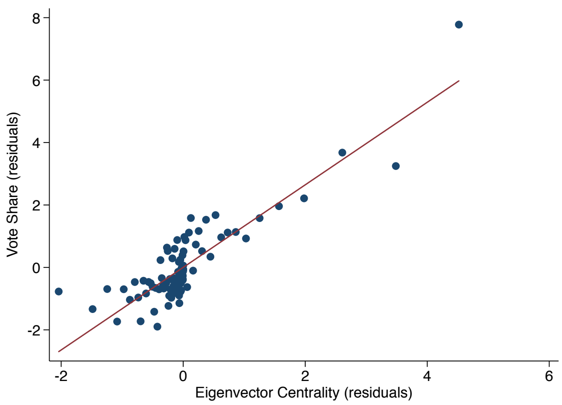
Politician Family Networks and Electoral Outcomes: Evidence from the Philippines

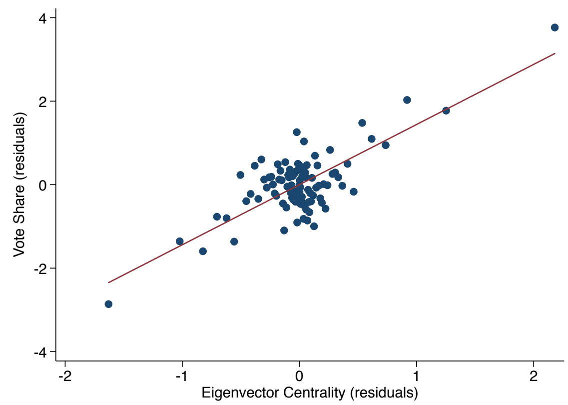
Online Appendix

Cesi Cruz, Julien Labonne, and Pablo Querubin

A.1 Additional Figures



Panel A: Candidate fixed-effects



Panel B: Controls from Column 4 in Table 2

Figure A.1: Scatterplots of binned residuals.

We plot binned residuals from regressions of candidate vote share and eigenvector centrality. In Panel A, the regressions only include candidate fixed effects. In Panel B, the regressions include candidate and village fixed-effects and control for the number of relatives, number of female relatives, number of relatives in each education category and the number of relatives in each occupation category.

A.2 Additional Tables

Table A.1: Correlation Centrality Measures

	Eigenvector (1)	Between (2)	Pagerank (3)	Katz (0.01) (4)	Katz (0.11) (5)	Katz (0.21) (6)	Katz (0.31) (7)
Panel A: Municipal-level measures (all families)							
Eigenvector	1						
Between	0.780	1					
Pagerank	0.736	0.894	1				
Katz (0.01)	0.871	0.781	0.686	1			
Katz (0.11)	0.483	0.367	0.366	0.544	1		
Katz (0.21)	0.256	0.185	0.223	0.334	0.333	1	
Katz (0.31)	0.138	0.0989	0.134	0.200	0.199	0.333	1
Panel B: Village-level measures (all families)							
Eigenvector	1						
Between	0.719	1					
Pagerank	0.619	0.777	1				
Katz (0.01)	0.717	0.747	0.574	1			
Katz (0.11)	0.822	0.674	0.499	0.871	1		
Katz (0.21)	0.568	0.482	0.357	0.599	0.580	1	
Katz (0.31)	0.285	0.314	0.245	0.406	0.335	0.313	1
Panel C: Village-level measures (2010 candidates)							
Eigenvector	1						
Between	0.811	1					
Pagerank	0.703	0.858	1				
Katz (0.01)	0.828	0.751	0.569	1			
Katz (0.11)	0.848	0.701	0.509	0.918	1		
Katz (0.21)	0.607	0.526	0.435	0.644	0.581	1	
Katz (0.31)	0.376	0.340	0.323	0.449	0.381	0.428	1

Notes: Correlation between the various centrality measures used in the paper. Authors' calculations. Panel A: n= 3,882,261. Panel B: n= 6,704,256. Panel C: n=50,228.

Table A.2: Descriptive Statistics - Municipal*Family Level [All Families]

Variable Name	Observations (1)	Mean (2)	Std. Dev. (3)
Ran for mayor in 2010 (*100)	3,882,261	0.09	(3.06)
Eigenvector	3,882,261	0.00	(1.00)
Betweenness	3,882,261	0.00	(1.00)
PageRank	3,882,261	0.00	(1.00)
Katz (0.01)	3,882,261	0.00	(1.00)
Katz (0.11)	3,882,261	0.00	(1.00)
Katz (0.21)	3,882,261	0.00	(1.00)
Katz (0.31)	3,882,261	0.00	(1.00)
Nb Relatives	3,882,261	6.73	(24.73)
Nb Female Relatives	3,882,261	3.31	(12.33)
Nb of relatives with education levels:			
No Grade Completed	3,882,261	0.58	(3.81)
Kinder or Daycare	3,882,261	0.03	(0.28)
Grade 1	3,882,261	0.17	(0.95)
Grade 2	3,882,261	0.24	(1.22)
Grade 3	3,882,261	0.33	(1.53)
Grade 4	3,882,261	0.41	(1.86)
Grade 5	3,882,261	0.40	(1.71)
Grade 6	3,882,261	1.16	(4.96)
1st Year High School	3,882,261	0.34	(1.38)
2nd Year High School	3,882,261	0.44	(1.75)
3rd Year High School	3,882,261	0.43	(1.73)
4th Year High School	3,882,261	1.11	(4.56)
1st Year College	3,882,261	0.21	(0.95)
2nd Year College	3,882,261	0.24	(1.13)
3rd Year College	3,882,261	0.11	(0.58)
4th Year College	3,882,261	0.07	(0.46)
College Graduate	3,882,261	0.44	(2.25)
Above (MA/PhD)	3,882,261	0.01	(0.18)
Nb of relatives with occupation:			
Special Occupations	3,882,261	0.06	(0.53)
Officials, Managers, Supervisors	3,882,261	0.09	(0.57)
Professionals	3,882,261	0.10	(0.66)
Technicians, Associate Professionals	3,882,261	0.02	(0.22)
Clerks	3,882,261	0.03	(0.24)
Service, Shop, Market Sales Workers	3,882,261	0.23	(1.37)
Farmers, Forestry Workers, Fishermen	3,882,261	1.39	(6.43)
Trades, Related workers	3,882,261	0.09	(0.65)
Plant, Machine Operators, Assemblers	3,882,261	0.06	(0.44)
Laborers, Unskilled Workers	3,882,261	0.79	(3.68)
None	3,882,261	2.81	(11.70)
Share of municipal land owned	2,908,192	0.01	(0.33)
Land Area	2,908,192	928.01	(24954.68)
Landowning status:			
Landowner [*100]	2,908,192	0.73	(8.50)
Top 50% Landowner [*100]	2,908,192	0.37	(6.04)
Top 25% Landowner [*100]	2,908,192	0.18	(4.23)
Top 10% Landowner [*100]	2,908,192	0.07	(2.60)
Top Landowner [*100]	2,908,192	0.01	(0.73)
Colonial status:			
Spanish Elite (municipal) [*100]	1,385,804	0.05	(2.17)
Spanish Elite (provincial) [*100]	2,950,234	0.31	(5.52)
Taft Elite (municipal) [*100]	493,859	0.06	(2.38)
Taft Elite (provincial) [*100]	1,364,295	0.50	(7.05)

Notes: Authors' calculations.

Table A.3: Descriptive Statistics - Precinct*Family-Level [Candidates only]

Variable Name	Observations (1)	Mean (2)	Std. Dev. (3)
Vote Share	50,228	25.85	(21.55)
Eigenvector	50,228	0.00	(1.00)
Betweenness	50,228	0.00	(1.00)
PageRank	50,228	0.00	(1.00)
Katz (0.01)	50,228	0.00	(1.00)
Katz (0.11)	50,228	0.00	(1.00)
Katz (0.21)	50,228	0.00	(1.00)
Katz (0.31)	50,228	0.00	(1.00)
Nb relatives	50,228	8.88	(23.73)
Nb Female Relatives	50,228	4.44	(12.16)
Nb of relatives with education levels:			
No Grade Completed	50,228	0.60	(3.87)
Kinder or Daycare	50,228	0.04	(0.27)
Grade 1	50,228	0.15	(0.75)
Grade 2	50,228	0.22	(0.94)
Grade 3	50,228	0.29	(1.12)
Grade 4	50,228	0.37	(1.40)
Grade 5	50,228	0.39	(1.38)
Grade 6	50,228	1.28	(4.09)
1st Year High School	50,228	0.39	(1.33)
2nd Year High School	50,228	0.54	(1.74)
3rd Year High School	50,228	0.57	(1.75)
4th Year High School	50,228	1.61	(4.87)
1st Year College	50,228	0.37	(1.25)
2nd Year College	50,228	0.47	(1.55)
3rd Year College	50,228	0.23	(0.84)
4th Year College	50,228	0.16	(0.71)
College Graduate	50,228	1.15	(3.87)
Above (MA/PhD)	50,228	0.04	(0.43)
Nb of relatives with occupation:			
Special Occupations	50,228	0.11	(0.79)
Officials, Managers, Supervisors	50,228	0.27	(1.20)
Professionals	50,228	0.28	(1.28)
Technicians, Associate Professionals	50,228	0.05	(0.35)
Clerks	50,228	0.06	(0.42)
Service, Shop, Market Sales Workers	50,228	0.32	(1.52)
Farmers, Forestry Workers, Fishermen	50,228	1.35	(5.37)
Trades, Related workers	50,228	0.16	(0.88)
Plant, Machine Operators, Assemblers	50,228	0.10	(0.54)
Laborers, Unskilled Workers	50,228	0.91	(3.34)
None	50,228	3.74	(11.76)
Share of village land owned	34,972	0.01	(0.06)
Land Area	34,972	2,928.14	(51660.64)
Landowning status:			
Landowner [*100]	34,972	1.95	(13.84)
Top 50% Landowner [*100]	34,972	1.26	(11.15)
Top 25% Landowner [*100]	34,972	0.81	(8.94)
Top 10% Landowner [*100]	34,972	0.66	(8.08)
Top Landowner [*100]	34,972	0.58	(7.58)
Colonial status:			
Spanish Elite (municipal) [*100]	21,587	3.12	(17.39)
Spanish Elite (provincial) [*100]	38,937	4.51	(20.75)
Taft Elite (municipal) [*100]	7,490	1.71	(12.96)
Taft Elite (provincial) [*100]	18,548	5.31	(22.42)

Notes: Authors' calculations.

Table A.4: Descriptive Statistics from Other Surveys

Variable Name	Observations	Mean	Std. Dev.
	(1)	(2)	(3)
Panel A: Variables from the NHTS-PR (village-level):			
Number of services	12,874	0.82	(0.56)
Philhealth	12,874	0.29	(0.23)
Panel B: Variables from the 2013 Ilocos Survey (candidate*village-level):			
Policy Alignment	629	58.25	(10.55)
Support candidate	658	2.63	(0.97)
Traits			
Honest	658	0.60	(0.29)
Approachable	658	0.66	(0.29)
Experienced	658	0.58	(0.36)
Connected	658	0.59	(0.32)
Panel C: Variables from the 2016 Ilocos Survey (individual-level):			
Vote Buying			
Overall	3,423	0.40	(0.49)
By Incumbent	3,189	0.24	(0.43)
By Challenger	3,189	0.16	(0.37)
Ease of Access to			
Endorsement Letter	3,462	6.78	(2.75)
Funeral Expense	3,463	7	(2.67)
Medical Expense	3,467	7.43	(2.59)
Police Clearance	3,470	8.55	(2.21)
Barangay Clearance	3,475	9.20	(1.78)
Death Certificate	3,463	7.53	(2.68)
Business Permit	3,462	6.18	(3.03)

Notes: Authors' calculations.

Table A.5: Candidate Networks and Precinct-Level Vote Share - Various Ways of Aggregating Centrality

	(1)	(2)	(3)	(4)
(Avg.) Eigenvector	1.663			
	(0.275)			
Eigenvector (Last Name)		1.106		1.352
		(0.232)		(0.242)
Eigenvector (Middle Name)			0.365	0.738
			(0.181)	(0.190)
Observations	50,228	50,228	50,228	50,228
R-squared	0.813	0.812	0.812	0.813

Results from precinct*candidate regressions. The dependent variable is vote share (measured as a proportion of the registered population). Eigenvector centrality is normalized to be mean 0 and standard deviation 1. All regressions include candidate and village fixed-effects. Regressions control for the number of relatives, number of female relatives, number of relatives in each education category and number of relatives in each occupation category. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.6: Candidate Networks and Precinct-Level Vote Share - Using Weighted Networks

	(1)	(2)	(3)	(4)
Eigenvector	1.322	1.030	0.954	1.441
	(0.116)	(0.136)	(0.132)	(0.251)
Observations	50,228	50,228	50,228	50,228
R-squared	0.784	0.785	0.786	0.812

Notes: Results from precinct*candidate regressions. The dependent variable is vote share (measured as a proportion of the registered population). Eigenvector centrality is normalized to be mean 0 and standard deviation 1. All regressions include candidate fixed-effects. Regressions control for the number of relatives (Columns 2-4), number of female relatives (Columns 2-4), number of relatives in each education category (Columns 3-4) and number of relatives in each occupation category (Columns 3-4). Village fixed effects are included in Column 4. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.7: Candidate Networks and Precinct-Level Vote Share - Controlling for Land Wealth

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Eigenvector	0.869	0.863	0.868	0.846	0.851	0.867	0.880	0.873
	(0.257)	(0.258)	(0.257)	(0.258)	(0.260)	(0.259)	(0.258)	(0.257)
Share land		6.145						
		(3.461)						
Land area			0.046					
			(0.046)					
Landowner				3.904				
				(1.258)				
Top 50% landowner					3.909			
					(1.418)			
Top 25% landowner						4.502		
						(1.936)		
Top 10% landowner							5.129	
							(2.193)	
Top landowner								4.746
								(2.243)
Observations	34,972	34,972	34,972	34,972	34,972	34,972	34,972	34,972
R-squared	0.838	0.838	0.838	0.838	0.838	0.838	0.838	0.838

Results from precinct*candidate regressions. The dependent variable is vote share (measured as a proportion of the registered population). Eigenvector centrality is normalized to be mean 0 and standard deviation 1. All regressions include candidate and village fixed-effects. Regressions control for the number of relatives, number of female relatives, number of relatives in each education category and number of relatives in each occupation category. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.8: Candidate Networks and Precinct-Level Vote Share - Excluding Landed Elites

	(1)	(2)	(3)	(4)	(5)
Eigenvector	1.027	1.208	1.160	0.991	0.994
	(0.374)	(0.339)	(0.313)	(0.297)	(0.297)
Exclude :	Any Landowner	top 50%	top 25%	top 10%	top
Observations	27,351	29,319	30,555	31,277	31,556
R-squared	0.868	0.855	0.850	0.850	0.849

Results from precinct*candidate regressions. The dependent variable is vote share (measured as a proportion of the registered population). Eigenvector centrality is normalized to be mean 0 and standard deviation 1. All regressions include candidate and village fixed-effects. Regressions control for the number of relatives, number of female relatives, number of relatives in each education category and number of relatives in each occupation category. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.9: Candidate Networks and Precinct-Level Vote Share - Interactions with Landed Elites

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Eigenvector	0.876	0.841	0.822	0.896	0.885	0.889	0.891
	(0.262)	(0.264)	(0.268)	(0.265)	(0.262)	(0.262)	(0.263)
Eigenvector*Land	-1.442	0.137	0.511	-0.770	-0.199	0.346	-1.100
	(3.971)	(0.075)	(1.155)	(1.113)	(1.872)	(2.036)	(2.119)
Land Measure:	Share	Land Area	Landowner	Top 50%	Top 25%	Top 10%	Top
Observations	34,972	34,972	34,972	34,972	34,972	34,972	34,972
R-squared	0.839	0.839	0.839	0.839	0.839	0.839	0.839

Results from precinct*candidate regressions. The dependent variable is vote share (measured as a proportion of the registered population). Eigenvector centrality is normalized to be mean 0 and standard deviation 1. All regressions include candidate and village fixed-effects. Regressions control for the number of relatives, number of female relatives, number of relatives in each education category and number of relatives in each occupation category. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.10: Candidate Networks and Precinct-Level Vote Share - Excluding Colonial elites

	(1)	(2)	(3)	(4)
Eigenvector	0.901	1.142	1.562	0.890
	(0.286)	(0.268)	(0.404)	(0.334)
Exclude :	Spanish elite		Taft commission	
	Municipal	Provincial	Municipal	Provincial
Observations	20,557	35,797	7,304	16,990
R-squared	0.848	0.847	0.851	0.848

Results from precinct*candidate regressions. The dependent variable is vote share (measured as a proportion of the registered population). Eigenvector centrality is normalized to be mean 0 and standard deviation 1. All regressions include candidate and village fixed-effects. Regressions control for the number of relatives, number of female relatives, number of relatives in each education category and number of relatives in each occupation category. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.11: Candidate Networks and Precinct-Level Vote Share - Interaction with Colonial elites

	(1)	(2)	(3)	(4)
Eigenvector	0.983	1.178	1.712	0.899
	(0.281)	(0.259)	(0.420)	(0.335)
Eigenvector*Elite	-1.026	-0.759	-0.230	-0.489
	(0.769)	(0.667)	(2.107)	(0.797)
Colonial Measure :	Spanish elite		Taft commission	
	Municipal	Provincial	Municipal	Provincial
Observations	20,557	35,797	7,304	16,990
R-squared	0.848	0.847	0.851	0.848

Results from precinct*candidate regressions. The dependent variable is vote share (measured as a proportion of the registered population). Eigenvector centrality is normalized to be mean 0 and standard deviation 1. All regressions include candidate and village fixed-effects. Regressions control for the number of relatives, number of female relatives, number of relatives in each education category and number of relatives in each occupation category. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.12: Candidate Networks and Precinct-Level Vote Share (Excluding "home" village)

	(1)	(2)	(3)	(4)
Eigenvector	1.025	0.782	0.732	0.870
	(0.119)	(0.137)	(0.133)	(0.245)
Observations	46,319	46,319	46,319	46,319
R-squared	0.792	0.792	0.793	0.827

Notes: Results from precinct*candidate regressions. The dependent variable is vote share (measured as a proportion of the registered population). Eigenvector centrality is normalized to be mean 0 and standard deviation 1. All regressions include candidate fixed-effects. Regressions control for the number of relatives (Columns 2-4), number of female relatives (Columns 2-4), number of relatives in each education category (Columns 3-4) and number of relatives in each occupation category (Columns 3-4). Village fixed effects are included in Column 4. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.13: Candidate Networks and Precinct-Level Vote Share - Excluding Families with Previous Electoral Experience

	(1)	(2)	(3)	(4)
Eigenvector	1.658	0.990	1.053	1.797
	(0.212)	(0.233)	(0.243)	(0.684)
Observations	15,394	15,394	15,394	15,394
R-squared	0.760	0.761	0.763	0.889

Results from precinct*candidate regressions. The dependent variable is vote share (measured as a proportion of the registered population). Eigenvector centrality is normalized to be mean 0 and standard deviation 1. All regressions include candidate fixed-effects. Regressions control for the number of relatives (Columns 2-4), number of female relatives (Columns 2-4), number of relatives in each education category (Columns 3-4) and number of relatives in each occupation category (Columns 3-4). Village fixed effects are included in Column 4. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.14: Candidate Networks and Precinct-Level Vote Share - Network Restricted to Individuals > 45

	(1)	(2)	(3)	(4)
Panel A: OLS with over 45				
Eigenvector	1.108	0.667	0.615	0.805
	(0.103)	(0.126)	(0.120)	(0.222)
Observations	49,108	49,108	49,1088	49,108
R-squared	0.783	0.783	0.785	0.814
Panel B: IV with over 45				
Eigenvector	1.376	1.050	0.987	1.359
	(0.125)	(0.184)	(0.186)	(0.306)
Observations	49,108	49,108	49,108	49,108

Notes: Results from OLS (Panel A) and IV (Panel B) precinct*candidate regressions. The dependent variable is vote share (measured as a proportion of the registered population). Eigenvector centrality is normalized to be mean 0 and standard deviation 1. In Panel B, eigenvector centrality from the network of individuals older than 45 is used as an instrument for eigenvector centrality in the full network. All regressions include candidate fixed-effects. Regressions control for the number of relatives (Columns 2-4), number of female relatives (Columns 2-4), number of relatives in each education category (Columns 3-4) and number of relatives in each occupation category (Columns 3-4). Village fixed effects are included in Column 4. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.15: Candidate Networks and Precinct-Level Vote Share - Use Percentile Rank

	(1)	(2)	(3)	(4)
Eigenvector (rank)	2.276	1.362	1.189	3.690
	(0.226)	(0.229)	(0.224)	(0.481)
Observations	50,228	50,228	50,228	50,228
R-squared	0.783	0.784	0.785	0.813

Notes: Results from precinct*candidate regressions. The dependent variable is vote share (measured as a proportion of the registered population). Eigenvector (rank) is the rank of the candidate's family in the distribution of eigenvector centrality in each village. All regressions include candidate fixed-effects. Regressions control for the number of relatives (Columns 2-4), number of female relatives (Columns 2-4), number of relatives in each education category (Columns 3-4) and number of relatives in each occupation category (Columns 3-4). Village fixed effects are included in Column 4. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.16: Candidate Networks and Precinct-Level Vote Share - Removing Outliers

	(1)	(2)	(3)	(4)
		Outliers		w/o
	1%	5%	10%	ARMM
Eigenvector	1.028	1.157	1.658	0.851
	(0.257)	(0.400)	(0.667)	(0.237)
Observations	49,341	47,717	45,207	42,299
R-squared	0.817	0.821	0.830	0.829

Results from precinct*candidate regressions. The dependent variable is vote share (measured as a proportion of the registered population). Eigenvector centrality is normalized to be mean 0 and standard deviation 1. All regressions include candidate and village fixed-effects. Regressions control for the number of relatives, number of female relatives, number of relatives in each education category and number of relatives in each occupation category. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.17: Strong Vs. Weak Candidates

	(1)	(2)	(3)	(4)
Panel A: Incumbent Vs. Challengers				
Eigenvector	1.429	1.073	1.014	1.404
	(0.131)	(0.154)	(0.154)	(0.265)
Eigenvector*Incumbent	-0.351	-0.038	-0.049	0.401
	(0.228)	(0.288)	(0.293)	(0.478)
Observations	50,228	50,228	50,228	50,228
R-squared	0.784	0.785	0.787	0.814
Panel B: Only 'Serious' Candidates				
Eigenvector	1.448	1.205	1.106	1.988
	(0.137)	(0.165)	(0.162)	(0.435)
Observations	34,441	34,441	34,441	34,441
R-squared	0.610	0.610	0.612	0.694

Notes: Results from precinct*candidate regressions. The dependent variable is vote share (measured as a proportion of the registered population). Eigenvector centrality is normalized to be mean 0 and standard deviation 1. All regressions include candidate fixed-effects. Regressions control for the number of relatives (Columns 2-4), number of female relatives (Columns 2-4), number of relatives in each education category (Columns 3-4) and number of relatives in each occupation category (Columns 3-4). In Panel A, all control variables are interacted with both the incumbent dummy and with eigenvector centrality. Village fixed effects are included in Column 4. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.18: Candidate Networks and Precinct-Level Vote Share - Alternative Centrality Measures

	(1)	(2)	(3)	(4)	(5)	(6)
	Between	Pagerank	Katz - Decay factor:			
			.01	.11	.21	.31
Centrality	1.371	1.555	2.014	1.073	0.763	0.798
	(0.240)	(0.299)	(0.403)	(0.327)	(0.169)	(0.152)
Observations	50,228	50,228	50,228	50,228	50,228	50,228
R-squared	0.812	0.813	0.812	0.812	0.812	0.812

Results from precinct*candidate regressions. The dependent variable is vote share (measured as a proportion of the registered population). The network measures are normalized. All regressions include candidate and village fixed-effects. Regressions control for the number of relatives, number of female relatives, number of relatives in each education category and number of relatives in each occupation category. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.19: Family Networks and the Decision to Run for Office - Robustness Checks

	(1)	(2)	(3)	(4)	(5)
	Non-Parametric	Assets	Land Wealth	Colonial Status	All
Eigenvector	0.002	0.003	0.002	0.002	0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	3,882,261	3,882,261	2,908,192	1,385,804	1,304,312
R-squared	0.109	0.042	0.028	0.030	0.155

Notes: Results from family-level regressions. The dependent variable is a dummy equal to one if someone with the family name ran in the 2010 mayoral elections. Eigenvector centrality is normalized to be mean 0 and standard deviation 1. Regressions control for the number of relatives, number of female relatives, number of villages where a relative lives, number of relatives in each education category and the number of relatives in each occupation category. In Column 1, the specification includes dummies for each distinct value of each control variable. In Column 2, the regression controls for the number of relatives in each asset category. In Column 3, the regression controls for the share of municipal land that the family owns. In Column 4, the regression controls for whether a family member was mayor in the municipality at the end of the 19th century. In Column 5, the regression includes all controls from Columns 1-4. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.20: Family Networks and the Decision to Run for Office - Controlling for Land Wealth

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Eigenvector	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.002
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Share land		0.328						
		(0.062)						
Land area			0.004					
			(0.001)					
Landowner :								
Any				0.011				
				(0.001)				
Top 50%					0.015			
					(0.002)			
Top 25%						0.020		
						(0.003)		
Top 10%							0.027	
							(0.005)	
Top								0.080
								(0.028)
Obs.	2,908,192	2,908,192	2,908,192	2,908,192	2,908,192	2,908,192	2,908,192	2,908,192
R-squared	0.027	0.028	0.028	0.028	0.028	0.028	0.027	0.027

Notes: Results from family-level regressions. The dependent variable is a dummy equal to one if someone with the family name ran in the 2010 mayoral elections. Eigenvector centrality is normalized to be mean 0 and standard deviation 1. Regressions include municipal fixed-effects and control for the number of relatives, number of female relatives, number of villages where a relative lives, number of relatives in each education category and the number of relatives in each occupation category. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.21: Family Networks and the Decision to Run for Office - Controlling for Colonial Elites Status

	(1)	(2)	(3)	(4)
Eigenvector	0.002	0.002	0.002	0.002
	(0.000)	(0.000)	(0.001)	(0.000)
Spanish elite (municipal)	0.041			
	(0.013)			
Spanish elite (provincial)		0.006		
		(0.001)		
Taft commission (municipal)			0.010	
			(0.013)	
Taft commission (provincial)				0.004
				(0.001)
Observations	1,385,804	2,950,234	493,859	1,364,295
R-squared	0.030	0.034	0.034	0.029

Notes: Results from family-level regressions. The dependent variable is a dummy equal to one if someone with the family name ran in the 2010 mayoral elections. Eigenvector centrality is normalized to be mean 0 and standard deviation 1. Regressions include municipal fixed-effects and control for the number of relatives, number of female relatives, number of villages where a relative lives, number of relatives in each education category and the number of relatives in each occupation category. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.22: Family Networks and the Decision to Run for Office - Exclude Landed Elites

	(1)	(2)	(3)	(4)	(5)
Eigenvector	0.001	0.001	0.002	0.002	0.002
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Exclude :	Any Landowner	top 50%	top 25%	top 10%	top
Observations	2,887,015	2,897,532	2,902,977	2,906,232	2,908,039
R-squared	0.019	0.022	0.023	0.025	0.026

Notes: Results from family-level regressions. The dependent variable is a dummy equal to one if someone with the family name ran in the 2010 mayoral elections. Eigenvector centrality is normalized to be mean 0 and standard deviation 1. Regressions include municipal fixed-effects and control for the number of relatives, number of female relatives, number of villages where a relative lives, number of relatives in each education category and the number of relatives in each occupation category. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.23: Family Networks and the Decision to Run for Office - Interactions with Landed Elites

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Eigenvector	0.002	0.002	0.001	0.001	0.002	0.002	0.002
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Eigenvector*Land	0.012	0.001	0.003	0.004	0.003	0.003	0.004
	(0.016)	(0.000)	(0.001)	(0.001)	(0.002)	(0.002)	(0.012)
Land Measure:	Share	Land Area	Landowner	Top 50%	Top 25%	Top 10%	Top
Observations	2,908,192	2,908,192	2,908,192	2,908,192	2,908,192	2,908,192	2,908,192
R-squared	0.033	0.032	0.031	0.030	0.030	0.029	0.030

Notes: Results from family-level regressions. The dependent variable is a dummy equal to one if someone with the family name ran in the 2010 mayoral elections. Eigenvector centrality is normalized to be mean 0 and standard deviation 1. Regressions include municipal fixed-effects and control for the number of relatives, number of female relatives, number of villages where a relative lives, number of relatives in each education category and the number of relatives in each occupation category. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.24: Family Networks and the Decision to Run for Office - Excluding Colonial Elites

	(1)	(2)	(3)	(4)
Eigenvector	0.002	0.002	0.002	0.002
	(0.000)	(0.000)	(0.001)	(0.000)
Exclude :	Spanish elite		Taft commission	
	Municipal	Provincial	Municipal	Provincial
Observations	1,385,150	2,941,221	493,579	1,357,473
R-squared	0.026	0.033	0.033	0.027

Notes: Results from family-level regressions. The dependent variable is a dummy equal to one if someone with the family name ran in the 2010 mayoral elections. Eigenvector centrality is normalized to be mean 0 and standard deviation 1. Regressions include municipal fixed-effects and control for the number of relatives, number of female relatives, number of villages where a relative lives, number of relatives in each education category and the number of relatives in each occupation category. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.25: Family Networks and the Decision to Run for Office - Interaction with Colonial Elites

	(1)	(2)	(3)	(4)
Eigenvector	0.002	0.002	0.002	0.002
	(0.000)	(0.000)	(0.001)	(0.000)
Eigenvector*Elite	-0.005	0.001	-0.001	0.001
	(0.005)	(0.002)	(0.004)	(0.002)
Colonial Measure :	Spanish elite		Taft commission	
	Municipal	Provincial	Municipal	Provincial
Observations	1,385,804	2,950,234	493,859	1,364,295
R-squared	0.037	0.036	0.043	0.033

Notes: Results from family-level regressions. The dependent variable is a dummy equal to one if someone with the family name ran in the 2010 mayoral elections. Eigenvector centrality is normalized to be mean 0 and standard deviation 1. Regressions include municipal fixed-effects and control for the number of relatives, number of female relatives, number of villages where a relative lives, number of relatives in each education category and the number of relatives in each occupation category. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.26: Family Networks and the Decision to Run for Office - Interactions with Previous Electoral Experience

	(1)	(2)	(3)	(4)
Eigenvector*New	0.001	0.001	0.001	0.001
	(0.000)	(0.000)	(0.000)	(0.000)
Eigenvector*Old	0.007	0.007	0.005	0.005
	(0.001)	(0.002)	(0.002)	(0.002)
Observations	3,882,261	3,882,261	3,882,261	3,882,261
R-squared	0.157	0.158	0.172	0.173

Notes: Results from family-level regressions. The dependent variable is a dummy equal to one if someone with the family name ran in the 2010 mayoral elections. Eigenvector centrality is normalized to be mean 0 and standard deviation 1. Regressions control for the number of relatives (Columns 2-4), number of female relatives (Columns 2-4), number of villages where a relative lives (Columns 2-4), number of relatives in each education category (Columns 3-4) and the number of relatives in each occupation category (Columns 3-4). All control variables are interacted with both the old dummy and with eigenvector centrality. Municipal fixed effects are included in Column 4. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.27: Family Networks and the Decision to Run for Office - Excluding Families with Previous Electoral Experience

	(1)	(2)	(3)	(4)
Eigenvector	0.001	0.001	0.001	0.001
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	3,872,133	3,872,133	3,872,133	3,872,133
R-squared	0.003	0.004	0.006	0.007

Notes: Results from family-level regressions. The dependent variable is a dummy equal to one if someone with the family name ran in the 2010 mayoral elections. Eigenvector centrality is normalized to be mean 0 and standard deviation 1. Regressions control for the number of relatives (Columns 2-4), number of female relatives (Columns 2-4), number of villages where a relative lives (Columns 2-4), number of relatives in each education category (Columns 3-4) and the number of relatives in each occupation category (Columns 3-4). Municipal fixed effects are included in Column 4. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.28: Family Networks and the Decision to Run for Office - Network Restricted to Individuals > 45

	(1)	(2)	(3)	(4)
Panel A: OLS with over 45				
Eigenvector	0.005	0.004	0.003	0.003
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	2,086,781	2,086,781	2,086,781	2,086,781
R-squared	0.017	0.019	0.036	0.038
Panel B: IV with over 45				
Eigenvector	0.003	0.003	0.003	0.003
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	2,086,781	2,086,781	2,086,781	2,086,781
R-squared	0.038	0.038	0.038	0.038

Notes: Results from OLS (Panel A) and IV (Panel B) family-level regressions. The dependent variable is a dummy equal to one if someone with the family name ran in the 2010 mayoral elections. Eigenvector centrality is normalized to be mean 0 and standard deviation 1. In Panel B, eigenvector centrality from the network of individuals older than 45 is used as an instrument for eigenvector centrality in the full network. Regressions control for the number of relatives (Columns 2-4), number of female relatives (Columns 2-4), number of relatives in each education category (Columns 3-4) and the number of relatives in each occupation category (Columns 3-4). Municipal fixed effects are included in Column 4. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.29: Family Networks and the Decision to Run for Office - Use Percentile Rank

	(1)	(2)	(3)	(4)
Eigenvector (rank)	0.004	0.001	0.001	0.001
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	3,882,261	3,882,261	3,882,261	3,882,261
R-squared	0.002	0.015	0.032	0.033

Notes: Results from family-level regressions. The dependent variable is a dummy equal to one if someone with the family name ran in the 2010 mayoral elections. Eigenvector (rank) is the rank of the candidate's family in the distribution of eigenvector centrality in each municipality. Regressions control for the number of relatives (Columns 2-4), number of female relatives (Columns 2-4), number of villages where a relative lives (Columns 2-4), number of relatives in each education category (Columns 3-4) and the number of relatives in each occupation category (Columns 3-4). Municipal fixed effects are included in Column 4. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.30: Family Networks and the Decision to Run for Office - Removing Outliers

	(1)	(2)	(3)	(4)
		Outliers		w/o
	1%	5%	10%	ARMM
Eigenvector	0.002	0.001	0.001	0.002
	(0.000)	(0.000)	(0.000)	(0.000)
Observations	3,843,079	3,687,890	3,494,055	3,173,779
R-squared	0.014	0.007	0.005	0.029

Notes: Results from family-level regressions. The dependent variable is a dummy equal to one if someone with the family name ran in the 2010 mayoral elections. Eigenvector centrality is normalized to be mean 0 and standard deviation 1. Regressions include municipal fixed-effects and control for the number of relatives, number of female relatives, number of villages where a relative lives, number of relatives in each education category and the number of relatives in each occupation category. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.31: Family Networks and the Decision to Run for Office - Alternative Centrality Measures

	(1)	(2)	(3)	(4)	(5)	(6)
	Between	Pagerank		Katz - Decay factor:		
			.01	.11	.21	.31
Centrality	0.0039	0.0042	0.0031	0.0004	0.0002	0.0001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Observations	3,882,261	3,882,261	3,882,261	3,882,261	3,882,261	3,882,261
R-squared	0.039	0.039	0.035	0.033	0.033	0.033

Notes: Results from family-level regressions. The dependent variable is a dummy equal to one if someone with the family name ran in the 2010 mayoral elections. The network measures are normalized. Regressions include municipal fixed-effects and control for the number of relatives, number of female relatives, number of villages where a relative lives, number of relatives in each education category and the number of relatives in each occupation category. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.32: Distance (capped) to the Incumbent Mayor and Clientelistic Practices

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dep Var:		Vote Buying					Ease of Access to:			
	Overall	by Incumbent	by Challenger	Endorsement Letter	Funeral Expense	Medical Expense	Police Clearance	Barangay Clearance	Death Certificate	Business Permit
Panel A: Village Fixed Effects										
Distance	-0.040	-0.035	-0.007	-0.256	-0.151	-0.139	-0.200	-0.153	-0.158	-0.223
	(0.008)	(0.008)	(0.005)	(0.047)	(0.047)	(0.047)	(0.037)	(0.035)	(0.043)	(0.049)
Observations	3,405	3,178	3,178	3,444	3,445	3,449	3,452	3,457	3,445	3,444
R-squared	0.215	0.201	0.265	0.157	0.127	0.124	0.130	0.110	0.102	0.122
Panel A: Village Fixed Effects and Household Controls										
Distance	-0.042	-0.035	-0.008	-0.230	-0.147	-0.142	-0.180	-0.141	-0.148	-0.202
	(0.008)	(0.008)	(0.005)	(0.051)	(0.051)	(0.050)	(0.039)	(0.038)	(0.046)	(0.054)
Observations	3,073	2,861	2,861	3,105	3,106	3,110	3,113	3,118	3,106	3,106
R-squared	0.221	0.211	0.275	0.167	0.138	0.132	0.136	0.119	0.115	0.137
Mean Dep. Var.	0.397	0.240	0.161	6.775	7.002	7.431	8.546	9.204	7.530	6.179

Notes: Results from individual-level regressions. The distance variable is capped at 5. The dependent variable is a dummy equal to one if the respondent was targeted for vote buying during the 2016 elections (Column 1), a dummy equal to one if the respondent was targeted for vote buying during the 2016 elections and declared voting for the incumbent (Column 2), a dummy equal to one if the respondent was targeted for vote buying during the 2016 elections and declared voting for the challenger (Column 3). The dependent variable is a 0-10 count capturing the ease with which the respondent would be able to request the following services from their local government: Endorsement Letter from the mayor for employment (Column 4), Funeral expenses from mayor (Column 5), Medical expenses from mayor (Column 6), Municipal police clearance (Column 7), Barangay clearance (Column 8), Death Certificate (Column 9), Business permit (Column 10). In Panel B, regressions control for household size, number of children under the age of 6, number of children between the age of 7 and 14, household head's gender, household head's age, household head's education level. All regressions include village fixed effects. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.33: Distance to the Incumbent Mayor and Clientelistic Practices [Exclude Relatives]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dep Var:	Vote Buying			Ease of Access to:						
	Overall	by Incumbent	by Challenger	Endorsement Letter	Funeral Expense	Medical Expense	Police Clearance	Barangay Clearance	Death Certificate	Business Permit
Panel A: Village Fixed Effects										
Distance	-0.038	-0.031	-0.007	-0.250	-0.158	-0.145	-0.195	-0.161	-0.160	-0.225
	(0.008)	(0.008)	(0.005)	(0.048)	(0.049)	(0.048)	(0.039)	(0.040)	(0.047)	(0.049)
Observations	3,168	2,959	2,959	3,203	3,203	3,206	3,209	3,213	3,201	3,201
R-squared	0.223	0.204	0.269	0.159	0.134	0.125	0.138	0.121	0.108	0.126
Panel A: Village Fixed Effects and Household Controls										
Distance	-0.040	-0.033	-0.008	-0.228	-0.154	-0.151	-0.182	-0.154	-0.154	-0.200
	(0.009)	(0.009)	(0.005)	(0.053)	(0.053)	(0.051)	(0.041)	(0.043)	(0.049)	(0.054)
Observations	2,853	2,659	2,659	2,882	2,882	2,885	2,888	2,892	2,880	2,881
R-squared	0.231	0.217	0.280	0.168	0.143	0.133	0.142	0.129	0.119	0.139
Mean Dep. Var.	0.389	0.234	0.157	6.771	6.993	7.435	8.539	9.206	7.519	6.167

Notes: Results from individual-level regressions. The sample excludes all relatives of the incumbent. The dependent variable is a dummy equal to one if the respondent was targeted for vote buying during the 2016 elections (Column 1), a dummy equal to one if the respondent was targeted for vote buying during the 2016 elections and declared voting for the incumbent (Column 2), a dummy equal to one if the respondent was targeted for vote buying during the 2016 elections and declared voting for the challenger (Column 3). The dependent variable is a 0-10 count capturing the ease with which the respondent would be able to request the following services from their local government: Endorsement Letter from the mayor for employment (Column 4), Funeral expenses from mayor (Column 5), Medical expenses from mayor (Column 6), Municipal police clearance (Column 7), Barangay clearance (Column 8), Death Certificate (Column 9), Business permit (Column 10). In Panel B, regressions control for household size, number of children under the age of 6, number of children between the age of 7 and 14, household head's gender, household head's age, household head's education level. All regressions include village fixed effects. The standard errors (in parentheses) account for potential correlation within municipality.

Table A.34: Distance (capped) to the Incumbent Mayor and Clientelistic Practices [Exclude Relatives]

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dep Var:	Vote Buying			Ease of Access to:						
	Overall	by Incumbent	by Challenger	Endorsement Letter	Funeral Expense	Medical Expense	Police Clearance	Barangay Clearance	Death Certificate	Business Permit
Panel A: Village Fixed Effects										
Distance	-0.044	-0.036	-0.007	-0.258	-0.154	-0.136	-0.196	-0.159	-0.160	-0.219
	(0.008)	(0.008)	(0.006)	(0.053)	(0.052)	(0.052)	(0.039)	(0.037)	(0.047)	(0.053)
Observations	3,168	2,959	2,959	3,203	3,203	3,206	3,209	3,213	3,201	3,201
R-squared	0.223	0.205	0.269	0.158	0.134	0.124	0.137	0.119	0.107	0.124
Panel A: Village Fixed Effects and Household Controls										
Distance	-0.046	-0.038	-0.009	-0.231	-0.149	-0.142	-0.179	-0.151	-0.153	-0.197
	(0.009)	(0.009)	(0.006)	(0.057)	(0.057)	(0.055)	(0.041)	(0.040)	(0.051)	(0.059)
Observations	2,853	2,659	2,659	2,882	2,882	2,885	2,888	2,892	2,880	2,881
R-squared	0.232	0.218	0.280	0.167	0.143	0.132	0.141	0.127	0.118	0.138
Mean Dep. Var.	0.389	0.234	0.157	6.771	6.993	7.435	8.539	9.206	7.519	6.167

Notes: Results from individual-level regressions. The distance variable is capped at 5. The sample excludes all relatives of the incumbent. The dependent variable is a dummy equal to one if the respondent was targeted for vote buying during the 2016 elections (Column 1), a dummy equal to one if the respondent was targeted for vote buying during the 2016 elections and declared voting for the incumbent (Column 2), a dummy equal to one if the respondent was targeted for vote buying during the 2016 elections and declared voting for the challenger (Column 3). The dependent variable is a 0-10 count capturing the ease with which the respondent would be able to request the following services from their local government: Endorsement Letter from the mayor for employment (Column 4), Funeral expenses from mayor (Column 5), Medical expenses from mayor (Column 6), Municipal police clearance (Column 7), Barangay clearance (Column 8), Death Certificate (Column 9), Business permit (Column 10). In Panel B, regressions control for household size, number of children under the age of 6, number of children between the age of 7 and 14, household head's gender, household head's age, household head's education level. All regressions include village fixed effects. The standard errors (in parentheses) account for potential correlation within municipality.

A.3 Technical Appendix

A.3.1 Family Network Centrality Measures

Once the networks of intermarriages are constructed within the localities, we compute different centrality measures for all families in the locality. Our primary measure is eigenvector centrality, which we use as a specific instance of Katz or Bonacich Centrality (Katz, 1953; Bonacich, 1972, 1987).

Eigenvector, degree, Katz, and Bonacich centrality are all part of a group of network measures that essentially start with the number of intermarriage connections your family has, and factors in whether these families connected to your family are also themselves well-connected. In a sense, these types of centrality measures are akin to a popularity contest, making them one of the most intuitive ways of thinking about centrality in a network. The main differences between the various measures in this group is in how much they weight the importance of close vs. distant connections: for degree centrality, for example, only direct ties matter and second- through nth-degree ties do not contribute to centrality, while at the other extreme, Katz and Bonacich parameters can be set to consider even the most distant ties as having a contribution to centrality.

Let F denote the adjacency matrix of family network f, such that $F_{ij} = 1$ if there is a tie between nodes i and j, and 0 otherwise. The Katz centrality $Katz_i(f)$ of node i is given by:

$$Katz_i(f) = \sum_{k=1}^{\infty} \sum_{j=1}^{n} \alpha^k (F^k)_{ji} \quad (1)$$

where α is a constant corresponding to the decay factor. When the decay factor is close to 0, distant connections become less important in determining centrality, and centrality is primarily determined by close connections, converging to degree centrality when $\alpha = 0$. When the decay factor is large, distant connections are more valuable and Katz centrality is influenced by the structural features of the network as a whole. Generally, decay factors are chosen between 0 and $1/\rho(F)$, where $\rho(F)$ is the largest eigenvalue of network F.[1] We follow Banerjee et al. (2013) and choose a prominent value of α: the inverse of the first eigenvalue of the adjacency matrix. For this particular value of α, Katz centrality coincides with eigenvector centrality.

Degree Centrality

Degree centrality is the simplest measure, counting the number of ties that the politician's family has to other families. Following Wasserman and Faust (1994), we use two variants, a raw measure of the total number of connections, as well as an indexed measure that compares the total connections to the family with the highest total number of connections in the network. Since our ties represent intermarriages, they are undirected–that is, observing a tie from family A to family B implies an intermarriage between the two families, but there is no directionality: family B is just as married to family A as family A is to family B. As a result, we do not need to consider in-degree (inward) and out-degree (outward) ties.

$$Degree_i(f) = \sum F_{ij} \quad (2)$$

where F is the adjacency matrix of family network f, such that $F_{ij} = 1$ if there is a tie between nodes i and j, and 0 otherwise.

[1] Bonacich is a generalization of this measure that allows for an additional parameter, as well as negative values of alpha.

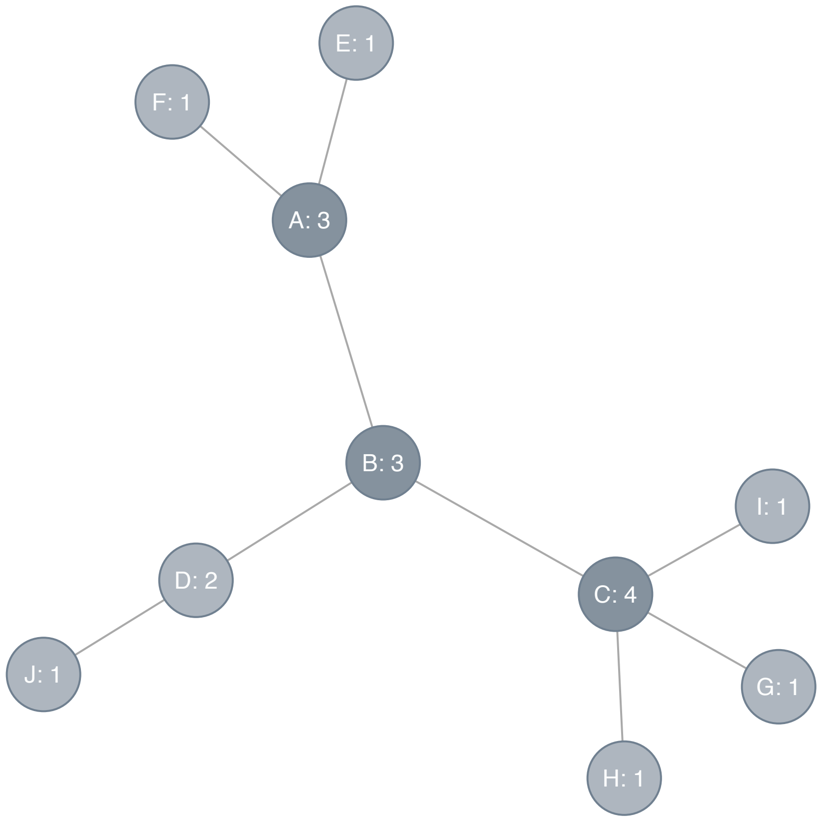


Figure A.2: Degree Centralities in a Network

Figure A.2 shows the degree centralities in a sample family network. Family *A* has a degree centrality of 3 because it has three ties through intermarriages, to families B, E, and F. Similarly, family *B* also has a degree centrality of 3 because of its intermarriage ties with families A, D, and C. The highest degree centrality belongs to family *C*, which has a degree centrality of 4, because it has intermarriage ties with four families: B, G, H, and I.

Eigenvector Centrality

Eigenvector centrality is a measure of centrality that accounts not only for the number of ties, but also whether these ties are themselves well connected (Bonacich, 1972, 1987; Jackson, 2010). Eigenvector centrality is computed recursively by calculating the prestige of a family weighted by whether the others connected to the family are themselves influential (see equation 3). Families that would be considered central using this measure are those families that have many ties to other well-positioned families. As noted above, this is one of the more intuitive measures of centrality and is often used to assess prestige and popularity.

$$Eigenvector_i(f) \propto \sum F_{ij} * Eigenvector_j(f) \tag{3}$$

where F is the adjacency matrix of graph f, such that $F_{ij} = 1$ if there is a tie between nodes i and j and 0 otherwise. This weights all of the ties to i by the connectedness of the tie (Bonacich, 1972, 1987).

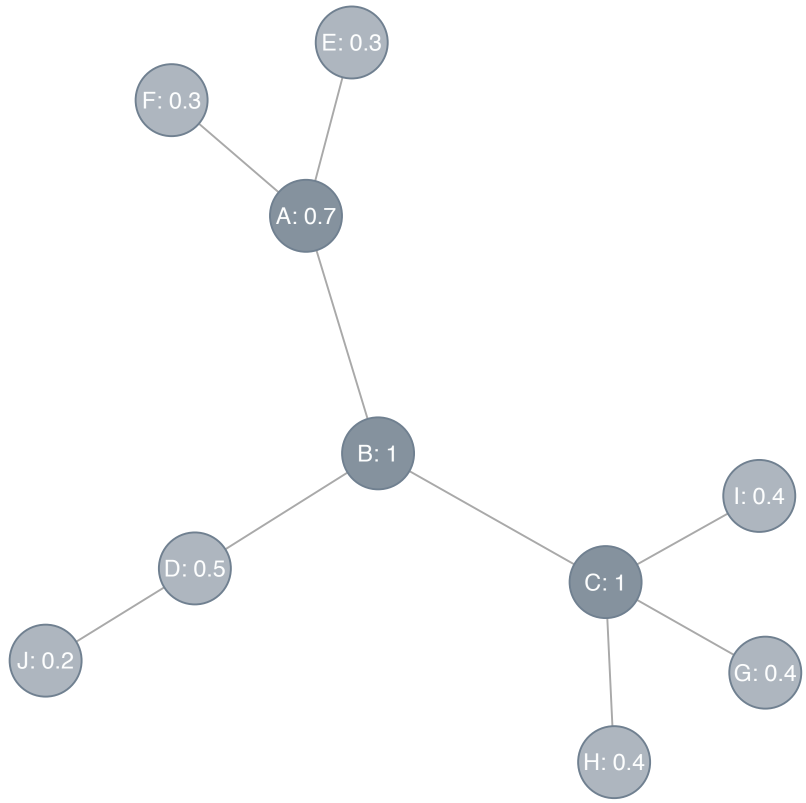


Figure A.3: Eigenvector Centralities in a Network

Figure A.3 shows the eigenvector centralities in the same sample family network. As in the eigenvector centrality measures used in the paper, this example re-scales the eigenvector centralities to have a maximum eigenvector centrality of 1. Recall that family A and family B both have degree centralities of 3 (figure A.2). However, because eigenvector centrality accounts for the not only the number of ties but whether those ties themselves are central, we can observe that family B has a higher eigenvector centrality than family A. This is because family B's ties, families A, D, and C, have eigenvector centralities of .7, .5, and 1, respectively. On the other hand, family A's ties are families B, F, and E, with eigenvector centralities of 1, .3, and .3, respectively–lower centrality than the ties of family B. Family B's eigenvector centrality is 1, while family A's eigenvector centrality is .7.[2]

Betweenness Centrality

Betweenness centrality is the extent to which the family serves as a link between different groups of families in the network. It assesses centrality by looking at whether the family is an important hub in the paths traversing the network and is calculated using the number of shortest paths in the network that necessarily pass through the family (Freeman, 1977). Betweenness for any single family is calculated in terms of its position compared to all other pairs of families (equation 4). Betweenness centrality has implications for the ability of the family to serve as a link between different groups (Padgett and Ansell, 1993).

Following the notation in Jackson (2010), in the family network f, let $P_i(kj)$ indicate the number of shortest paths between family k and family j that necessarily pass through family i, while $P(kj)$ is the

[2]Family B's eigenvector centrality is actually .97 compared to family C's eigenvector centrality of 1. The values were rounded in this example for simplicity.

total number of shortest paths between k and j.

The ratio $P_i(kj)/P(kj)$ approximates the importance of family i in connecting k and j. If $P_i(kj) = P(kj)$, yielding a ratio of 1, then family i lies on all of the shortest paths connecting families k and j. Conversely, if $P_i(kj) = 0$, then family i is not important for connecting families k and j.

Betweenness centrality is calculated by averaging this ratio across all nodes (Freeman, 1977).

$$Betweenness_i(f) = \sum \frac{P_i(kj)}{P(kj)} \tag{4}$$

In our analysis, we normalize betweenness centrality for comparability:

$$Betweenness_i(f) = \sum \frac{P_i(kj)/P(kj)}{(n-1)(n-2)/2} \tag{5}$$

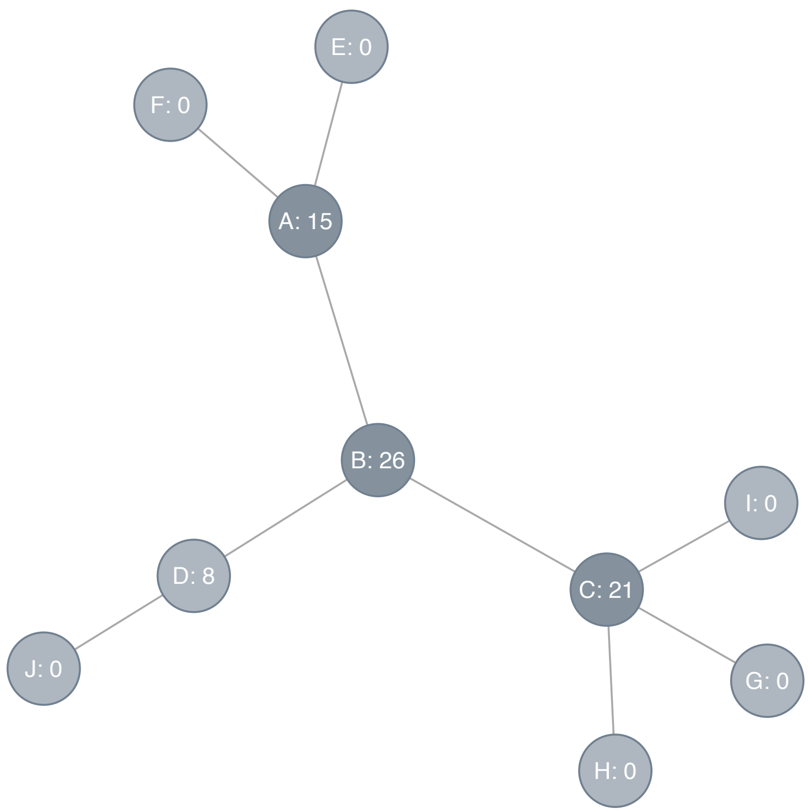


Figure A.4: Betweenness Centralities in a Network

Figure A.4 shows the betweenness centralities in the same sample family network. As indicated above, betweenness centrality is calculated first by counting the number of shortest paths through the network that necessarily pass through the family. Using family D as an example, we can see that D lies on 8 of the shortest paths through the network: all of the paths that originate from family J to all of the other nodes in the network.

As in the previous two examples, while betweenness centrality does tend to be correlated with the other centrality measures, it does not always produce the same results as eigenvector and degree. For example, from figure A.3, we know that families B and family C both have eigenvector centralities of 1, the maximum in the network. However, they have different values of betweenness centrality because of the number of shortest paths through the network that necessarily pass through them. Family *C* is on 21 shortest paths through the entire network, as it is a link from families G, H, and I to the rest of the families in the network. Family *B* has the highest betweenness centrality because it links families C, H, G, and I with the rest of the network; D and J with the rest of the network; and A, E, and F with the rest of the network. Note that family *B* does not lie on the shortest path when linking families within these clusters (i.e., family *B* is not needed to link I and H or C and G), but only when linking across clusters.

References

Banerjee, Abhijit, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson. 2013. "The diffusion of microfinance." *Science*, 341(6144).

Bonacich, Philip. 1972. "Factoring and weighting approaches to clique identification." *Journal of Mathematical Sociology*, 2: 113–120.

Bonacich, Philip. 1987. "Power and Centrality: A Family of Measures." *American Journal of Sociology*, 92(5): 1170–1182.

Freeman, L.C. 1977. "A Set of Measures of Centrality Based on Betweenness." *Sociometry*, 40: 35–41.

Jackson, Matthew O. 2010. *Social and Economic Networks. Princeton University Press*, Princeton University Press.

Katz, Leo. 1953. "A new status index derived from sociometric analysis." *Psychometrika*, 18(1): 39–43.

Padgett, John F., and Christopher K. Ansell. 1993. "Robust Action and the Rise of the Medici, 1400-1434." *American Journal of Sociology*, 98(6): 1259–1319.

Wasserman, Stanley, and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications.* Cambridge:Cambridge University Press.