Friends in High Places: Online Appendix

Lauren Cohen
Harvard Business School and NBER

Christopher J. Malloy
Harvard Business School and NBER

In this Appendix we describe in more detail the method and data cut-offs we use to classify bills into industries. We also describe additional variations in vote samples that we have explored.

A.1 Industry Classification, Keywords, and Cut-offs

As described in the data section, we first download the full text of all bills jointly from the Government Printing Office (GPO) and Congress's Thomas database. We then parse each bill's entire text, and use a list of matching words to classify each bill into the industries to which it applies. Table A1 displays the words we use to classify into the Fama-French 49 industries, for three sample industries. We are happy to provide the entire list upon request, for all 49 industries (but including them all in the appendix table made this a 13 page table). Again, the Fama-French 49 industries are somewhat analogous to the SIC 2 digit industry classification, with some improvements and aggregations of similar SIC 2 sub-industry components. As Table A1 shows, we obviously attempt to use a number of keywords to capture the bill's relevance to a given industry. However, we balance this by not choosing too many keywords to induce false positives. In the table, we include when a given industry (or keyword) was removed because it was capturing too many false positives in the industry assignment process.

To give a few examples, we remove the word "soda" from the "Candy and Soda" industry, as it kept matching with "soda ash" and "soda mountain" from a number of bills, both having nothing to do with the desired industry. As another example, for the "Personal Services Industry," we initially included the keyword "beauty shop." Unfortunately, nearly all of the instances of this keyword in bills refer to the "House Beauty Shop," referencing a (debate about) and the eventual closing of this service in one of the House of Representative buildings, and so we remove this keyword as well.

Another important aspect of this table is that after deciding upon keyword roots, we then go through each extension and conjugation that we see in the bills in order to determine which extensions and conjugations reasonably refer to the given industry. So, for instance, for the "Utilities" industry, we use the keyword root "utilit-." While this matches correctly "utility" and "utilities," it incorrectly picks up "utilize" and "utilitarian," which also appear in bills. We thus remove all of the final two matches from the bill matched sample to Utilities through "utilit-." We do this for every keyword root in every industry to ensure that the given keyword root matches to the intended industry.

The last element of the process is then choosing threshold frequencies for each keyword appearing in a given bill relative to that keyword's use across all bills, in order to classify a given bill as referring to that keyword's industry. We use two potential methods for this, the first is the absolute count of the keyword, and the second is the ratio of that word to the entire number of words in the bill. For instance, the word "electricity" has a frequency cut-off of 11 times, representing the 95th percentile of that keyword's distribution amongst bills. We have used cut-offs for both measures ranging from the 75th-95th percentile, and the results in the paper are unaffected. All results reported in the paper are for the middle of this range, 85th percentile, using the absolute number of keyword appearances.

The outcome of this process is a match of relevant industries to each bill considered in congress. We believe we have a quite conservative match process, but match fairly definitively 20% of all bills to a relevant industry (or industries).

A.2 Vote Subsamples

All results that we report in the paper's tables include all votes. However, it is important to note that we have looked at a number of other subsets of votes as well, and that the results are robust across all of these subsamples. First, looking only at final votes on bills that eventually are passed into law gives roughly identical results in terms of magnitude and significance.[1] The reason we report results for all votes in the paper is that we believe vote-trading within the social network may be going on across many types of votes.

Second, in Table VI in the paper we show a number of vote subsamples according to important economic sub-classifications within our sample, and show that no single subsample drives our results. Specifically, we split our sample out separately for every school, every Senator, and every Congress-Session. We then run our tests separately for every sub-set (for instance, in the school case, we run our tests separately for every school that appears in our sample), and report average coefficients across the subsets (along with the cross-sectional t-stat across all school estimates). Using this method drastically reduces power, but *equally weights* across each subsample we consider (in the case of Congress-Session, for example, we only have 20 estimates (20 years) that enter into the average coefficient estimate, and 20 observations entering into the standard error measurement). Thus, if our results were driven by a certain subsample, or a small set of subsamples, the equally-weighted average of the coefficients across subsamples would look much different in magnitude than the pooled regressions (and likely significance, given the then implied differences in estimates across subsamples), and so these tests would yield different implications. From Table VI, we see that our school network effects are remarkably consistent across all

[1] Note that this separation into measures is very similar to Theriault (2006), which separates procedural and non-procedural votes. Our measures category matches closely to his "Substantive Votes" category.

subsamples, using school, Senator, or Congress-Session. Again, we are happy to provide the sub-sample estimates for every Senator, every school, and every Congress-Session.

Lastly, we have used variation in "important" bills to each Senator. We considered using measures of important votes identified by Mayhew (1991) and updated on his website,[2] and also Edwards et al. (1997, 2000). These are certainly valid measures of important bills at the bill level. However, we instead opt to use a measure of "important" bills defined at the Senator-bill level. In other words, we allow the *same* bill to be important or unimportant for *different* sets of Senators. We define "importance" quite flexibly throughout the paper, using important to the given Senator's party, state (through industries domiciled there), and ideology. We then interact these measures of important bills for the given Senator (or use varying subsamples),[3] and show how our estimated impacts vary. Throughout the paper, we show that school networks do have quite a different impact across important and unimportant votes to the given Senator (and also important and unimportant votes to other Senators in the given Senator's school network). We think these are important and strong validating pieces of evidence. In sum, the subsample analyses we have done helps to pin down and strengthen the mechanism of school network influence.

[2] http://pantheon.yale.edu/~dmayhew/data3.html

[3] These two methods are obviously nearly equivalent, except that the subsample method allows all regression coefficients to vary (be freely estimated) across subsamples, whereas the interaction method (including main effects) only allows the intercept and interacted coefficient to vary. In Table IV where we use both, you can see that the two methods (not surprisingly) yield nearly identical results in our sample.

Table A1: Industry Assignment Keywords and Cut-offs

This table shows the keywords used in assigning the full text of each bill in our sample to the resultant industries covered by the bill, along with the cut-offs for the percentile in the distribution of that keyword for the entire sample. We assign the given industry to a bill if any one of its keywords is above the 85th percentile cut-off given in the table. We choose a subset of the 49 industries (Fama-French Industry Classification) that we use, as the table would otherwise be prohibitively long. We are happy to provide the entire table of keywords and cut-offs upon request.

Fama-French Industry # / Industry Name	Keyword	Count Greater Than / Equal To	Count Percentile
1 – Agriculture	agricultur-	12	85
	animal feed	7	85
	corn	4	85
	crop(s)	14	85
	farm(s)(land)	11	85
	fishing	8	85
	livestock	7	85
	wheat	8	85
26 – Defense	air force	31	85
	Ammunition	15	85
	armed force(s)	10	85
	Army	13	85
	gun(s)(runners)(powder)	8	85
	marine corps	30	85
	Military	11	85
	missile(s)	23	85
	national guard	30	85
	Navy	19	85
	Ordnance	7	85
	space vehicle(s)	3	85
	Tanks	9	85
	weapon(s)	15	85
48 – Trading	broker dealer(s)	3	85
	closed end	2	85
	commodity broker(s)	14	85
	financial services firm(s)	2	85
	investment bank(s)	8	85
	investment firm(s)	2	85
	investment management	6	85
	investment trust(s)	12	85
	mutual fund(s)	3	85
	reit(s)	44	85
	broker-dealer(s)	No Keyword Count Information Available	
	closed-end	No Keyword Count Information Available	
	security broker(s)	Keyword removed : Only 2 bills with the keyword, and all appear in definition clauses	
	unit trust(s)	No Keyword Count Information Available	

Table A2: Industry Assignments by State

This table shows the 3 most important industries for each state at the beginning, midpoint, and endpoint of our sample. “Importance” is measured by summing up the market equity of all publicly traded firms in each industry residing in a state, and then ranking industries. We thus show below the three largest industries operating in each given state over each Congress. We choose a subset of states and Congresses, as the table would otherwise be prohibitively long. We are happy to provide the entire table of states, industries operating in those states, and most important industries for each state and Congress upon request.

State	Fama-French Industry #	Industry Name	Congress
TX	30	Oil	101
TX	31	Utilities	101
TX	32	Telecom	101
TX	30	Oil	105
TX	32	Telecom	105
TX	35	Computers	105
TX	30	Oil	110
TX	31	Utilities	110
TX	32	Telecom	110
NY	45	Banks	101
NY	46	Insurance	101
NY	48	Trading	101
NY	45	Banks	105
NY	46	Insurance	105
NY	48	Trading	105
NY	45	Banks	110
NY	46	Insurance	110
NY	48	Trading	110
CA	32	Telecom	101
CA	35	Computers	101
CA	43	Retail	101
CA	35	Computers	105
CA	36	Software	105
CA	37	Electronic Equipment	105
CA	35	Computers	110
CA	36	Software	110
CA	37	Electronic Equipment	110

Table A3: Expanded Information on School Networks

This table shows an expanded version of Table II. It includes a more detailed breakdown of political affiliations within each network, along with the same information for religion affiliations within each network. Additionally, it includes a list of the states that are represented in each network, along with the most represented state within each network.

Schools Represented in the top 10 School Networks in the US Senate														
Rank	Academic institution	#degrees	%total	#senators	# Rep	# Dem	# Indep	Most Represented Religions	# in that religion	total # reported religion	# unique states	Most Represented State	# from that state	All States
1	Harvard University	35	9.33	30	12	17	1	Catholic, Methodist, Jewish	4	25	24	MI	3	AK, CO, DE, FL, HI, ID, IL, LA, MA, MD, MI, NC, NH, NJ, NM, NY, PA, RI, SD, TN, VT, WA, AI, WV
2	Yale University	23	6.13	20	8	10	2	Episcopalian, Presbyterian, Jewish	3	15	13	PA	3	CA, CT, FL, MA, MN, MO, NY, OH, OK, PA, RI, VT, WY
3	University of Virginia	10	2.67	9	4	5	0	Episcopalian	2	5	7	VA	3	FL, IN, MA, MO, RI, TX, VA
4T	Stanford University	8	2.13	8	1	7	0	Jewish	2	6	6	OR, CA	2	CA, CO, MT, ND, NM, OR
4T	Georgetown University	8	2.13	6	2	4	0	Roman Catholic	4	5	6	AK, IL, ME, VA, VT, WY	1	AK, IL, ME, VA, VT, WY
6T	Oxford University	7	1.87	7	3	4	0	Methodist	3	6	7	IN, LA, MD, NC, NJ, OK, TX	1	IN, LA, MD, NC, NJ, OK, TX
6T	Vanderbilt University	7	1.87	5	2	3	0	Baptist	2	4	1	TN	5	TN
6T	University of Chicago	7	1.87	7	2	4	1	Presbyterian, Assemblies of God, United Church of Christ, Roman Catholic	1	4	6	MO	2	IL, MN, MO, NJ, PA, VT
9T	Princeton University	6	1.6	6	3	3	0	Episcopalian, Presbyterian	2	5	5	MO	2	MD, MO, NJ, RI, TN
9T	University of Georgia	6	1.6	4	3	1	0	Methodist	2	2	2	GA	3	GA, TX
9T	University of Alabama	6	1.6	4	3	1	0	Presbyterian, Methodist	2	4	2	AL	3	AL, CA
9T	University of Mississippi	6	1.6	3	3	0	0	Southern Baptist	3	3	1	MS	3	MS
9T	University of Minnesota	6	1.6	4	2	1	1	Congregationalist, Lutheran	1	2	3	MN	2	CO, MN, ND
All Degrees		375	100											

Table A4: Top Senators by School Connected Voting

This table lists the top 20 Senators in our sample in terms of their propensity to vote with their school networks. In particular, the ranking was made using the separate regressions run for each Senator described in Table VIII (Column 4). The ranking is then based on their coefficient estimate on School Connected Votes (SCV). The top 20 Senators based on this ranking over our sample period of the 101st-110th Congresses are then listed below in alphabetical order of last name.

Name	Birthdate	Birthplace	Party	Employment	State	Undergraduate	Graduate
John B. Breaux	3/1/1944	Crowley, LA	Democrat	Attorney	LA	University of Southwestern LA	Louisiana State University (LSU)
Thomas Richard Carper	1/23/1947	Beckley, WV	Democrat	Public official	DE	Ohio state University	University of Delaware
John H. Chafee	10/22/1922	Providence, RI	Republic	Attorney	RI	Yale University	Harvard University
Larry E. Craig	7/20/1945	Council, ID	Republic	Rancher	ID	University of Idaho	
Wyche Jr. Fowler	10/6/1940	Atlanta, GA	Democrat	Attorney	GA	Davidson College	Emory University
H. John III Heinz	10/23/1938	Pittsburgh, PA	Republic	Businessperson	PA	Yale University	Harvard University
Daniel K. Inouye	9/7/1924	Honolulu, HI	Democrat	Attorney	HI	University of Hawaii	George Washington University
Johnny Isakson	12/28/1944	Atlanta, GA	Republic	Realtor	GA	University of Georgia	
James M. Jeffords	5/11/1934	Rutland, VT	Democrat	Attorney	VT	Yale University	Harvard University
Dirk Kempthorne	10/29/1951	San Diego, CA	Republic	Public Official	ID	University of Idaho	
Mary L. Landrieu	11/23/1955	Arlington, VA	Democrat	Real Estate Agent	LA	Louisiana State University (LSU)	
Harlan Mathews	1/17/1927	Walker County, AL	Democrat	Public Official	TN	Jacksonville State University	Vanderbilt University
James A. McClure	12/27/1924	Payette, ID	Republic	Attorney	ID	Idaho State University	University of Idaho
Zell Bryan Miller	2/24/1932	Young Harris, GA	Democrat	Public Official	GA	University of Georgia	University of Georgia
Earl Benjamin (Ben) Nelson	5/17/1941	McCook, NE	Democrat	Attorney	NE	University of Nebraska	University of Nebraska
Samuel A. Nunn	9/8/1938	Perry, GA	Democrat	Attorney	GA	Emory University	Emory University
David Pryor	8/29/1934	Camden, AR	Democrat	Attorney	AR	University of Arkansas	University of Arkansas
Mark Pryor	1/10/1963	Fayetteville, AR	Democrat	Attorney	AK	University of Arkansas	University of Arkansas
Arlen Specter	2/12/1930	Wichita, KS	Republic	Attorney	PA	University of Pennsylvania	Yale University
Steven D. Symms	4/23/1938	Nampa, ID	Republic	Farmer	ID	University of Idaho	

Table A5: Top Schools by School Connected Voting

This table lists the top 20 schools in our sample in terms of their propensity to have Senators vote with their school networks. In particular, the ranking was made using the separate regressions run for each school described in Table VIII (Column 3). The ranking is then based on their coefficient estimate on School Connected Votes (SCV). The top 20 schools based on this ranking over our sample period of the 101st-110th Congresses are then listed below in alphabetical order name. The second column of the table then lists whether the school was also on the top 10 list of schools most represented in the 101st-110th Congresses (shown in Table II).

University	One of the Most Represented Schools in the Senate
University of Alabama	Yes
Colorado School of Mines	
Davidson College	
University of Delaware	
Emory University	
University of Georgia	Yes
George Mason University	
Huntington College	
University of Idaho	
Idaho State University	
Jacksonville State University	
Louisiana State University	
University of Mississippi	Yes
Mount Saint Agnes College	
Oregon State University	
University of Pennsylvania	
University of Southwestern LA	
Stanford University	Yes
St. Joseph's University	
University of Washington	
Yale University	Yes

Table A6: The Impact of School Ties on Voting Behavior in the U.S. House of Representatives

This table reports panel regressions of individual votes on the voting behavior of U.S. Representatives. The dependent variable is equal to 1 if the Representative voted "Yea," and zero otherwise. In Columns 1, School Connected Votes is the percentage of Representatives from the same school as the Representative in question who voted yes on the given bill; in Column 2, School Connected Votes is the sum of Representatives from the same school as the Representative in question who voted yes on the given bill. State Votes is the percentage (in Column 1), or sum (in Column 2), of Representatives from the same state as the Representative in question who voted yes on the given bill. Party Votes is the percentage (in Column 1), or sum (in Column 2), of Representatives from the same party as the Representative in question who voted yes on the given bill. Congress fixed effects, Congress-Session-Vote (C-S-Vote) fixed effects, and Representative -fixed effects are included where indicated. All standard errors are adjusted for clustering at the Representative level, and these clustered standard errors are included in parentheses below the coefficient estimates. ***Significant at 1%; **significant at 5%; *significant at 10%.

Dependent Variable: Vote(Yes/No)		
	(1)	(2)
Votes Sample	House	House
Measure of Connections	%	Sum
School Connected Votes	0.019***	0.004***
	[0.004]	[0.001]
School Connected Votes (School and Degree)		
School Connected Votes (School, Degree, and Year)		
State Votes	0.160***	0.004***
	[0.012]	[0.000]
Party Votes	0.995***	0.005***
	[0.001]	[0.000]
Fixed Effects	C-S-Vote	C-S-Vote
Fixed Effects	Rep	Rep
Adjusted R^2	0.57	0.54
No. of Obs.	3444036	3444036