Online Appendix for Viral Voting: Social Networks and Political Participation

Nicholas Eubank,[*]
Guy Grossman,[†]
Melina Platas,[‡]
Jonathan Rodden[§]

June 23, 2020

Online Appendix

A Network Survey

We collect data from a set of Ugandan villages that took part in a multi-year program called Governance, Accountability, Participation, and Performance (GAPP), which was implemented by RTI International and funded by the United States Agency for International Development (USAID) in Arua district, Uganda. 16 villages were selected from a set of over 131 villages that were part of the U-Bridge program, the maximum number that could be enumerated considering budget constraints. Half of the villages had a relatively high level of adoption of the U-Bridge program given village characteristics, and half of which had low levels of adoption. the process of selecting the highest and lowest performers was as follows. We regressed village level adoption of the U-Bridge technology on village-level predictors, and generate a set of predicted values for the dependent variable. We then calculated the difference between the predicted value and the actual value of U-Bridge adoption. Using these residuals, we selected the 8 highest performing and the 8 lowest performing villages with respect to U-Bridge adoption.

We conducted a census in each village in order to collect complete network data, interviewing every available adult who was a resident in the village. This involved a village listing prior to enumeration, the purpose of which was to create a written record of all of the names and household locations for all adults in the villages. To do this the enumerator met with the village leader (in Uganda, called an LC1) and two other village leaders (usually a member of the Village Health Team and an additional village elder or

[*]Center for the Study of Democratic Institutions, Vanderbilt University, nick@nickeubank.com
[†]Department of Political Science, University of Pennsylvania and EGAP, ggros@sas.upenn.edu
[‡]Department of Political Science, New York University Abu Dhabi, mplatas@nyu.edu
[§]Department of Political Science, Stanford University, jrodden@stanford.edu

community leader). Together they drew a map of the village identifying every location of a household and major geographical features (e.g., rivers, churches, etc.). Then the group created a list—using their shared knowledge of those households—to identify and name every adult in the village. The enumerator entered this information into a tablet along with other key identifying information such as quadrant (an arbitrary division of the village created by the enumerator—designed to divide the village into four equal portions), age and gender of the potential respondent. These names were then used in the network section of the survey, where respondents were asked about four types of social ties. The exact question wording for the social ties is as follows:

"In each of the following questions, we will ask you to think about people in your community and their relationships to you."

- Family: "Think about up to five family members in this village not living in your household with whom you most frequently spend time. For instance, you might visit one another, eat meals together, or attend events together."
- Friends: "Think about up to five of your best friends in this village. By friends I mean someone who will help you when you have a problem or who spends much of his or her free time with you. If there are less than five, that is okay too."
- Lender: "Think about up to five people in this village that you would ask to borrow a significant amount of money if you had a personal emergency."
- Problem solver: "Imagine there is a problem with public services in this village. For example, you might imagine that a teacher has not come to school for several days or that a borehole in your village needs to be repaired. Think about up to five people in this village whom you would be most likely to approach to help solve these kinds of problems."

B Turnout Estimation

As noted above, our turnout measure assumes that votes at a given precinct (polling station) are evenly distributed among voters registered at that precinct from different villages.

To illustrate, assume that at Precinct 1, 200 votes were cast. If 75% of the voters registered at Precinct 1 come from Village A and 25% come from Village B, we assume that Village A contributed 150 votes and Village B contributed 50 votes. Turnout for each village is then calculated as the sum of votes we infer to have been case by its residents at all Precincts. By assuming that at a given precinct all villages have the same turnout, this estimation should bias our analysis in favor of not finding differences in turnout across villages.

One constraint of this measures is that the accuracy of these estimates will be related to the correspondence of villages to precincts. If each village sends all residents to its own polling place (i.e. that polling place is only attended by residents of one village), inferences about village voting will be perfect. If, by contrast, all villages send their voters to a single polling place, effectively no information can be learned about how individual villages voted.

This correspondence can be summarized using a concentration statistic, where higher values mean the mapping from precincts to villages is more precise. For a village $v \in V$ and a polling place (precinct) $p \in P$, let $voters_v$ be the set of voters who live in v, let $voters_p$ be the set of voters who vote at polling place p, and let $voters_{v,p}$ be the set of voters who reside in v and voted at p. Then:

$$concentration_{v,p} \equiv \frac{\#voters_{v,p}}{\#voters_p} \tag{1}$$

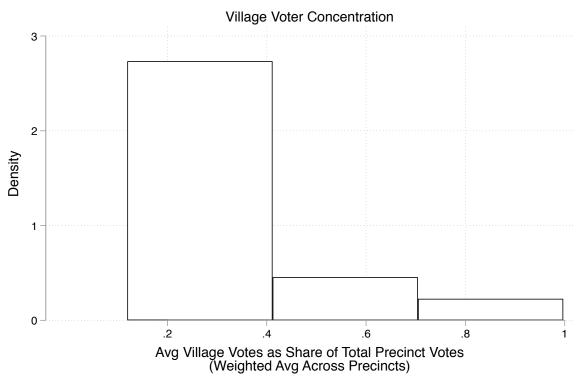
In other words, $concentration_{v,p}$ is the share of voters at a precinct from village v.

Since villages send people to multiple polling places, however, we must then calculate a weighted average of $concentration_{v,p}$ across polling places where the weight for each polling place is the share of voters from each village going to that polling place. Formally:

$$concentration_v \equiv \sum_{p \in P} concentration_{v,p} * \frac{\#voters_{v,p}}{\#voters_v} \tag{2}$$

The distribution of this statistic are presented in Figure 1:

Figure 1: Distribution of $concentration_{v,p}$ Measure Across Villages



Distribution of village-level estimated turnout as share of adult populations are presented below in Figure 2. In addition, we also plot $concentration_v$ against TPP in Figure 3.

Figure 2: Estimated Village Turnout as a Share of Adult Population

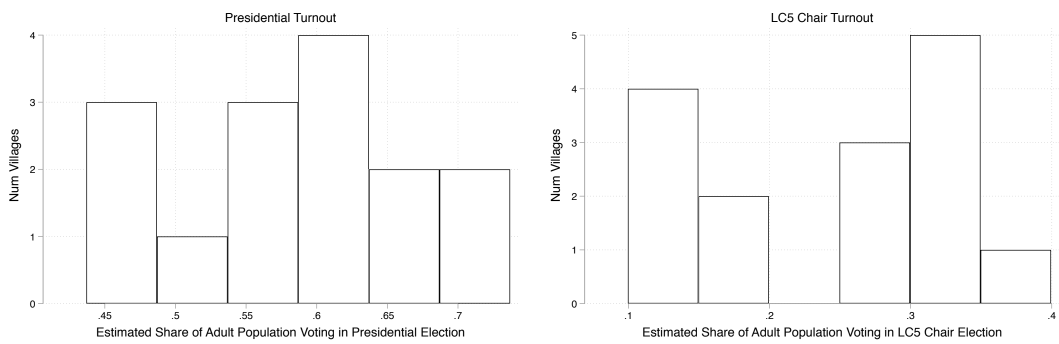
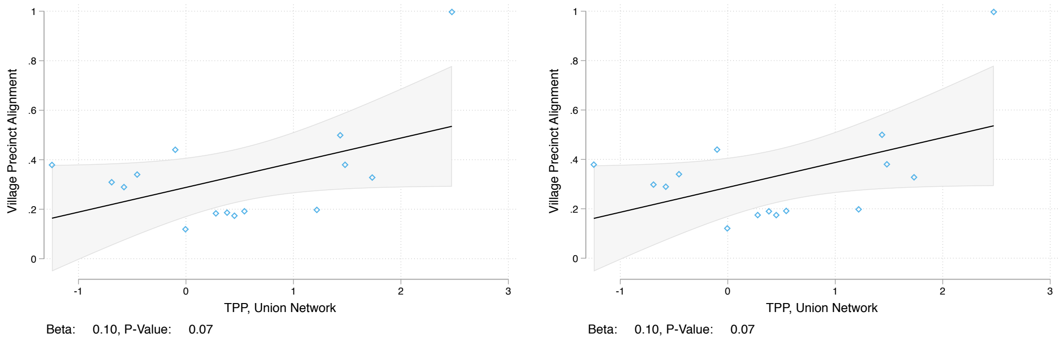


Figure 3: Precinct / Village Alignment (*$concentration_v$*) and TPP



Notes: The above plot presents the partial correlation between Theoretically-Predicted Participation (TPP) and precinct / village alignment ($concentration_v$). Grey bands indicate 95% confidence intervals. As detailed in Section ??, TPP is operationalized as the first principle component of normalized TPP scores across all parameter choices (as TPP is highly correlated across parameters). Presidential and LC5 Chair are concentration are correlated at 0.9998, which is why the plots are nearly identical.

C TPP Simulation Notes

C.1 Participation Updating

Updating of the $participation_{v,t+1}$ is accomplished by iterating through all vertices in the network in random order and having each vertex update its value of lpr and its participation $participation_{v,t+1}$ sequentially rather than simulatenously. This is the one departure from Siegel (2009). When all vertices update lpr simultaneously, it is possible to converge to a "flashing" state in which at time t a portion of the network is planning to vote while another portion is not planning to vote, at time $t+1$, these two groups flip inclinations, and at time $t+2$ they return to their initial state. This is caused by knife-edge simultaneity of updating, which seems unrealistic, since real updating is almost certainly sequential. Thus simulation uses sequential updating.

C.2 Convergence

The simulation is run until no more than 1% percent of the vertices in the network change participation status for at least 20 consecutive periods. Results below are averaged across 2,500 runs for each set of parameter values.

D Parameter Choices

These parameter values are chosen because they effectively cover the range of values that give rise to interesting dynamics. Significantly higher values of β_{mean} tend to result in convergence to full participation, while substantially lower values lead to non-dynamic simulations (those with values of $\beta > 1$ participate, but they are rare and others tend to have very low proclivities to participate, as a result of which almost no vertices flip from non-participation to participation). Similarly, larger values of β_{sd} increase the share of individuals whose behavior is unaffected by the behavior of other so much that the simulations tend not to be dynamic. In these non-dynamic settings, all networks are essentially comparable, as participation ends up being roughly equal to the share of nodes with $\beta_{mean} > 1$, which is the same for all networks in expectation.

Note that we exclude one parameter pair from those sets ($\beta_{mean} = 0.5, \beta_{sd} = 0.25$), as it generates almost no unconditional participators, and thus no dynamics.

E Social Context Simulation Validity

Table 1: Correlations across Parameter Values, Union Network

Variables	Mean 0.5, SD 0.5	Mean 0.6, SD 0.5	Mean 0.6, SD 0.25	Mean 0.7, SD 0.5	Mean 0.7, SD 0.25
Mean 0.5, SD 0.5	1.00				
Mean 0.6, SD 0.5	0.93	1.00			
Mean 0.6, SD 0.25	0.51	0.56	1.00		
Mean 0.7, SD 0.5	0.94	0.98	0.55	1.00	
Mean 0.7, SD 0.25	0.35	0.36	0.84	0.39	1.00

Table 2: Correlations across Parameter Values, Family Network

Variables	Mean 0.5, SD 0.5	Mean 0.6, SD 0.5	Mean 0.6, SD 0.25	Mean 0.7, SD 0.5	Mean 0.7, SD 0.25
Mean 0.5, SD 0.5	1.00				
Mean 0.6, SD 0.5	0.96	1.00			
Mean 0.6, SD 0.25	0.86	0.89	1.00		
Mean 0.7, SD 0.5	0.96	0.99	0.86	1.00	
Mean 0.7, SD 0.25	0.91	0.90	0.83	0.93	1.00

Table 3: Correlations across Parameter Values, Friends Network

Variables	Mean 0.5, SD 0.5	Mean 0.6, SD 0.5	Mean 0.6, SD 0.25	Mean 0.7, SD 0.5	Mean 0.7, SD 0.25
Mean 0.5, SD 0.5	1.00				
Mean 0.6, SD 0.5	0.98	1.00			
Mean 0.6, SD 0.25	0.94	0.98	1.00		
Mean 0.7, SD 0.5	0.99	1.00	0.97	1.00	
Mean 0.7, SD 0.25	0.98	0.99	0.96	0.99	1.00

Table 4: Correlations across Parameter Values, Lender Network

Variables	Mean 0.5, SD 0.5	Mean 0.6, SD 0.5	Mean 0.6, SD 0.25	Mean 0.7, SD 0.5	Mean 0.7, SD 0.25
Mean 0.5, SD 0.5	1.00				
Mean 0.6, SD 0.5	0.98	1.00			
Mean 0.6, SD 0.25	0.92	0.97	1.00		
Mean 0.7, SD 0.5	0.97	0.99	0.97	1.00	
Mean 0.7, SD 0.25	0.95	0.98	0.97	0.99	1.00

Table 5: Correlations across Parameter Values, Solver Network

Variables	Mean 0.5, SD 0.5	Mean 0.6, SD 0.5	Mean 0.6, SD 0.25	Mean 0.7, SD 0.5	Mean 0.7, SD 0.25
Mean 0.5, SD 0.5	1.00				
Mean 0.6, SD 0.5	0.95	1.00			
Mean 0.6, SD 0.25	0.89	0.95	1.00		
Mean 0.7, SD 0.5	0.96	0.98	0.96	1.00	
Mean 0.7, SD 0.25	0.85	0.93	0.92	0.91	1.00

F Turnout and TPP by Network Type

Table 6 presents the regressions underlying Figure ??.

Table 6: TPP and Turnout Regressions

	(1) Presidential	(2) LC5 Chair	(3) Pooled
Eqm Participation Index (Union)	0.0048	0.045*	0.0048
	(0.023)	(0.021)	(0.026)
LC5 Election * Eqm Participation			0.040
			(0.023)
LC5 Chair Election			-0.35***
			(0.019)
Observations	15	15	30

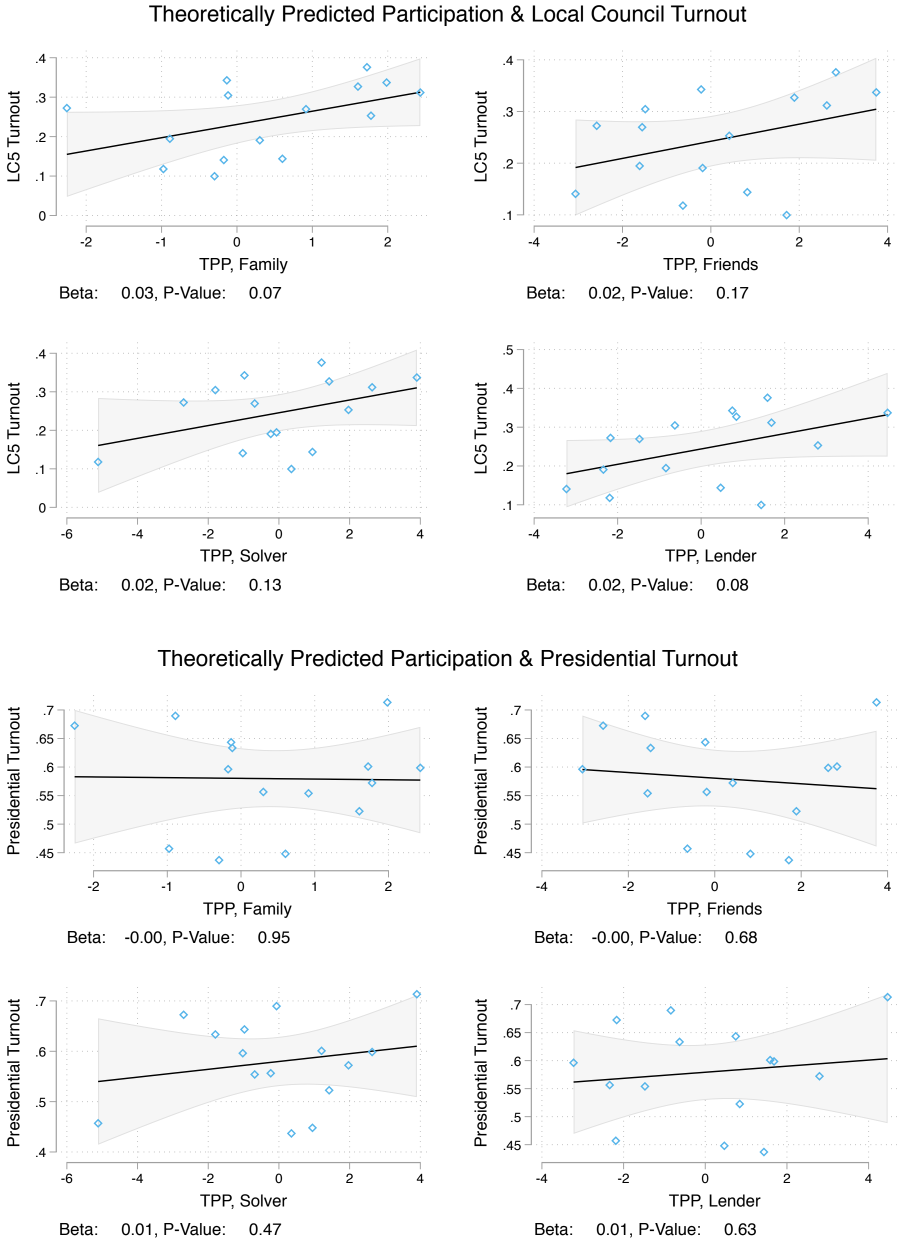
Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Pooled estimates clustered at village level.

Figure 4 below shows the relationship between turnout and TPP simulated separately on the four different types of networks that contribute to the Union network presented in the main paper. We find that our main results are generally insensitive to the type of network used to estimate TPP.

Figure 4



G Robustness Regression Tables

The following tables show robustness of our primary result when controlling for ELF (Column 2), controlling for education (Column 3), subsetting on the half of villages with the best polling-place-village correspondence (Column 4), and when including the exceptionally small 16th village surveyed (Column 5).

Table 7: Robustness for LC5 Election

	(1) Basic	(2) ELF	(3) Educ	(4) Size	(5) Concentrate	(6) W/16th	(7) Village
TPP (Union)	0.045*	0.044*	0.046*	0.038	0.041*	0.014	0.038
	(0.021)	(0.022)	(0.022)	(0.028)	(0.020)	(0.011)	(0.029)
ELF		-0.068					
		(0.13)					
Educ			0.058				
			(0.19)				
(Log) Network Size				0.060			-0.10
				(0.17)			(0.12)
Observations	15	15	15	15	8	16	16

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Robustness for Pres Election

	(1) Basic	(2) ELF	(3) Educ	(4) Size	(5) Concentrate	(6) W/16th	(7) Village
TPP (Union)	0.0048	0.0020	0.0041	0.0016	0.0058	-0.027**	0.0012
	(0.023)	(0.023)	(0.024)	(0.031)	(0.026)	(0.012)	(0.032)
ELF		-0.17					
		(0.14)					
Educ			-0.036				
			(0.21)				
(Log) Network Size				0.030			-0.13
				(0.18)			(0.13)
Observations	15	15	15	15	8	16	16

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

H Considering Non-Reciprocal Ties

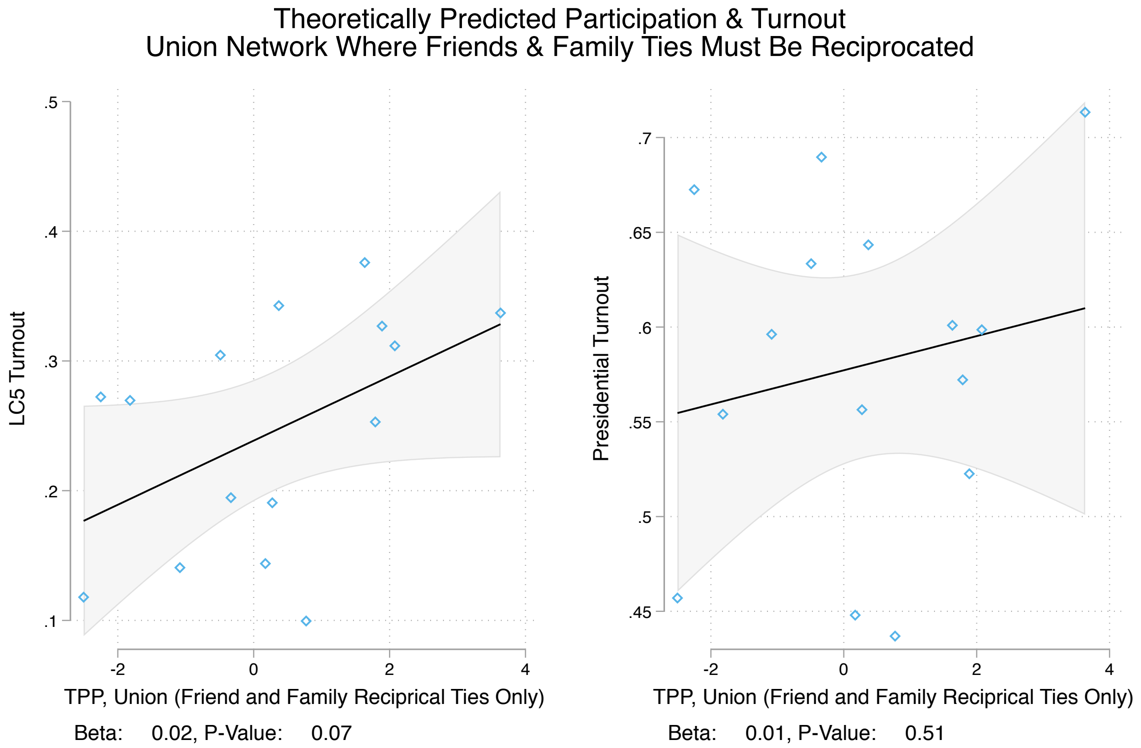
Figure 5 presents results when networks are created using only reciprocated ties to form Friends and Family ties (note the other two inputs into the Union network – the Lender / Solver networks – cannot be restricted in an analogous manner). The figures show results quite similar under this restriction.

As shown in Table 9, however, it is not clear that these restrictions are reasonable given the low average degree they generate in the Friend and Family networks. This may be due to censoring caused by the limited number of people individuals are allowed to list (5 family members and 5 friends), or failures to recall individuals.

Table 9: Network Summary Statistics: Including Only Reciprocated Friends and Family Ties

	Union	Friends	Family	Lender	Solver
Average Size	210.3	210.3	210.3	210.3	210.3
Average Num Connections	924.1	34.3	156.5	403.3	450.2
Average Degree	8.7	0.3	1.5	3.8	4.2
Min Size	160.0	160.0	160.0	160.0	160.0
Max Size	283.0	283.0	283.0	283.0	283.0

Figure 5



Notes: The above plot presents the partial correlation between Theoretically-Predicted Participation (TPP) and voter turnout in the Ugandan Presidential and LC5 Chair Elections where ties are only added between friends and family if ties are reciprocated. Grey bands indicate 95% confidence intervals. As detailed in Section ??, TPP is operationalized as the first principle component of normalized TPP scores across all parameter choices (as TPP is highly correlated across parameters). Turnout is shares of the adult village population. Regressions corresponding to these plots, as well as tests for the statistical significance of differences across elections can be found in Appendix F, along with analogous plots for different sub-networks. Adjustments for measurement /estimation error in TPP have not been made in these estimates; as a result their statistical difference from zero is likely under-stated, as measurement error in independent variables results in attenuation bias.

I Heterogeneous Effects

An important question is whether the social context effects we observe are driven by pressure to vote, or by pressure to coordinate around a given candidate. Table 10 presents tests for heterogeneous effects of TPP by (a) the degree to which village candidate preferences appear to be homogenous (i.e. everyone votes for the same candidate), and (b) the degree to which down-ballot races are competitive (there is no cross-village variation in the competitiveness of top-of-ticket races).

Vote homogeneity is measured using a simple Herfindahl index (which takes on a value of 1 if everyone votes for the same candidate, and a value of 0 if no one votes for the same candidate), while competitiveness is a fragmentation index (one minus the Herfindahl index of candidate vote shares across their entire electoral district – a value of 0 for races where one candidate won all votes or stood in an uncontested election, and values approaching 1 where votes are distributed evenly across a large number of candidates).

Note that the results in Column 4-5 are for down-ballot LC5 council seats, which are elected at the Sub-County level, so some villages faced the same slate of candidates. For that reason, results are clustered by sub-county (there are 10 clusters across the 15 villages). Similarly, Column 6 shows the relationship between competitiveness and TPP for down-ballot LC3 races, where villages in the same Parish face the same candidates, so results are clustered by parish (there are 13 clusters across the 15 villages). Column names report the term for the election in which candidates appeared using the election naming convention used throughout this paper (both LC5 council and LC3 candidates stood during the same election in which the LC5 Chair was selected).

First, as shown in Columns 2 and 4, we find that social context effects are somewhat smaller in villages where voters' candidate preferences are more homogeneous. Second, while the effect of TPP is slightly larger among villages with more competitive down-ballot LC3 local elections (Column 6), there is no evidence of a heterogeneous impact of TPP for villages facing more competitive down-ballot LC5 council seat elections (Column 5). While only suggestive (given our limited statistical power), taken together these results point towards network effects supporting a social norm of political participation, rather than facilitating strategic mobilization around a certain party or candidate.

Table 10: Social Norms and Political Mobilization

	(1) Pres	(2) Pres	(3) LC5Chair	(4) LC5Chair	(5) LC5Chair	(6) LC5Chair
TPP (Union)	0.0048	0.22	0.045*	0.086	0.0063	-0.052
	(0.023)	(0.17)	(0.021)	(0.14)	(0.048)	(0.046)
Pres Vote Homogeneity		0.32				
		(0.25)				
TPP X Vote Homogeneity		-0.42				
		(0.33)				
LC5Chair Vote Homogeneity				-0.13		
				(0.19)		
TPP X Vote Homogeneity				-0.086		
				(0.26)		
LC5 Council Competitive					0.27***	
					(0.048)	
TPP X LC5 Competitive					0.024	
					(0.093)	
LC3 Competitive						-0.098
						(0.097)
TPP X LC3 Competitive						0.21*
						(0.099)
Std of Block Voting		0.10		0.13		
Std of Competitiveness					0.24	0.17

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

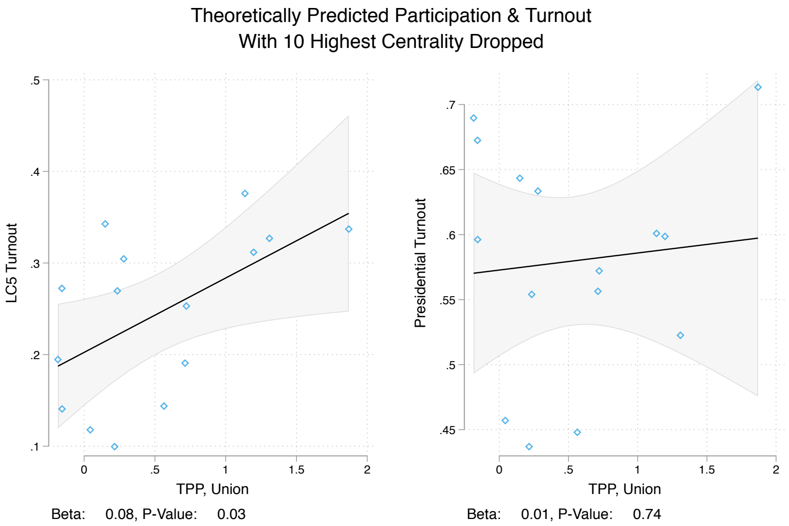
J Divide-The-Dollar Game

The divide-the-dollar game was organized as follows: first, subjects were given ten 100UGX coins. Subjects were then advised that they could split these coins between themselves and a stranger, who they were told will be "someone from Arua whom you do not know personally. We chose the stranger by randomly selecting someone living in Arua district from a long list."

K Dropping Highest Centrality Nodes

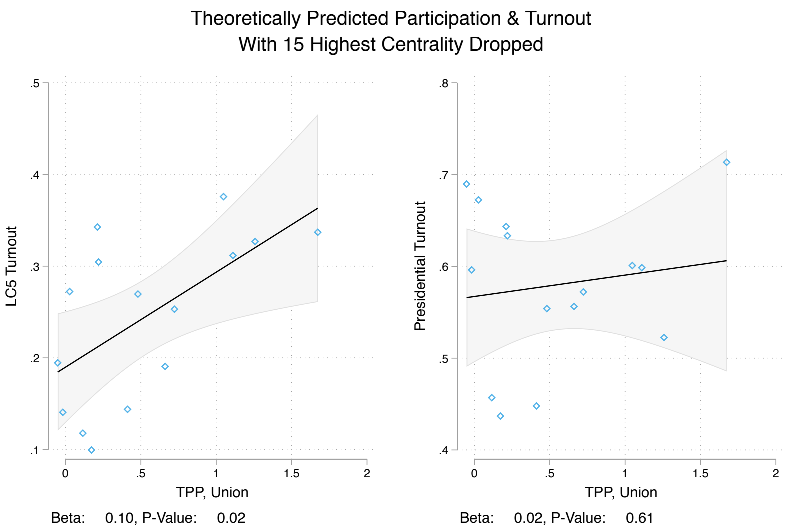
This section presents robustness checks to the analysis presented in Section ??. As shown in Figures 6 and 7, the results presented when we drop the 5 most central nodes in the network are similar to those found when we drop the 10 most central and 15 most central as well.

Figure 6



Notes: The above plot presents the partial correlation between voter turnout in the Ugandan Presidential and LC5 Chair Elections and a modified version of TPP. Grey bands indicate 95% confidence intervals. In particular, TPP has been re-calculated by removing the ten individuals with highest eigenvector centrality from each network and re-running TPP simulations on those networks.

Figure 7



Notes: The above plot presents the partial correlation between voter turnout in the Ugandan Presidential and LC5 Chair Elections and a modified version of TPP. Grey bands indicate 95% confidence intervals. In particular, TPP has been re-calculated by removing the fifteen individuals with highest eigenvector centrality from each network and re-running TPP simulations on those networks.

L Information Diffusion Simulation

We measure the ability of a network to efficiently diffuse information by running a simple diffusion model on our on empirical village networks. We then examine the average speed with which information spreads for each village.

Our decision to simulate this process is due to the fact that — as with social context influences — this is no simple statistic (like average degree or average shortest path length) which reliably summarizes the ability of a network to efficiently diffuse information when information spread is at least partially stochastic (Newman, N.d., p. 19-35).

Our simulation proceeds as follows:

1. At $t = 0$, one vertex v_0 in the network (selected with uniform probability) is endowed with a unique piece of knowledge. It is thus "informed" ($I(v_0) = 1$). All other vertices are assumed to be ignorant of this knowledge ($I(v_j) = 0 \;\; \forall j \in V \backslash 0$).
2. At $t = 1$, information spreads from v_0 to each of the neighbors of v_0, denoted $N(v_0)$ with i.i.d. probability $\frac{p}{|N(v_0)|} \in (0, 1)$.
3. Step 2 is then repeated indefinitely, where at each stage all "informed" vertices spread their knowledge to neighbors with i.i.d. probability p.

The ability of the network to support diffusion can then be specified as the number of people in the network that have become "informed" after s steps of the diffusion model. The larger the number of people "informed" for a given number of steps s, the more efficient a village's network.

Note that the probability of information diffusion from a vertex to her neighbors is normalized by the number of neighbors. This can be thought of as approximating the idea that individuals can only have so many interactions in a given period of time. This normalization more closely approximates the idea that all individuals have the same probability of interacting and sharing information with at least a friend in a given period, a dynamic suggested by recent work on information diffusion elsewhere in Uganda (Larson and Lewis, 2017). With that said, results look similar without the normalization.

L.1 Information Diffusion Summary Statistics

Table 11 below shows the correlation in the share of individuals in each village informed at different step thresholds, with different spread probabilities, and with different network specifications. As the table shows, inter-parameter correlations are quite high, and so an index is created for expositional ease by taking the first component of a PCA index for each network specification.

Table 11: Diffusion Correlations across Parameter Values

Variables	p 0.60, 10 steps, Union	p 0.60, 20 steps, Union	p 0.35, 10 steps, Union	p 0.35, 20 steps, Union
p 0.60, 10 steps, Union	1.00			
p 0.60, 20 steps, Union	0.76	1.00		
p 0.35, 10 steps, Union	0.96	0.60	1.00	
p 0.35, 20 steps, Union	0.97	0.83	0.92	1.00

M Information Measures, Turnout, and TPP Regressions

Table 12: Turnout and TPP with Information Measure Controls

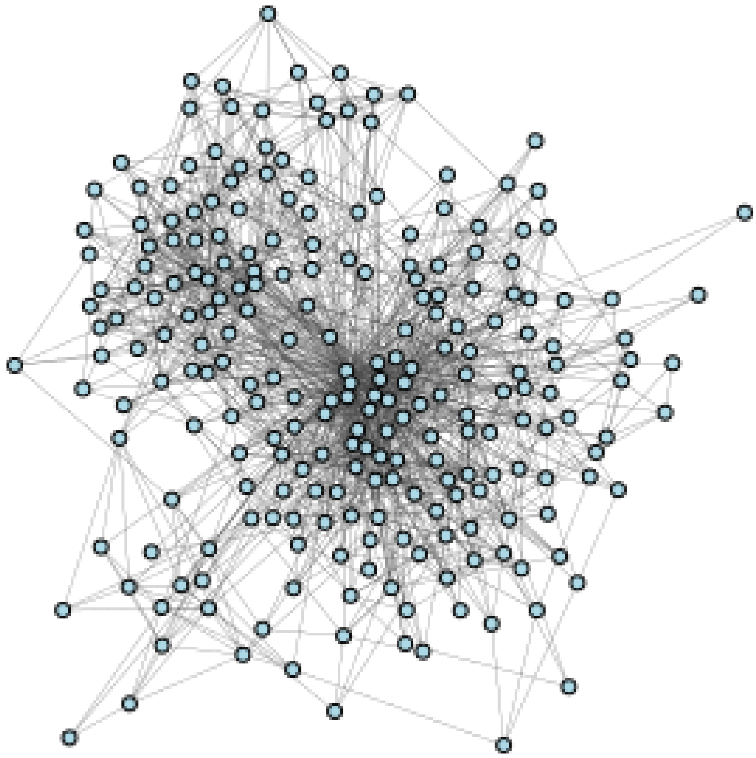
	(1) LC5 Chair	(2) LC5 Chair	(3) Presidential	(4) Presidential
Eqm Participation Index (Union)	0.055**	0.044*	0.016	0.0053
	(0.019)	(0.022)	(0.021)	(0.025)
Share of Village Aware of UBridge	0.33*		0.37*	
	(0.16)		(0.18)	
Info Diffusion Simulation 1st Component		-0.0020		0.0011
		(0.012)		(0.014)
Observations	15	15	15	15

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

N Village Network Plots

Figure 8: 237 Vertices, 1,457 Edges. Eqm Participation Index Value: -1.25



References

Larson, Jennifer M and Janet I Lewis. 2017. "Ethnic Networks." American Journal of Political Science 61(2):350–364.

Newman, Mark. N.d. 17 Epidemics on networks. Networks: An Introduction Oxford University Press.

Siegel, David A. 2009. "Social Networks and Collective Action." American Journal of Political Science 53(1):122–138.

Figure 9: 229 Vertices, 1,723 Edges. Eqm Participation Index Value: 0.38

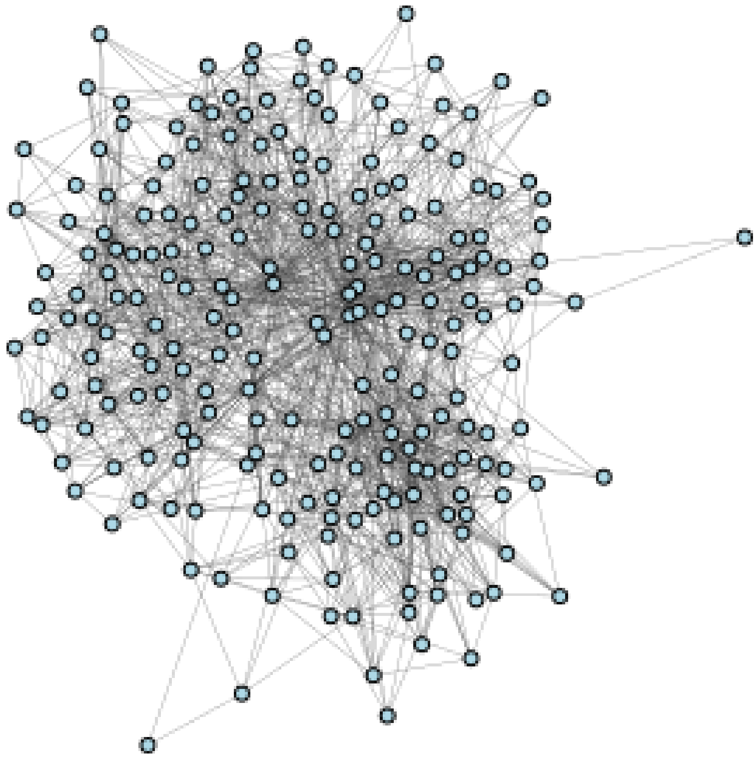


Figure 10: 204 Vertices, 1,618 Edges. Eqm Participation Index Value: 0.28

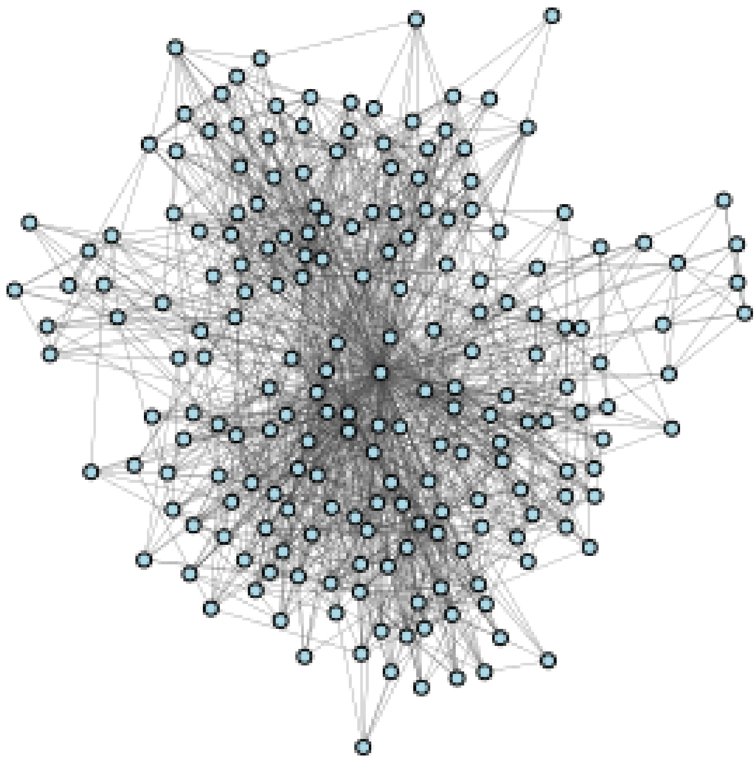


Figure 11: 197 Vertices, 1,283 Edges. Eqm Participation Index Value: -0.69

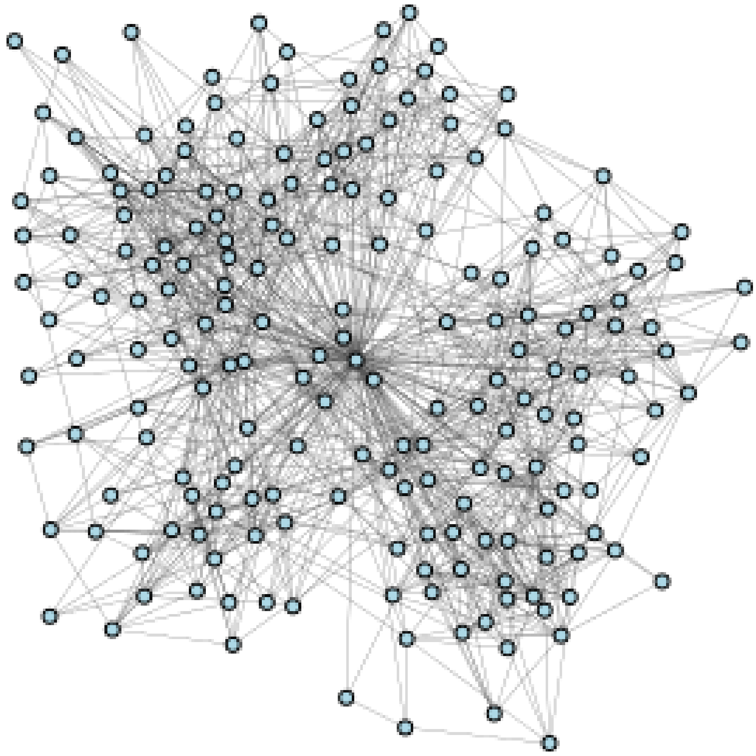


Figure 12: 30 Vertices, 176 Edges. Eqm Participation Index Value: -6.93

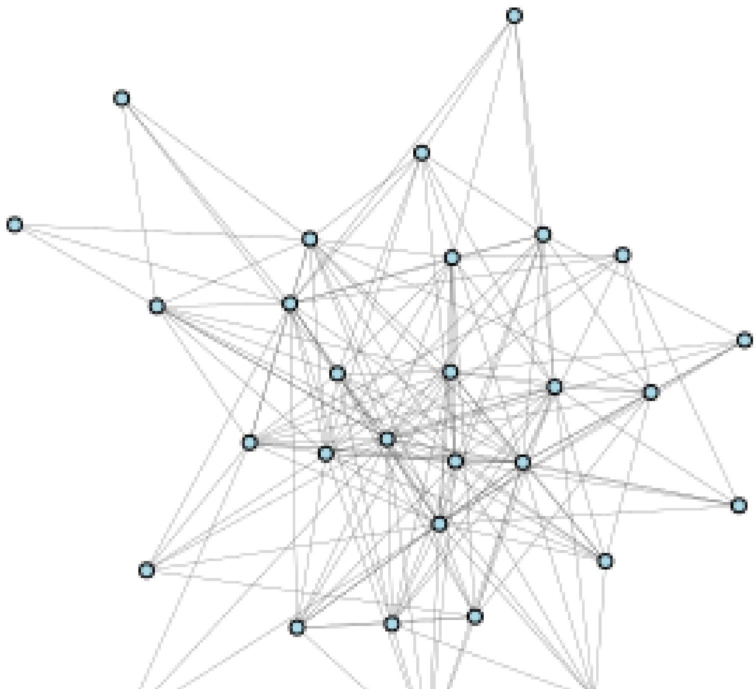


Figure 13: 189 Vertices, 1,370 Edges. Eqm Participation Index Value: -0.00

Figure 14: 283 Vertices, 2,797 Edges. Eqm Participation Index Value: 2.47

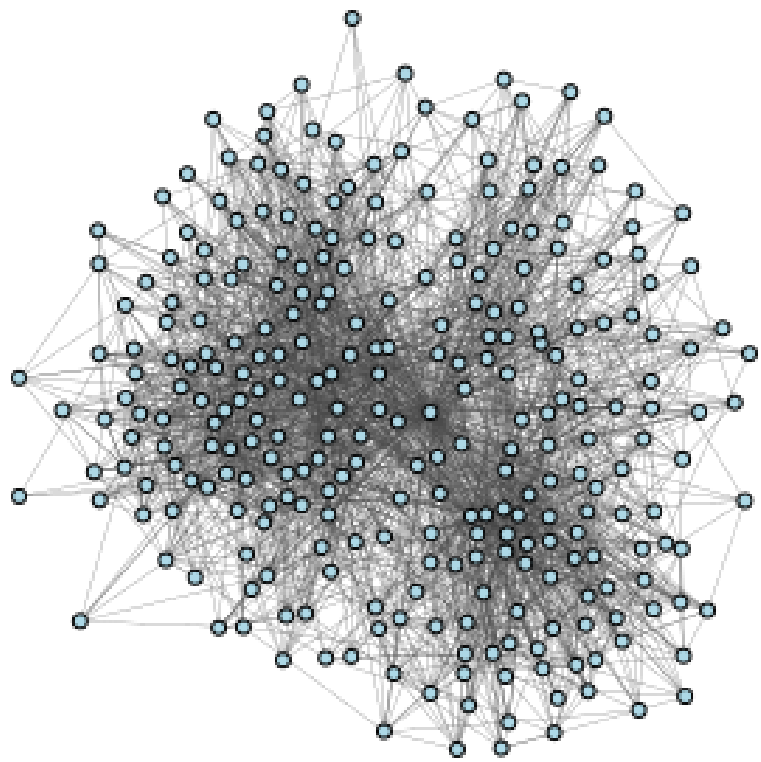


Figure 15: 192 Vertices, 1,490 Edges. Eqm Participation Index Value: 0.45

Figure 16: 163 Vertices, 1,247 Edges. Eqm Participation Index Value: -0.10

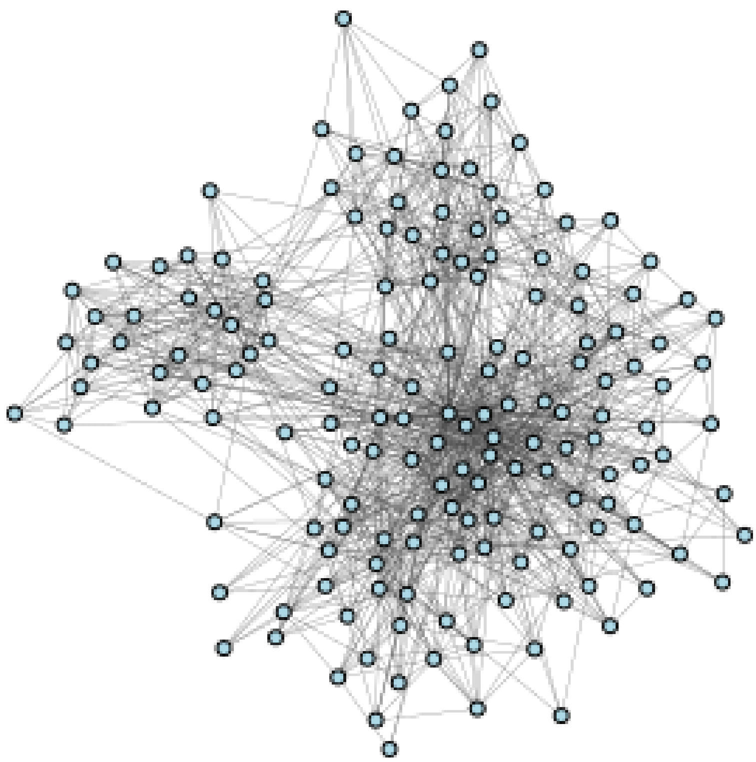


Figure 17: 168 Vertices, 1,189 Edges. Eqm Participation Index Value: -0.58

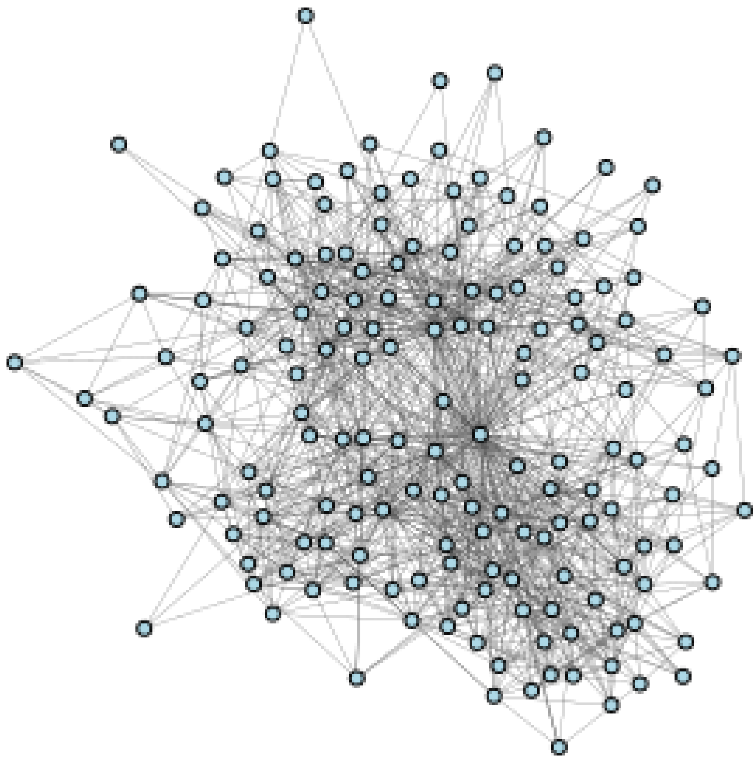


Figure 18: 254 Vertices, 2,423 Edges. Eqm Participation Index Value: 1.73

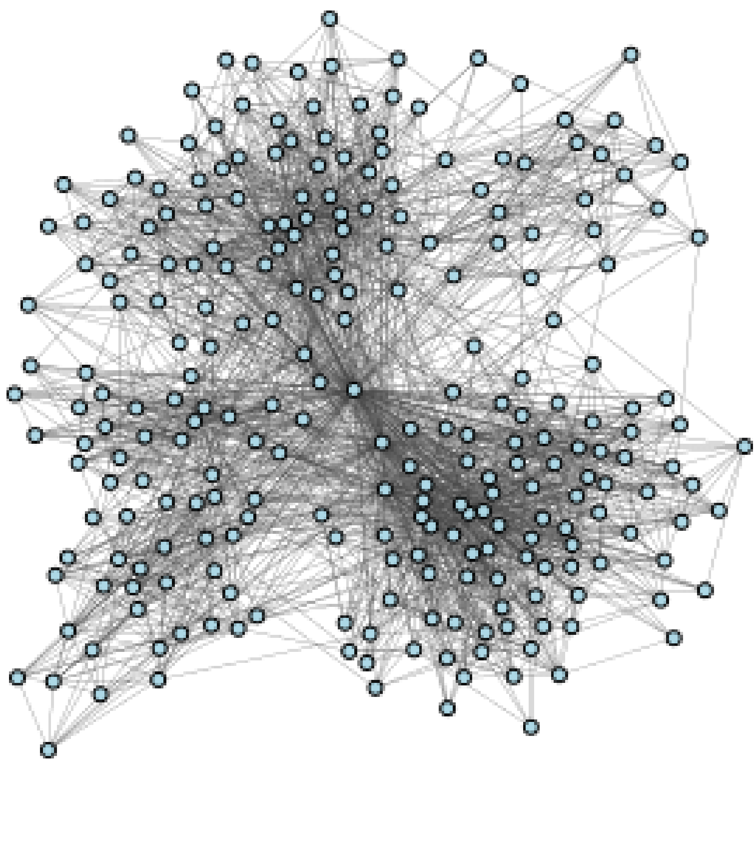


Figure 19: 225 Vertices, 2,016 Edges. Eqm Participation Index Value: 1.22

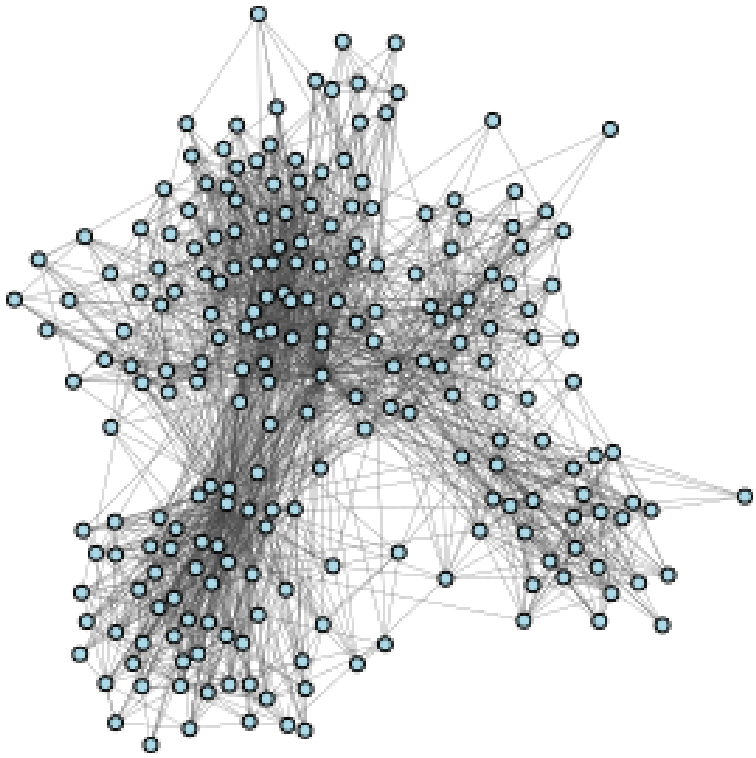


Figure 20: 205 Vertices, 1,857 Edges. Eqm Participation Index Value: 1.48

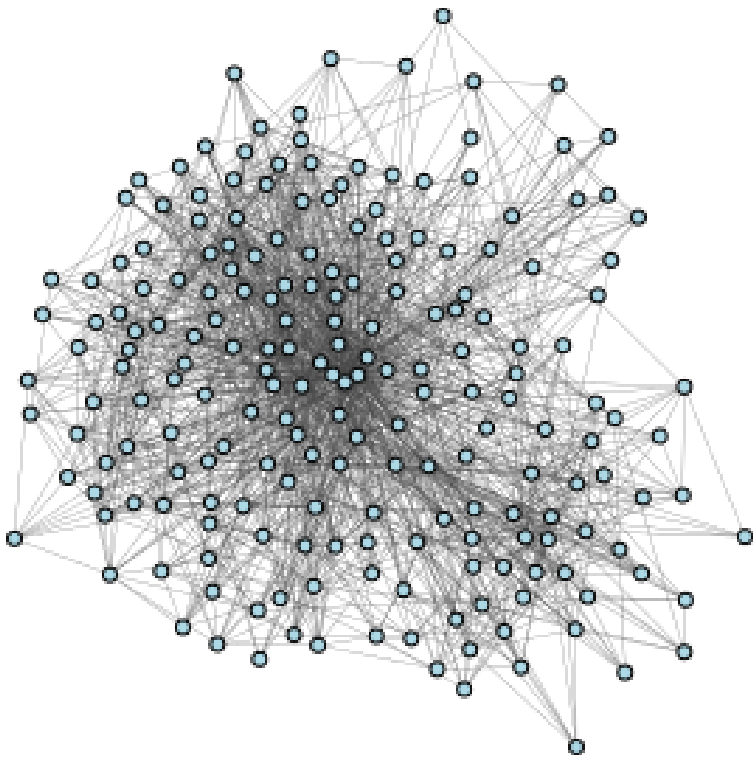


Figure 21: 263 Vertices, 2,272 Edges. Eqm Participation Index Value: 1.44

Figure 22: 185 Vertices, 1,516 Edges. Eqm Participation Index Value: 0.54

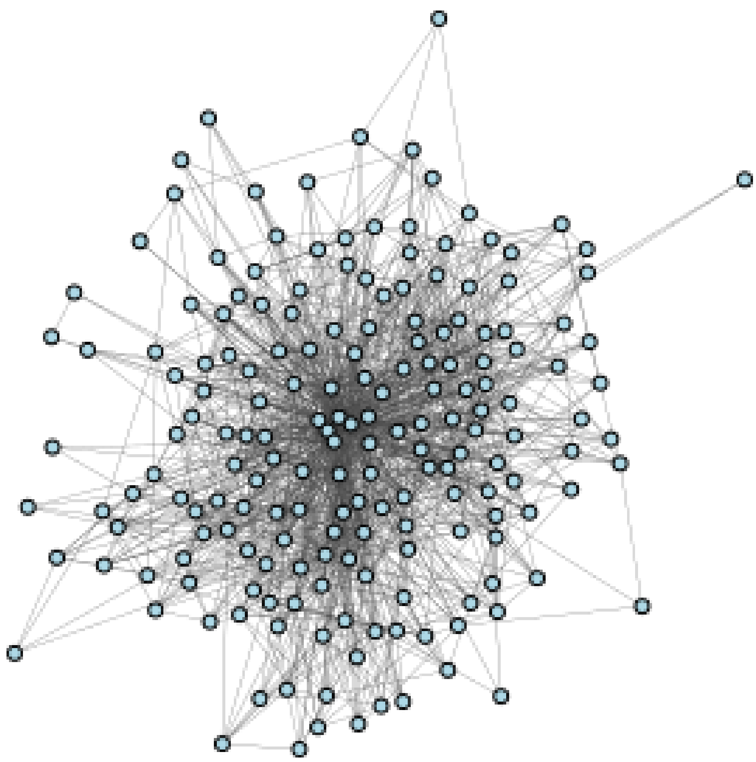


Figure 23: 160 Vertices, 1,150 Edges. Eqm Participation Index Value: -0.45