Online Appendix A: Distribution of Theses

Although we have data from 1938 through 2013, we limit our analysis to dissertations filed between the years 2000 and 2013, because 95.6% of the dissertations were filed during or after 2000. There is a clear jump in dissertations filed in 2000, suggesting a natural cut-off point. This does not represent all the dissertations filed in this span of time, but we have no reason to believe the missingness is systematic.

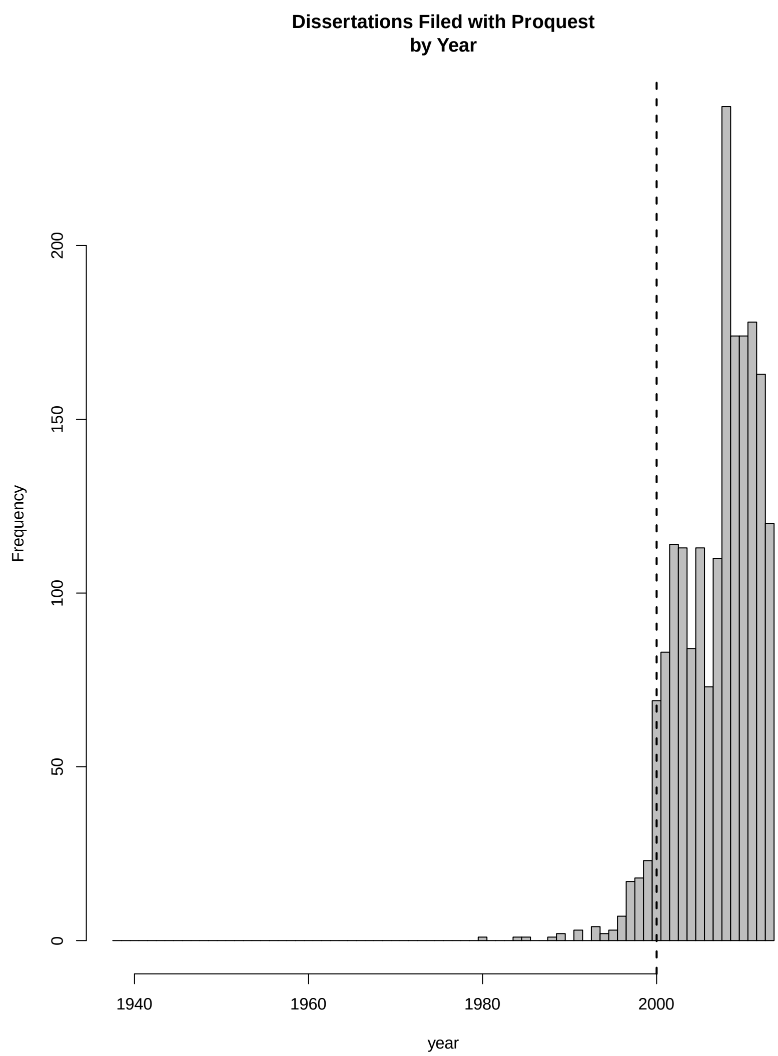

Figure A1: Number of Dissertations Per Year

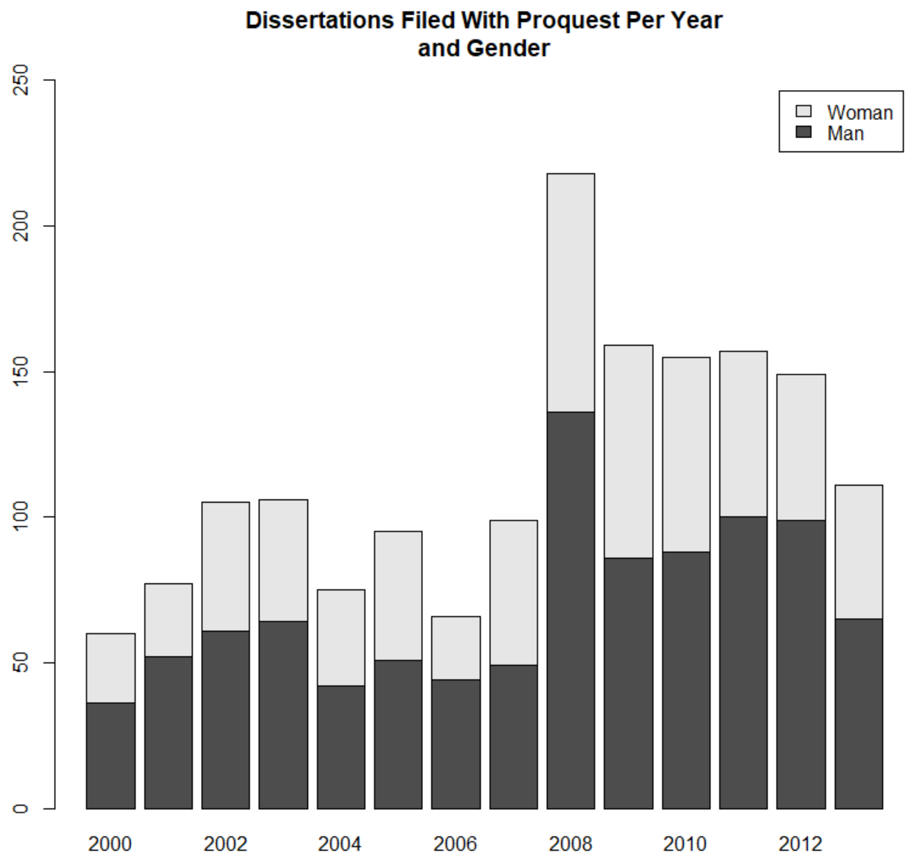

Figure A2: Number of Dissertations Per Year

Online Appendix B: Determining Author Gender

There are three reasons we may be unable to predict an author's gender. First, because the gender predictor underlying the tool uses social media data to develop its probability distributions, names that are not sufficiently common do not receive a prediction. Second, some authors have an initial for a first name and no given middle name, which makes it impossible to predict author gender without more information. Third, some names occur in roughly equal proportions among men and women. Rather than risk mis-gendering those in this third category, we assign as "woman" any name with a greater than 0.7 probability of belonging to a woman, "man" any name with a less than 0.3 probability of belonging to a woman, and NA to anyone with a probability between 0.3 and 0.7. If we instead split it at 0.5 and only fail to predict those in the first two categories, we have 2076 women, 3012 men, and 384 missing values. This represents a loss of about 3.7% of our data. For a list of names coded as NA under the stricter, 0.7 cut-off, see Table B1. The tradeoff here is between inclusion and mis-gendering, and we err on the side of not mis-gendering.

Table B1: Given Names With Pr(Woman) between 0.3 and 0.7, with Number of Observations in Parentheses

Li (6)	Jamie (5)	Jan (4)	Ji (4)	Jin (4)
Jody (4)	Deniz (3)	Hong (3)	Lei (3)	Loren (3)
Qi (3)	Robin (3)	Shiko (3)	Yan (3)	Buu (2)
Carroll (2)	Casey (2)	Cheng (2)	Chin-en (2)	Dominique (2)
Guhn-Choon (2)	Han (2)	Jae (2)	Jean (2)	Ke (2)
Kenly (2)	Ko-Yu (2)	Kuan-Shun (2)	Kyoung (2)	Luky (2)
Nien-he (2)	Page (2)	Piya (2)	Quoc-Anh (2)	Rama (2)
Ritu (2)	Riza (2)	Sang-Hyun (2)	Sascha (2)	Seiji (2)
Shao-Chee (2)	Shino (2)	Tai-Li (2)	Thuong (2)	Toni (2)
Yi-Cheng (2)	Yu (2)	Yuqing (2)	Athar (1)	Audria (1)
Bentley (1)	Bing (1)	Blair (1)	Chan (1)	Chang (1)
Chang-Chou (1)	Chao-Chi (1)	Chi (1)	Choon-Shan (1)	Dal (1)
Diarra (1)	Fang-Yi (1)	Fatos (1)	Fatu (1)	Gwynne (1)
Han-Pu (1)	Hee-jin (1)	Hong-Chi (1)	Hua-Chen (1)	Hui-Ru (1)
Hung-Hsu (1)	Hyun (1)	Jaime (1)	Jianmin (1)	Joo (1)
Jung-Hsiang (1)	Ka-Ping (1)	Kalen (1)	Kan (1)	Keegan (1)
Kennedy (1)	Kitti (1)	Kris (1)	Kusuma (1)	Maranatha (1)
Melvis (1)	Mi-Kyung (1)	Morgan (1)	Morgen (1)	Mushfiq-Us (1)
Noe (1)	Nong (1)	Patrice (1)	Rahsaan (1)	Rei (1)
Robynn (1)	Rumi (1)	Sai (1)	Seung-Whan (1)	Sheng (1)
Shiru (1)	Shun-jie (1)	Sina (1)	So (1)	Sook-Jin (1)
Sule (1)	Sultan (1)	Sung-youn (1)	Syu-ping (1)	Tal (1)
Tse-hsin (1)	Tsung-han (1)	Tsung-Sheng (1)	Tsz (1)	Uche (1)
Vu (1)	Weiwei (1)	Wen-Chin (1)	Woo (1)	Xiao-Xiong (1)
Yi (1)	Yi-Ru (1)	Yong (1)	Yoon (1)	Yoon-kyung (1)
Yuan-Chuan (1)	Yuan-Hsin (1)	Yue (1)	Yung-Ming (1)	Yuri (1)

Online Appendix C: Topic Identification

To determine the number of topics (*K*) most likely to appear in the data, we use the algorithm of Lee and Mimno (2014) built into the `stm` package (Roberts et al. 2014). As each implementation produced a slightly different number of topics, we ran the model 100 times[1] and assessed the frequency with which each number of topics was arrived upon (Figure C1).[2] The median number of topics in our simulations was 61 and the mean was 61.72. For that reason, we set our model to find $K = 61$ topics.

Some of the FREX words create easily identifiable topics. For example, we label as "international political economy" a topic identified by the word fragments "bank" , "invest", "foreign", "monetary", "trade", "financ" and "investor". Similarly, we label "vote", "voter", "choic", "turnout", "elect", "elector", and "chapter" as "voting" and "judici", "court", "judg", "complianc", "legal", "execut", and "decis" as judicial politics.

While the previous examples illustrate that many of the word groupings can be readily identified as a subfield or research topic in political science, some of them are less easily identifiable by their FREX words alone. A topic characterized by the words "china", "chines", "technolog", "water", "aid", "foreign", and "enterpris", for instance, appears, at first glance, to most likely be about Chinese politics. Upon further examination of the dissertations with the highest prevalence of this topic, however, it becomes clear that the unifying theme is actually technological capacity (whether for firms, water management, or other aims), and the topic is often (but not always) investigated in China. For this reason, we label topics based not only on their FREX words, but also based on a careful reading of the dissertation abstracts that have the highest prevalence of the given topic.

[1] For the sake of efficiency, and because the number of topics the model arrives at does not depend on it, we do not run any of them to convergence.

[2] We also used the `searchK` function, but it did not produce a conclusive answer.

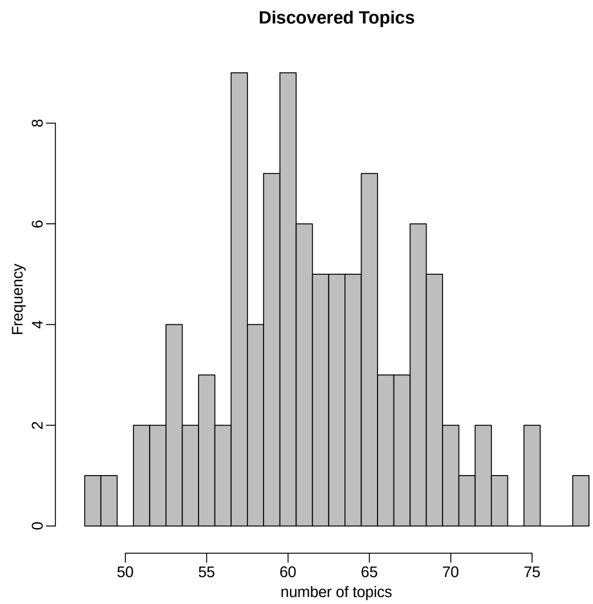

Figure C1: Number of topics discovered in 100 iterations

Table C1: Topics 1-61: A list of frequent and exclusive words and our label for them

#	Topic Label	FREX Words
1	Bargaining	polici,coalit,chang,outcom,player,decis,shift
2	Branches of Government	legislatur,execut,branch,profession,california,membership,legisl
3	Bureaucracy	agenc,bureaucrat,bureaucraci,enforc,regul,administr,regulatori
4	Campaigns	campaign,candid,consult,voter,elect,primari,presidenti
5	Canada	canada,canadian,thesi,labour,constitut,feder,german
6	Citizen Engagement	delib,particip,participatori,citizen,civic,deliber,engag
7	Citizenship/National Identity	citizenship,justic,religi,global,turkey,duti,religion
8	Civil Conflict	conflict,domest,intern,peac,territori,threat,rival
9	Congress	congress,committe,bill,legisl,congression,hous,senat
10	Corruption	corrupt,countri,variabl,quantit,cross-nat,cross-sect,regress
11	Critical Theory	critic,modern,conting,contemporari,theori,conceptu,theoret
12	Culture/Values	role,play,valu,cultur,univers,american,approach
13	Disputes	mediat,disput,cost,formal,inform,predict,incent
14	Environmental Politics	environment,program,climat,scienc,scientif,sustain,paradigm
15	Ethnic Groups	ident,latino,ethnic,identif,mobil,attitud,group
16	Federalism	plan,local,collabor,govern,feder,intergovernment,implement
17	Finance	exploit,coordin,neoliber,space,procedur,constitut,crisi
18	Fiscal Policy	tax,properti,redistribut,fiscal,revenu,cost,spend
19	Foreign Policy	militari,innov,doctrin,capabl,forc,intellig,civilian
20	Freedom	societi,democrat,freedom,civil,democraci,liberti,liber
21	Healthcare	care,health,reform,medic,law,adopt,access
22	History	norm,indigen,coloni,histor,financi,right,japanes
23	Human Development	right,incom,human,growth,inequ,econom,sanction
24	Immigration	immigr,migrant,restrict,skill,incorpor,respons,enforc
25	Interest Groups	group,interest,lobbi,activ,strategi,advocaci,organiz
26	International Political Economy	bank,invest,foreign,monetari,trade,financi,investor
27	Interstate War	terror,war,terrorist,oper,likelihood,theori,novel
28	Judicial Politics	judici,court,judg,complianc,legal,execut,decis
29	Labor	labor,market,busi,firm,economi,union,global
30	Land	land,agricultur,industri,sector,economi,rural,oil
31	Legitimacy	oblig,liber,legitimaci,action,normat,affirm,reason
32	Local/Urban politics	communiti,citi,urban,neighborhood,resid,local,civic
33	Media	media,news,prime,coverag,approv,issu,messag
34	Military and Police	offic,polic,school,interview,review,staff,field
35	Minority Participation	minor,indian,behavior,american,aggress,counti,empower
36	Narrative/Discourse	narrat,stori,young,victim,discours,histori,symbol
37	Non-Democracies/New Democracies	regim,authoritarian,opposit,leader,dictat,rule,patronag

38	Parties	parti,elector,faction,competit,system,domin,ideolog
39	Partisanship	partisan,polar,elector,elect,incumb,district,advantag
40	Personal Judgment	individu,good,judgment,peopl,qualiti,subject,experiment
41	Political Attitudes	communic,onlin,messag,internet,inform,televis,knowledg
42	Political Psychology	emot,attitud,survey,citizen,sophist,belief,psycholog
43	Political Theory	moral,thought,ethic,virtu,life,human,philosophi
44	Political Trust	institut,trust,perform,democraci,polit,across,level
45	Presidential Action	presid,frame,presidenti,agenda,issu,press,speech
46	Public Goods/Services	board,servic,provis,deliveri,educ,voluntari,contact
47	Public Opinion	public,opinion,crime,mass,fear,poll,assess
48	Race/Gender	women,gender,black,racial,white,poverti,race
49	Recruitment	environ,paper,recruit,appeal,select,cohes,behavior
50	Regional Politics	region,european,integr,latin,transport,trade,america
51	Representation	represent,constitu,style,repres,attribut,home,member
52	Repression & Violence	violenc,ethnic,insurg,civil,repress,violent,war
53	Security	secur,transnat,migrat,border,brazil,physic,boundari
54	Social Movements	movement,network,activist,organ,social,protest,mobil
55	Spending	welfar,fund,safeti,health,risk,spend,program
56	State Capacity	south,korea,africa,decentr,african,taiwan,consolid
57	State Politics	state,unit,adopt,governor,feder,variabl,determin
58	Technological Capacity	china,chines,technolog,water,aid,foreign,enterpris
59	Terrorism	attack,model,terrorist,target,threat,game,measur
60	Transitions	allianc,transit,communist,europ,post-communist,eastern,transform
61	Voting	vote,voter,choic,turnout,elect,elector,chapter

References:

Lee, Moontae and David Mimno. 2014. "Low-dimensional Embeddings for Interpretable Anchor- based Topic Inference." In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* pp. 1319–1328.

Roberts, Margaret E., Brandon M. Stewart and Dustin Tingley. 2014. "stm: R Package for Structural Topic Models." *Journal of Statistical Software*.