 = UPPER MIDDLE; 4 = UPPER)					
R is White?	1483	0.82	0.39	0	1
(0 = NO; 1 = YES)					
R is Black?	1483	0.04	0.19	0	1
(0 = NO; 1 = YES)					
R is Female?	1522	0.62	0.49	0	1
(0 = NO; 1 = YES)					
Zone 1 Mean Turnout in 2010 Election	1575	0.21	0.22	0	1
(MEAN AMONG R'S NAMED FRIENDS)					
Zone 2 Mean Turnout in 2010 Election	1575	0.21	0.21	0	1
(MEAN AMONG FRIENDS OF R'S NAMED FRIENDS)					
Nodes	1575	12.46	3.94	3	27
(NUMBER OF PEOPLE IN R'S TWO-STEP NETWORK)					
Edges	1575	19.48	8.16	3	53
(NUMBER OF RELATIONSHIPS IN R'S TWO-STEP NETWORK)					
Indegree	1575	2.14	1.68	0	15
(NUMBER OF STUDENTS NAMING R AS A FRIEND)					
Local Centrality	1575	1.14	1.99	0	21
(NUMBER OF TWO-STEP NEIGHBORHOODS WHERE R IS MOST CENTRAL)					

Note: Summary statistics are restricted to survey respondents with at least one Zone 1 and one Zone 2 associate.

Figure A1. Centrality within locally defined networks by global centrality (indegree).



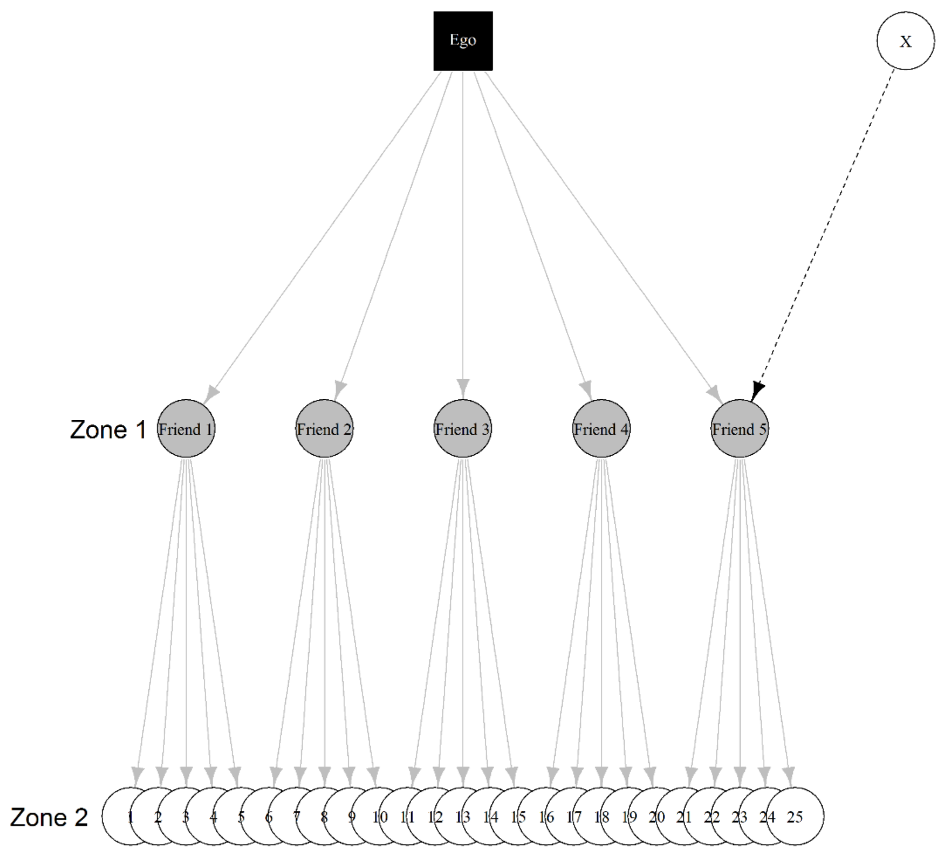
Note: Estimates include respondents and all students named as a friend by at least one respondent (N = 4,293). Dots jittered to show frequency. The x-axis shows global centrality as measured by indegree—the number of respondents nominating an individual as a friend. Ninety-four percent of the students have an indegree between 0 and 4 and approximately 2.5 percent have an indegree exceeding 5. Only three students are identified by ten or more respondents. The y-axis shows local centrality as measured by the number of two-step neighborhoods in which an individual is the most frequently named friend. Roughly 30 percent of students are identified as being most central in one two-step neighborhood—most typically their own. More than 20% are most central in two neighborhoods and almost 10% are central in three. The relationship between these two measures is strong, but by no means perfect; the measures appear to tap different dimensions of centrality.

Appendix B: Creating the Networks

To create the two-step neighborhoods, we begin by constructing a directed network which includes as nodes all students at the university. An edge from student i to j is present in the network if i named j as a friend via the name generator. Using this complete network, we then identify the two-step neighborhoods of all respondents who identified at least one friend using the name generator. The number of steps in the neighborhood refers to the number of relationships needed to bridge the gap between a pair of individuals. We use only outgoing ties in this process because influence in friendship networks is limited in cases where the alter names the ego as a friend but the ego does not reciprocate (Christakis and Fowler 2009). Figure B1 provides a hypothetical example. In the figure, the ego is one step away from each of the five friends she identified using the name generator (Zone 1) and two steps away from each of these friends' friends (Zone 2). Thus, the ego's two-step neighborhood includes all five friends in Zone 1 and all 25 individuals in Zone 2. Using only outgoing ties, the ego's neighborhood does not include the student labeled 'X', because he has not been named as a friend by the ego or any of her five friends.

Figure B1. Example of a two-step neighborhood.

The network includes the ego, all five friends named by the ego (Zone 1), and each of these friends' five friends (Zone 2). It does not include X, the person who identified Friend 5 as a friend, because X was not identified as a friend by the ego or any of his five friends.

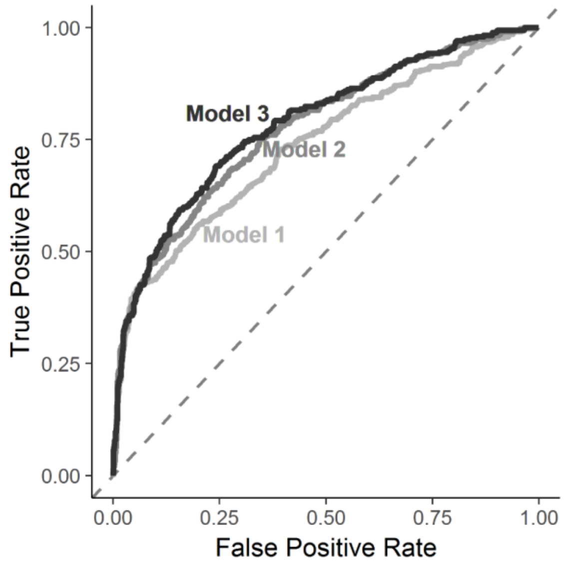


Appendix C: Predictive Accuracy

Just how much does moving beyond close friends improve our models? One way to answer this question is to examine the accuracy of our predictions using the true and false positive rates. In the context of Table 4 from the main text, which models students' participation in the 2010 election, the true positive rate tells us the proportion of voters whom the model predicts would vote. The false positive rate tells us the proportion of non-voters the model predicts to vote. By these metrics, better models have greater true positive rates and lower false positive rates. When these two metrics are equal, the model performs only as well as random guesses.

Since the regression generates probabilistic rather than binary predictions, one cannot calculate these quantities without first choosing a probability threshold that can separate predicted voters from predicted non-voters. For instance, we may classify as predicted voters anyone whose predicted probability exceeds 0.5. This threshold is arbitrary and other thresholds that improve predictive accuracy often exist. Figure C1 plots these quantities for the models in Table 4 across all possible thresholds, placing the true and false positive rates on the y and x axis. The plot facilitates comparisons of different models because a model's performance improves as its curve tends more toward the upper-left corner of the plot, where correct predictions are maximized and incorrect predictions are minimized.

Figure C1. A model including Zone 2 and network structure improves predictions of turnout.

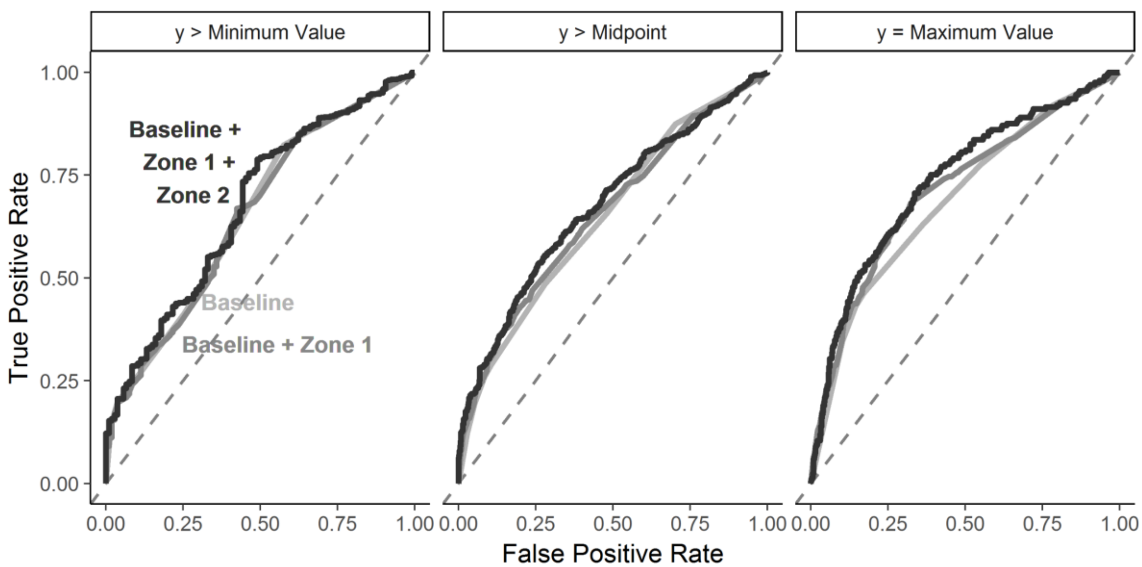


Note: This figure is based on predictions from the models in Table 4 of the main text. Model 1 includes only attributes of the respondent. Model 2 includes these individual covariates as well as the average turnout in Zone 1. Model 3 includes the Model 2 covariates as well as average turnout in Zone 2 and measures of the respondent's two-step network structure.

The plot suggests that all three models perform better than random guessing because each curve is well above the 45-degree line. More importantly, Model 3 contains more area under its curve than Models 2 or 1. This result suggests that Model 3, which includes the Zone 2 and network structure measures, performs better than Model 1, which excludes all network measures, and Model 2, which includes only the Zone 1 network. When Model 3 maximizes the difference between true and false positives, it correctly classifies 69% of voters (true positives) and incorrectly classifies 24% of non-voters (false positives). To obtain a comparable rate of correct voter classification using model 1 or 2, we would also have to accept an increased false positive rate, incorrectly classifying 29% of non-voters with Model 2 and 38% with Model 1. Alternatively, to obtain with model 1 or 2 a comparable false positive rate, we would have to accept a reduced rate of correctly classified voters. Though at its best Model 3 classifies correctly 69% of voters, obtaining a similarly-low false positive rate would yield 64% correctly-classified voters with Model 2 and 58% with Model 1. Thus Model 3 improves over Model 2 by about five percentage points in either the true of false positive rate and improves over Model 1 by just over 10 percentage points on either metric. By comparison, when Model 2 maximizes the difference between the true and false positive rates, it improves over Model 1 by about seven percentage points on either metric. Thus, a model that considers close friends improves predictions considerably over an atomistic model, and moving beyond this small group to include the broader network improves predictions by almost as much.

Figure C2 provides analogous plots for models from Table 3. The figure provides three sets of curves because the outcome variable, awareness of the noise ordinance, is ordered rather than binary. The first panel plots predictions classifying whether a student will exceed the minimum awareness level. The second panel plots predictions classifying whether a student will exceed the midpoint, rating themselves as somewhat or very familiar with the ordinance. The third panel plots predictions classifying whether a student will rate themselves as very familiar, the maximum possible value.

Figure C2. A model including Zone 2 and network structure improves predictions of noise-ordinance awareness.



Note: This figure is based on predictions from the models in Table 3 of the main text. The Baseline model, not displayed in Table 3, includes only attributes of the respondent. The Baseline + Zone 1 model, Model 1 from the table, includes these individual covariates as well as the proportion of Zone 1 reporting having attended a party cited by police for violating the noise ordinance. The Baseline + Zone 1 + Zone 2 model, Model 3 from the table, includes the aforementioned covariates as well as the proportion of Zone 2 reporting having attended a party cited by police and measures of the respondent's two-step network structure.

Even the first model in Table 3 includes the Zone 1 measures. Thus, the plots also include a baseline model run on the same subset of data as the Table 3 models, but omitting the network covariates. The plots compare this baseline model to Table 3 Model 1, which includes the baseline covariates and Zone 1, and Table 3 Model 3, which includes these covariates as well as the Zone 2 and the structure measures. When predicting which students will exceed the minimum (the left panel), adding Zone 1 provides little improvement over the baseline model, but adding the Zone 2 and the structure measures provides modest improvement. When predicting which students exceed the midpoint (middle panel) or meet the maximum (right panel), the Zone 1 model improves predictions somewhat over the baseline model and the Zone 2 and structure model improves further. Thus, moving beyond Zone 1 improves predictions modestly, but consistently across the range of noise-ordinance awareness.

Appendix D: Sensitivity Analysis

The cross-sectional survey provides only limited controls and therefore we subject our estimates of social influence to sensitivity analysis, as developed by VanderWeele (2011). In the analysis shown in Table D1, we assume that some binary variable, U, is omitted from the regression. For our purposes, U may represent some shared environmental influence (e.g., a mobilization effort contacting the main respondent and much of her network), source of homophily (e.g., people who enjoy parties forming relationships with other partygoers), or reverse causality bias (i.e., the main respondent's attitudes/behavior affecting that of her friends). We then allow gamma, the effect of U on Y, to vary and observe the social influence estimates we would have obtained had we controlled for U. Table B allows gamma to vary from 1—corresponding to no effect of U on Y—to 5—an effect of U corresponding to a five-fold increase in the odds of Y = 1.[1]

For these estimates, we also must make assumptions about the prevalence of U = 1 across the range of the explanatory variable X. We assume a strong positive relationship between U and X. For example, in the analysis of the Zone 1 Mean 2010 Turnout estimate from Table 4, we assume U = 1 for 70% of respondents with all Zone 1 friends voting; we assume U = 1 for only 30% of respondents with no Zone 2 friends voting. We use these 70%/30% prevalence differences for all corrected estimates.

The table shows that the omitted variable or variables would need to produce a very strong effect on Y to fully explain away the apparent social influence. It would take an odds ratio approaching 3 to explain away the Zone 2 estimate from Table 3. All other estimates are robust

[1] When gamma = 1, we recover the original estimates from Tables 4 and 5. Due to the positive coefficients associated with the uncorrected social influence estimates, the corrected estimates of social influence decrease as gamma increases. With values of gamma less than one, the corrected estimates of social influence are larger than the uncorrected estimates and thus we focus only on instances where gamma is greater than or equal to 1.

to much larger levels of bias, sustaining in the presence of odds ratios well in excess of 5. Looking at the Table 4 Zone 2 estimate for example, if U produced a fivefold increase in the odds of Y = 1 (Gamma = 5), the corrected estimate suggests respondents will have 64% higher odds of voting if all of their Zone 2 friends vote than if none vote. Thus the estimates are robust to large levels of endogeneity or omitted variable bias.

Table D1. Sensitivity Analysis for Social Influence Estimates From Tables 3 and 4

			Gamma (Odds ratio for omitted variable)				
Model	Outcome Variable	Explanatory Variable	1	2	3	4	5
Table 3, Model 3	Noise-Ordinance Awareness	Zone 1 Mean Attended	1.98	1.52	1.32	1.22	1.15
Table 3, Model 3	Noise-Ordinance Awareness	Zone 2 Mean Attended	1.52	1.16	1.01	0.93	0.88
Table 4, Model 3	Turnout	Zone 1 Mean Turnout	5.33	4.08	3.56	3.27	3.09
Table 4, Model 3	Turnout	Zone 2 Mean Turnout	2.83	2.17	1.89	1.74	1.64

Note: The Table applies sensitivity analysis to the estimates from Tables 3 and 4. Cell entries represent the estimates of an explanatory variable's odds ratio we would obtain after controlling for omitted variable bias of strength gamma.

Appendix E: Robustness Checks for the Turnout Models

Instrumental Variables Analysis

In Table 4, we regressed main respondents' turnout on their friends' turnout. To the extent that we are interested only in friends influence on the main respondents, these coefficients are biased due to the endogenous relationship between the behavior of respondents and their friends. This type of endogeneity bias is common in the study of social influence and is often addressed through an instrumental variables approach (e.g., Huckfeldt and Mendez 2008; Huckfeldt and Sprague 1991, 1995; Ikeda and Richey 2009; Kenny 1992; Mutz 2002; Rogowski and Sinclair 2012). With this approach, we hope to purge the portion of variation in friends' turnout that is influenced by main respondents' behavior. We can do so if we can find an instrument or set of instruments that are uncorrelated with main respondents' turnout, but correlated with friends' turnout (Greene 2012, 223).

The endogenous regressors we are concerned with are the Zone 1 and Zone 2 mean turnout measures. We instrument for these measures using Zone 1 and Zone 2 mean validated turnout in the previous year's gubernatorial election and measures of the proportion of Zone 1 and Zone 2 alters registered to vote in Williamsburg by January 1, 2010.[2] Like 2010 turnout, the 2009 turnout and registration measures both come from state records rather than self reports. Prior participation provides a good instrument because it should influence subsequent turnout (Gerber, Green, and Shachar 2003), particularly for the young (Coppock and Green 2015). Yet it

[2] We use this date because it is well before the 2010 municipal election campaign and registration is thus temporally prior to any conversations emerging about the campaign. Instead of using as the cutoff January 1, 2010, we have also replicated the analysis using December 1, 2009 and February 1, 2010. Both alternative specifications lead to the same conclusions of substantive and statistical significance.

cannot be influenced by main respondents' 2010 participation. Registration provides a useful instrument because it imposes a cost on voting in the city, specifically. Even politically-interested students may fail to vote if they are unregistered. Nonetheless, registration status is unlikely to be a topic of conversation for most college students. Thus, registration influences friends' chances of voting in 2010, but should be largely exogenous to main respondents' chances.

Table E1 presents the results from the instrumental variable analysis. In the table, the first two columns of coefficients present first-stage estimates of the endogenous regressors: Zone 1 and Zone 2 Mean Turnout in 2010. The last two coefficient columns present the second stage equations estimated via probit and least squares. We use robust standard errors with the least squares estimate to account for the heteroskedasticity produced by the linear probability model. The linear estimates are useful because they provide several diagnostic checks on the key assumptions of the instrumental variables approach.

Consistent with the Second-Order-Friends Hypothesis, the coefficients for Zone 2 turnout and registration are significant in the first-stage equations. The Zone 2 coefficient remains large in the second-stage results, though the instrument introduces too much uncertainty to reject the null ($p = 0.054$). This finding suggests that even after purging the main respondent's influence from the measure, Zone 2's behavior still has a strong positive relationship with the main respondent's 2010 turnout. Several diagnostic checks from the linear estimates bolster our evidence. First, the F test for weak instruments (see Sovey and Greene 2011) is large and statistically significant for both first-stage models, suggesting an adequately strong relationship between our instruments and the endogenous regressors. Second, the insignificant Sargan test fails to reject the null that the included instruments are endogenous to the model residuals.

Together, these tests suggest the instruments are sufficiently related to friends' turnout and sufficiently exogenous to main respondents' turnout.

Applying the sensitivity analysis introduced in Appendix D to these estimates, the linear estimate of Zone 2 mean turnout is robust to unaccounted bias equivalent to a 55 percentage point increase in individuals' propensity to vote.[3] Thus, to explain away these results, the bias produced by homophily or shared environments would need to be stronger than the estimated impact of any other explanatory variable in the model.

[3] For linear estimates, the sensitivity analysis requires an assumption about the relationship between the explanatory variable X and the biasing variable U. This relationship δ is expressed as a difference in prevalence in U = 1 associated with a one-unit increase in X (VanderWeele 2011, 245). We assume a strong positive relationship, with $\delta = .4$. Substantively, we are assuming a 40 percentage point increase in the prevalence of the biasing variable U = 1 for respondents with all Zone 2 friends voting relative to respondents with no Zone 2 friends voting.

Table E1. Instrumental Variables Analysis of the Relationship between Turnout of Main Respondents and their Zone 1 and 2 Friends.

	First Stage		Second Stage	
	Zone 1	Zone 2	Probit	Linear
Intercept	0.029	0.059	-1.888 *	-0.073
	(0.031)	(0.031)	(0.257)	(0.065)
R's Noise-Ordinance Awareness	-0.002	0.010	0.084	0.023
	(0.007)	(0.007)	(0.053)	(0.014)
R Attended Party Cited by Police	-0.013	-0.017	0.109	0.026
	(0.013)	(0.012)	(0.099)	(0.027)
R's Academic Year	-0.045 *	-0.033 *	-0.112 *	-0.029 *
	(0.005)	(0.005)	(0.039)	(0.011)
R's Interest in National Politics	-0.002	0.015 *	-0.033	-0.009
	(0.007)	(0.007)	(0.057)	(0.015)
R Validated Turnout in the 2009 Gubernatorial Election	-0.009	0.013	0.611 *	0.183 *
	(0.011)	(0.011)	(0.083)	(0.026)
R's Family Economic Status	0.012 *	0.009	0.001	0.003
	(0.006)	(0.006)	(0.050)	(0.013)
R is White? (Asian/Latino/Other is reference cat.)	0.028	-0.006	0.173	0.041
	(0.014)	(0.014)	(0.119)	(0.030)
R is Black? (Asian/Latino/Other is reference cat.)	0.002	-0.021	-0.644 *	-0.104 *
	(0.028)	(0.028)	(0.314)	(0.044)
R is Female	-0.019	-0.016	-0.033	-0.011
	(0.010)	(0.010)	(0.082)	(0.023)
Nodes	-0.000	-0.006 *	0.038 *	0.011 *
	(0.002)	(0.002)	(0.018)	(0.005)
Edges	0.003 *	0.003 *	-0.016	-0.005
	(0.001)	(0.001)	(0.009)	(0.003)
Indegree	0.004	0.005	0.101 *	0.028 *
	(0.005)	(0.005)	(0.038)	(0.011)
Local Centrality	0.002	-0.003	-0.033	-0.009
	(0.004)	(0.004)	(0.029)	(0.009)
Zone 1 Mean Turnout in 2009 Gubernatorial Election	0.286 *	0.015		
	(0.032)	(0.032)		
Zone 2 Mean Turnout in 2009 Gubernatorial Election	0.103 *	0.279 *		
	(0.033)	(0.033)		
Zone 1 Mean Registered Voters	0.267 *	-0.013		
	(0.023)	(0.023)		
Zone 2 Mean Registered Voters	-0.044	0.252 *		
	(0.024)	(0.024)		
Zone 1 Mean 2010 Turnout (Instrumented)			1.634 *	0.482 *
			(0.361)	(0.108)
Zone 2 Mean 2010 Turnout (Instrumented)			0.799	0.224
			(0.413)	(0.116)
N	1420	1420	1420	1420
Weak instruments F statistic	79.117	50.008		
Weak instruments p-value	< .001	< .001		
Sargan statistic			3.78	
Sargan p-value			0.15	

The table presents estimates for two-stage models with endogenous regressors. The first two columns of coefficients present first-stage estimates of the endogenous regressors: Zone 1 and Zone 2 Mean Turnout in 2010. The last two coefficient columns present the second stage equations estimated via probit and least squares (with robust standard errors). Standard errors in parentheses. * $p < 0.05$ (two-tail)

Spatial Regression Analysis

A related issue with the Table 4 estimates is that observations are not independent. Indeed, our theory predict a reciprocal relationship between whether individuals and their friends turnout to vote and we therefore *expect* to violate the independence assumption. We deal with this problem with spatial regression techniques, which have been developed to account for various forms of interdependence. One such technique—the spatial lag model—is appropriate when values of an outcome are expected to be influenced by values of the outcome for “neighboring” observations. An alternative technique—the spatial errors model—is appropriate when some variable (or set of variables) influencing Y is omitted from the model and its values covary with network position, creating correlation in the residuals of neighboring observations.

For our purposes, neighboring observations are those connected by friendship. Thus, spatial lag models account for otherwise unmodeled relationships between the turnout behavior of friends. Given that we have tried to model this relationship directly with the Zone 1 and Zone 2 turnout means, this model can test whether our specification has fully accounted for the relationship between the turnout of main respondents and that of their friends. The spatial errors models can account for unobserved similarities between friends, reducing bias created by latent homophily or unobserved environmental confounds.

Estimating the models requires specifying a set of weights that each observation places on neighboring observations. We assign respondents’ Zone 1 friends a weight of two, Zone 1 friends a weight of one, and all other observations a weight of zero.[4] Non-linear models of these

[4] We have replicated the analysis using several other weighting strategies. These include (1) placing non-zero weights only on Zone 1 friends, (2) weighting Zone 1 and Zone 2 friends equally, and (3) weighting all other individuals in inverse proportion to the shortest outgoing path from the main respondent to the individual. While the spatial parameters change depending

types often lack closed-form solutions (Ward and Gleditsch 2002) and, in light of this challenge, linear probability models provide a reasonable approximation (See Beron and Vijverger 2004 for monte carlo comparisons of these two approaches).[5] The results are presented in Table E2. The table models respondents' turnout as a function of the same covariates used in Table 4, Model 3.

The spatial lag model includes the rho term, which measures the extent to which main respondents' turnout is related to that of their friends, conditional on the other covariates in the model. Our estimate of rho is almost zero and statistically insignificant, suggesting that our Zone 1 and Zone 2 mean turnout variables have fully captured this form of interdependence.[6] In the spatial errors model, the statistically significant lambda term suggests a correlation between the residuals of respondents and their friends. After correcting for this correlation, the estimates of Zone 1 and 2 mean turnout remain positive and statistically significant. Applying the sensitivity analysis from Appendix D to these estimates, the estimate of Zone 2 mean turnout is robust to unaccounted bias equivalent to a 44 percentage point increase in individuals' propensity to vote.[7] Thus, to explain away these results, the bias produced by homophily or shared environments would need to be as strong as the estimated impact of the individual's own turnout in the previous election ($\hat{\beta}_{2009\,Turnout} = .45$).

on the weighting scheme, all strategies we have tried lead to the same substantive conclusions about our hypotheses.

[5] Estimating linear probability models with individual turnout as the outcome variable is common in the voting behavior literature (e.g., Ansolabehere and Hersh 2012; Chong et al. 2015; Fowler 2013).

[6] Indeed, if we exclude the Zone 1 and Zone 2 covariates, rho is positive and significant (p-value < .001), indicating a positive relationship between the turnout behavior of respondents connected as friends.

[7] Again assuming a 40 percentage point increase in the prevalence of the biasing variable for people with all Zone 2 friends voting relative to people with no Zone 2 friends voting.

Table E2. Spatial Regression Analysis of the Relationship between Turnout of Main Respondents and their Zone 1 and 2 Friends.

Outcome Variable is Validated Turnout in the 2010 Municipal Election

	Lag Model	**Errors Model**
Intercept	-0.005	-0.054
	(0.062)	(0.058)
R's Noise-Ordinance Awareness (0 = Not aware; 1 = Not very; 2 = Somewhat; 3 = Very)	0.012	0.011
	(0.013)	(0.013)
R Attended Party Cited by Police (0 = No; 1 = Yes)	0.019	0.022
	(0.025)	(0.024)
R's Academic Year (0 = Freshman; 1 = Sophomore; 2 = Junior; 3 = Senior)	-0.036*	-0.031*
	(0.009)	(0.008)
R's Interest in National Politics (0 = Not at all; 1 = Not very; 2 = Somewhat; 3 = Very)	-0.003	-0.004
	(0.014)	(0.014)
R Validated Turnout in the 2009 VA Gubernatorial Election (0 = No; 1 = Yes)	0.459*	0.448*
	(0.028)	(0.028)
R's Family Economic Status	0.005	0.006
	(0.012)	(0.012)
R is White? (Asian/Latino/Other is reference cat.)	0.051	0.056*
	(0.029)	(0.028)
R is Black? (Asian/Latino/Other is reference cat.)	-0.105	-0.092
	(0.056)	(0.055)
R is Female (0 = No; 1 = Yes)	-0.029	-0.024
	(0.021)	(0.019)
Zone 1 Mean Turnout in Williamsburg 2010 election	0.258*	0.397*
	(0.048)	(0.045)
Zone 2 Mean Turnout in Williamsburg 2010 election	0.142*	0.180*
	(0.049)	(0.047)
Nodes (Number of people in R's two-step network)	0.009	0.011*
	(0.005)	(0.004)
Edges (Number of relationships in R's two-step network)	-0.003	-0.004
	(0.002)	(0.002)
Indegree (Number of students naming R as a friend)	0.022*	0.021*
	(0.010)	(0.010)
Local Centrality (Number of two-step neighborhoods where R is most central)	-0.006	-0.005
	(0.008)	(0.008)
N	1436	1436
AIC (Linear model)	1281.471	1281.471
AIC (Spatial model)	1283.300	1276.357
Log Likelihood	-623.650	-620.178
Rho statistic	0.003	
Rho p-value	0.679	
Lambda statistic		-0.027
Lambda p-value		0.008

Maximum Likelihood Estimates from linear probability spatial regression models. Standard errors in parentheses. * $p < 0.05$ (two-tail)

References

Beron, Kurt J., and Wim P. M. Vijverberg. 2004. "Probit in a Spatial Context: A Monte Carlo Analysis." In *Advances in Spatial Econometrics*, Advances in Spatial Science, eds. Dr Luc Anselin, Dr Raymond J. G. M. Florax, and Dr Sergio J. Rey. Springer Berlin Heidelberg, 169–95.

Chong, Alberto, L. Ana, Dean Karlan, and Leonard Wantchekon. 2015. "Does Corruption Information Inspire the Fight or Quash the Hope? A Field Experiment in Mexico on Voter Turnout, Choice, and Party Identification." *The Journal of Politics* 77(1): 55–71.

Christakis, Nicholas A., and James H. Fowler. 2009. *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*. Back Bay Books, Little, Brown.

Coppock, Alexander, and Donald P. Green. 2015. "Is Voting Habit Forming? New Evidence from Experiments and Regression Discontinuities." *American Journal of Political Science*: forthcoming.

Fowler, Anthony. 2013. "Electoral and Policy Consequences of Voter Turnout: Evidence from Compulsory Voting in Australia." *Quarterly Journal of Political Science* 8(2): 159–82.

Gerber, Alan S., Donald P. Green, and Ron Shachar. 2003. "Voting May Be Habit-Forming: Evidence from a Randomized Field Experiment." *American Journal of Political Science* 47(3): 540–50.

Greene, William H. 2012. *Econometric Analysis*. Seventh. Upper Saddle River: Pretice Hall.

Huckfeldt, Robert, and John Sprague. 1991. "Discussant Effects on Vote Choice: Intimacy, Structure, and Interdependence." *The Journal of Politics* 53(1): 122–58.

Ikeda, Ken'ichi, and Sean Richey. 2009. "The Impact of Diversity in Informal Social Networks on Tolerance in Japan." British Journal of Political Science 39(03): 655–68.

Kenny, Christopher B. 1992. "Political Participation and Effects from the Social Environment." *American Journal of Political Science* 36(1): 259.

Mutz, Diana C. 2002. "Cross-Cutting Social Networks: Testing Democratic Theory in Practice." *American Political Science Review* 96(01): 111–26.

Rogowski, Jon C., and Betsy Sinclair. 2012. "Estimating the Causal Effects of Social Interaction with Endogenous Networks." Political Analysis 20(3): 316–28.

Sovey, Allison J., and Donald P. Green. 2011. "Instrumental Variables Estimation in Political Science: A Readers' Guide." *American Journal of Political Science* 55(1): 188–200.

Ward, Michael D., and Kristian Skrede Gleditsch. 2007. "An Introduction to Spatial Regression Models in the Social Sciences." *Manuscript at http://www. faculty. washington. edu/mdw.*