THREE STUDIES OF NETWORK DYNAMICS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Yongren Shi

May 2016

THREE STUDIES OF NETWORK DYNAMICS

Yongren, Ph. D.

Cornell University 2016

This dissertation examines three topics on network dynamics. The first topic is the polarization and politicization of the scientific consumption in the U.S.. Passionate disagreements about climate change, stem cell research, and evolution raise concerns that science has become a new battlefield in the culture wars. Data are derived from millions of online co-purchases as a behavioral indicator for whether a shared interest in science bridges political differences or selective attention reinforces existing divisions. Findings reveal partisan preferences both within and across scientific disciplines. The paper concludes that the political left and right share an interest in science in general, but not science in particular.

The second paper examines the dynamics of organizational membership. A long line of research documents the essential role of social networks in mediating the recruitment and retention of members in organizations. But organizations also comprise a primary context where people form social ties. The paper investigates how the network structure that an organization creates among its members influences its ability to grow and reproduce itself. In particular, we propose that two dimensions of organizational strategy influence affiliation dynamics: (1) the extent to which the organization induces social interaction among its members (social encapsulation), and (2) the time and energy that the organization demands of its members. Results show a curvilinear relationship between membership growth and the rate of social encapsulation. For most types of organizations attaining sustained growth requires a

balance between open networks (for recruitment) and network closure (for retention).

The third paper explores a structural similarity measure based on bipartite graphs. Simulation analysis holding underlying similarity constant shows that two widely used measures – Jaccard and cosine similarity – are robust to increases in network size but biased by the distribution of out-degree. However, an alternative measure, the Standardized Co-incident Ratio (SCR), is unbiased. We apply SCR to members of Congress, musical artists, and professional sports teams to show how massive co-following on Twitter can be used to map meaningful affiliations among cultural entities, even in the absence of direct connections to one another.

BIOGRAPHICAL SKETCH

Yongren Shi was born and raised in Suzhou and spent his college life in Wuhan, China. His interest diverged from mechanical engineering in which he majored in college and became passionate about the complexity science after he graduated. He finished his Master Degree in Industrial Engineering, studying manufacture systems and supply chains. He continued his journey of understanding complex social systems at Cornell, with the support of great mentors and colleagues.

Yongren's current interests revolve around creating network-based analytic strategies to understand sociological problems, ranging from opinion dynamics and polarization, organizations and ecology, and sociology of culture. He is excited about using big data to answer questions, considering himself as a computational social scientist.

To Ivy, Eric and Yanjue

ACKNOWLEDGMENTS

I am extremely grateful for the guidance of my dissertation advisor, Michael Macy. I have been fortunate to have an advisor who provides great intellectual stimulation and freedom during my years at Cornell. Without your help, this dissertation would not have been in current form. I also would like to thank the support of my committee members Douglas Heckathorn and Edward Lawler, whose generous and helpful comments greatly improves the quality of the dissertation.

My years at Cornell are also benefited from great colleagues and collaborators, from whom I learned knowledge, skills, and wisdom. Among my collaborators, I would thank Fedor Dokshin, Daniel Dellaposta, Mathew Brashears, Chan Suh, Ningzi Li, Chris Cameron, Minsu Park, Michale Genkins, Kai Mast and Rachel Behler. I am also grateful to members in Social Dynamic Laboratory, including Patrick Park, with who I share the office and converse on many topics, Tony Sirianni, Mario Molina and George Berry.

I am also grateful to Susan Meyer, Eric Giese, Alice Murdock and Marty White for their assistance and help over the years.

I am especially indebted and thoroughly grateful to my wife, Yanjue Jiang, who supports me over the years, motivates me when I am depressed, and raises two amazing children, Ivy and Eric, in the past six years. The dissertation would not be possible without the incredible, enduring support of my family.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

MILLIONS OF ONLINE BOOK CO-PURCHASES REVEAL PARTISAN DIFFERENCES IN INTEREST IN SCIENCE[1]

Abstract: Passionate disagreements about climate change, stem cell research, and evolution raise concerns that science has become a new battlefield in the culture wars. We used data derived from millions of online co-purchases as a behavioral indicator for whether a shared interest in science bridges political differences or selective attention reinforces existing divisions. Findings reveal partisan preferences both within and across scientific disciplines. Across fields, customers for liberal or "blue" political books prefer basic science (e.g., physics, astronomy, and zoology), while conservative or "red" customers prefer commercially applied science (e.g., medicine, criminology, and geophysics). Within disciplines, red books tend to be co-purchased with a narrower subset of science books on the periphery of the discipline. We conclude that the political left and right share an interest in science in general, but not science in particular.

[1] This is a paper coauthored with Feng Shi, Fedor A. Dokshin, James A. Evans, and Michael W. Macy. Yongren Shi and Feng Shi are co-first authors, contributing equally to this work.

In its quest for an objective understanding of the world (Daston and Galison 2007), modern science has practiced two distinct forms of political neutrality: as an apolitical "separate sphere" detached from ideological debates, and as a "public sphere" relevant to political issues but with balanced political engagement that facilitates reasoned deliberation and deference to evidence (Habermas 1991; Shapin 1994; Shapiro 2003; Sutton 2005). Recent surveys support the view that science contributes not only to human knowledge but also to social integration, both as a voice of reason and also as a shared value. Joint surveys conducted by the American Association for the Advancement of Science (AAAS) and the Pew Research Center in 2009 and 2014 found that science remains near the top in public rankings of professions, well above that of clergy, despite the prevalence of liberals among scientists (Kohut *et al*. 2009; Funk *et al.* 2015). Although nearly two-thirds of respondents question evolution, even those who see conflict with issues of personal faith overwhelmingly support scientific contributions to public well-being (67%). In a highly polarized electorate, these responses invite reassurance that science continues to command political deference as a voice of reason that bridges the partisan divide. We may disagree on hot button social issues but at least we can agree on science.

Political and cultural polarization within the United States, however, raises questions about the validity of this interpretation (Fiorina 2008). A less comforting possibility is that verbal survey responses may simply echo an Enlightenment commitment to value-free scientific inquiry that masks underlying skepticism about science. In recent years, conservative politicians and pundits have challenged

scientific positions on evolution, cosmology, climate change, and the perceived liberal bias in policies advocated by social scientists. For example, the conservative-funded scientific counter-movement in climate change research suggests the possibility of politically driven scientific polarization (Farrell 2015; Suhay and Druckman 2015). Survey data shows little overall change in public confidence in science since 1970, but beneath the surface there is a dramatic change in polarization: conservatives in the Vietnam era were more confident in science than liberals, but today that pattern has reversed (Gauchat 2012) (Fig. S1). Does public exposure to science play an integrative role by encouraging and informing empirical validation? Or has selective attention instead reinforced the "Big Sort" of American politics (Bishop 2009; Bakshy, Messing, and Adamic 2015)?

Much previous research has used surveys to investigate political alignments of the *producers* of science (important exceptions include Yeo *et al*. 2015; Jelveh, Kogut, and Naidu 2014). We focus instead on the *consumers* of science, using online co-purchases of books on science and politics as a behavioral indication of preferences held by customers who "vote with their pocketbook," in contrast to survey responses that are costless. Surveys measure what researchers think is important, not what respondents care about, whereas online consumers can register their preferences by purchasing books on any topic they choose. Retrospective self-reports are vulnerable to lapses of memory, while online sellers track every purchase. Survey responses are difficult to align across instruments that ask different questions and ask questions differently, whereas books from different stores can be classified using consistent

typologies (e.g., Library of Congress). Surveys are vulnerable to response bias from participants reluctant to reveal views regarded as politically incorrect. Books purchased online arrive cloaked in cardboard. Finally, while surveys can use stratified random samples to generalize results to the underlying population, which is not possible with data from a convenience sample, rates of nonresponse are rising in landline-administered surveys, which raises concerns about their external validity (Massey and Tourangeau 2013).

We addressed concerns about generalizability in two ways – by replicating our analysis using two independent samples of purchasing behavior from two online merchants (Amazon and Barnes & Noble), and also by the size of these samples, collectively comprising hundreds of millions of online customers, including members of hidden populations (e.g. those without landlines) who may be undercounted in surveys based on at most a few thousand respondents.

In sum, online co-purchase behavior among a diverse population of tens of millions of consumers provides an unprecedented opportunity to study the entire audience for science in ways that are not possible using traditional methods. These data do not speak to the partisan alignment of scientists, the policy relevance of scientific research, or the political polarization of science as an institution. Nor do we address the political preferences of the consumers of science. Rather, our attention is focused exclusively on the science preferences of those who purchase liberal and conservative political books. To what extent are purchasers of political books also interested in science, and in what parts of science are they most interested? A shared

interest in science might provide a bridge across partisan divisions, while selective attention to "convenient truths" risks reinforcement of existing political identities.

To find out, we constructed two undirected co-purchase networks of books from the American domain of the world's two largest online book stores, Amazon and Barnes & Nobel, following an approach pioneered by Valdis Krebs (Eakin 2004; Krebs 1999; 2003). Up to 100 unranked books are listed on each book page under the heading "Customers Who Bought This Item Also Bought." These are based on a collaborative filtering algorithm initially seeded by customers' co-purchase behavior (Linden, Smith, and York 2003). Recommender algorithms are closely guarded industry secrets, however, and the list of "Customers…Also Bought" books likely supplements co-purchasing with other information unrelated to customer preferences. We tested for possible bias by examining the consistency of our findings separately on two datasets from bookstores that each use proprietary algorithms. Book co-purchase mentions are not necessarily reciprocated, and directionality may be an artifact of the algorithm (e.g. the relative sales volume of each book) rather than a property of co-purchase behavior, which is inherently undirected at the level of an individual customer's multiple purchases. We therefore ignore direction and reciprocation and define an undirected *co-purchase link* between two books as the level of bi-directed co-purchasing required to trigger a co-purchase mention in either direction.

Beginning with a variety of seed books, we collected data recursively by tracing co-purchase links, iterating the search until no new titles could be identified. In total we collected 26,467,385 co-purchase links among 1,303,504 books from

amazon.com after consolidating multiple editions. (For details see SM; these counts refer to the number of titles, not purchases; we use “books” to refer to titles and never to the physical objects.) From this collection, we identified three groups: political books, science books, and non-science books. 3530 politically relevant titles were identified from three sources: Amazon’s “Conservatism and Liberalism” topic, closely related topics as indicated by Amazon cross-listings with “Conservatism and Liberalism,” and books by prominent political leaders including members of the U.S. Congress since 1993 and major party presidential candidates since 1992. Using preview text, two independent coders (with a third as tie-breaker) identified 673 conservative (“red”) books, 583 liberal (“blue”) books, and discarded 2274 neutral books. As an additional validation, we imputed red and blue codings based on the relative number of links to other red and blue books and compared these with the hand codings. Over 96% were in agreement. The network among blue and red books is visualized in Fig. 1A. The monochromatic clustering reveals the political echo chambers in a highly polarized population. (See SM for methodological details, red/blue book comparisons that show similarity in publication year and sales rank, and information about the handful of "misfits" in each cluster). We then identified all political and science books from Amazon in barnesandnoble.com, to test the consistency of co-purchasing patterns between the two bookstores. These co-purchase networks, composed of millions of distinct purchases, are different in important ways: only 9% of Amazon co-purchase links are found in the Barnes & Noble network, and only 21% of Barnes & Noble links are found in Amazon. Nevertheless, the number of political links with each book in Amazon and Barnes & Noble is highly correlated

(0.60). The consistency of our findings in these two environments suggests robust patterns of co-purchase behavior, in which the political preferences of consumers of science books are very similar even though the expression of those preferences in the purchase of particular books differs across the two web sites.

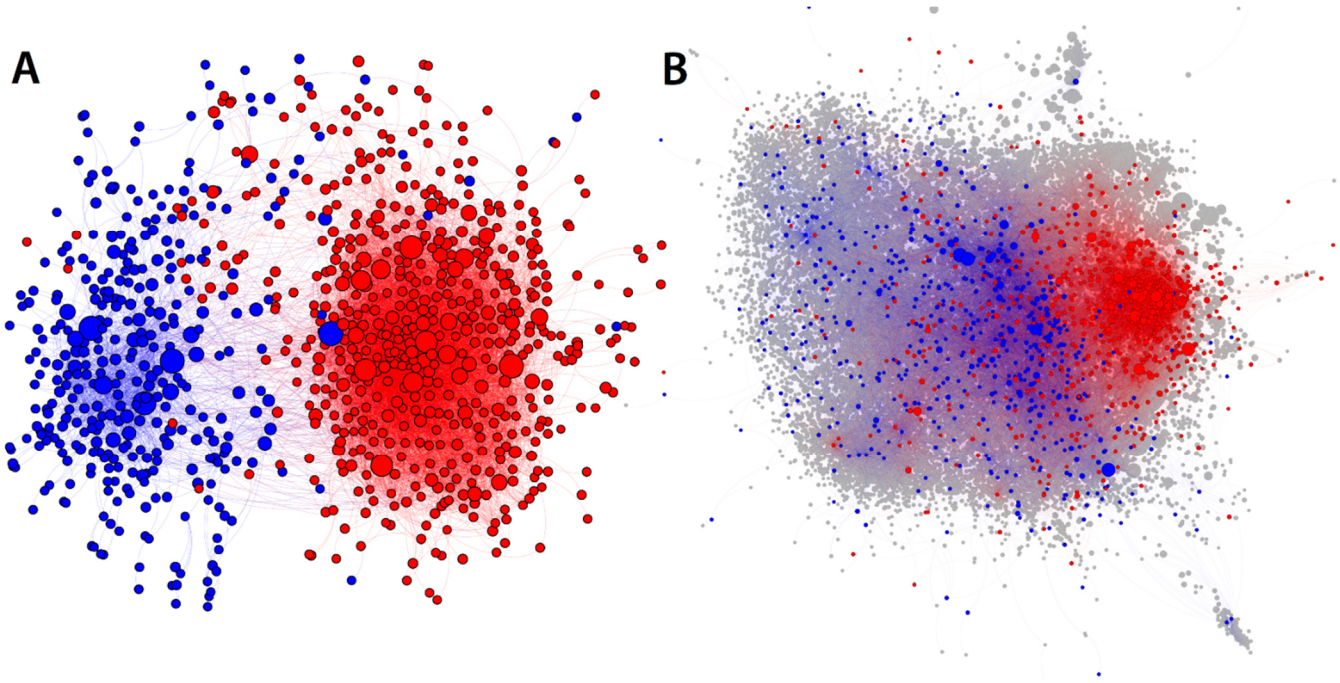


Figure 1.1. Visualization of the co-purchase network among 583 liberal (blue) and 673 conservative (red) books (A) and science (gray) books (B).
In (A), 97.2% of red books linked to other reds and 93.7% of blue books linked to other blues. Note a small number of books that were more likely to be co-purchased with books of a different color. We subjected these blinded books to additional judges and found that the original codings were nearly all correct. A number of red "orphans" were written by moderate Republicans critical of the religious right while blue "orphans" were written by progressive community organizers like Saul Alinsky (*Rules for Radicals*), later rediscovered by the Tea Party who reference their effective ethnic blue-collar organizing tactics. In (B), the broader spread of blue among science books indicates that blue books have a significantly greater scientific breadth and connect more centrally in the network of co-purchased science books.

We identified 428,433 titles that appeared under science categories in the Library of Congress (LC) and Dewey Decimal (DD) Classifications. We grouped these into 27 exclusive high-level topics, corresponding to broadly defined scientific disciplines (e.g. Physics, Chemistry, Medicine, Economics, etc.). These 27 disciplines fall under four major scientific "schools" (humanities, physical, life, and social sciences). An additional 494,278 non-science titles were grouped in four major topics – Arts, Sports, Literature (fiction and poetry), and Religion – as a baseline for assessing co-purchase links between science and politics (see SM for detailed categories).

We used co-purchase links to measure the political relevance, alignment, and polarization among online customers. *Political relevance* is the probability of a co-purchase link between political books (whether red or blue) and books about science and other topics outside politics. We estimate this probability through a Bayesian framework with a prior distribution on the probability induced by the configuration model (Molloy and Reed 1995), in which links between science and political books are randomly generated, given the network degree of each book. The higher the political relevance, the greater the likelihood that purchasers of political books will purchase a book on science compared to a randomly chosen customer. As additional context, we also measured the political relevance of books outside science (e.g. sports, fiction, religion, and performing arts).

Political alignment measures the probability that books in a particular discipline will link to red books, conditioned on their links to political books (red or

blue). Alignment is used to measure partisan interest in each discipline on the red-blue spectrum, where purple (alignment=0.5) could indicate the balanced political interest required for a public sphere of reasoned discourse. Alternatively, a "purple" discipline could also be internally divided, with equal interest from left and right but in separate subsets of books.

Political polarization identifies this latter possibility, as a function of the number of books within a discipline linked to both red and blue books, compared with a null model in which red and blue links are randomly assigned to books in the topic. Polarization equals zero when red and blue books are co-purchased with disciplinary books uniformly at random, but increases as the sets of red- and blue- linked books diverge, indicating red and blue preferences for distinct books. (See SM for formal mathematical descriptions of the measures).

In addition to these three measures, we also measured characteristics of scientific books and fields to account for diverging red and blue scientific interest. We scored fields as basic or applied science by tabulating the number of times journals within a discipline have been cited by the US patent database from Google (1976-2014), normalized by the number of journal citations to control for discipline size and activity. We also consider whether books are "academic" or "popular science," based on publication by an academic or popular press. Finally, we measured the *scientific breadth* of political co-purchase links and the *network location* of books with ties to red and blue relative to the core or periphery of the co-purchase network within the discipline. *Scientific breadth* is the proportion of disciplinary titles that link to a red or

blue book. For example, if red books link to a narrow subset of books within a discipline, while an equal number of blue books connect to a large and diverse subset of disciplinary books, those purchasing blue books have exposure to a wider range of science books—and likely a wider range of scientific perspectives—than those purchasing red books. We measure *core location* of red and blue with respect to a discipline as the closeness centralities of red- and blue-linked books within that discipline (Sabidussi 1966). For example, if red-linked books have low centrality, that indicates that conservative interest tends to be focused on books in a discipline that are less likely to be co-purchased by any random customer with other books in the discipline. (See SM for details on the measures.)

Our analysis proceeds in three steps: We first assess relevance, alignment, and polarization as measures of political interest in science compared to political interest in books outside science; second, we report differences in these measures across scientific disciplines, broken down by the relative importance of applied and basic science; third, we report the scientific breadth and location of red- and blue-linked books in the core or periphery within disciplines.

First, compared to the number of purchases expected by chance, liberal and conservative customers are more likely to buy books on science than on major topics outside science, but this interest in science does not appear to be insulated from the "Big Sort" of American politics. Figure 2 reports political relevance and polarization of science, religion, sports, arts, and literature, with alignment indicated by color. Results for political relevance show that political readers have greater interest in

science relative to non-science topics, due largely to books on social science. The physical and life sciences do not attract markedly greater interest among political readers compared to topics outside science, but this interest is significantly more polarized than for Arts and Sports, indicating that liberals and conservatives are less likely to read the same book.

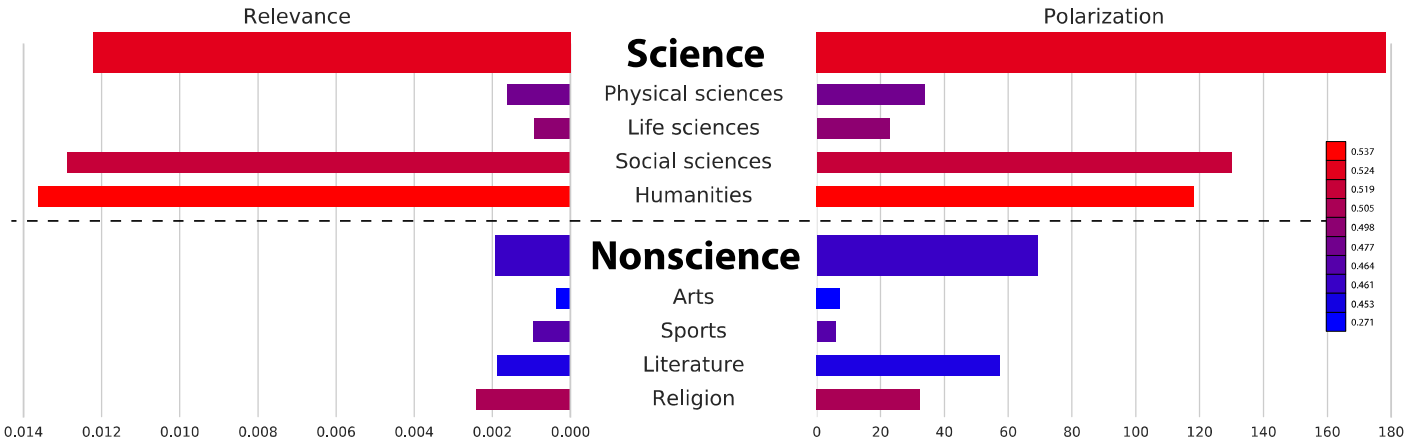


Figure 1.2. Comparisons of political relevance, polarization, and alignment between science and non-science books.
Color indicates alignment from conservative (red) to liberal (blue). Science books are more politically relevant and polarized than non-science, due largely to the social sciences and humanities, while the physical and life sciences are similar to non-science overall. Books on the performing arts and sports have low political relevance and polarization compared to literature, religion or science.

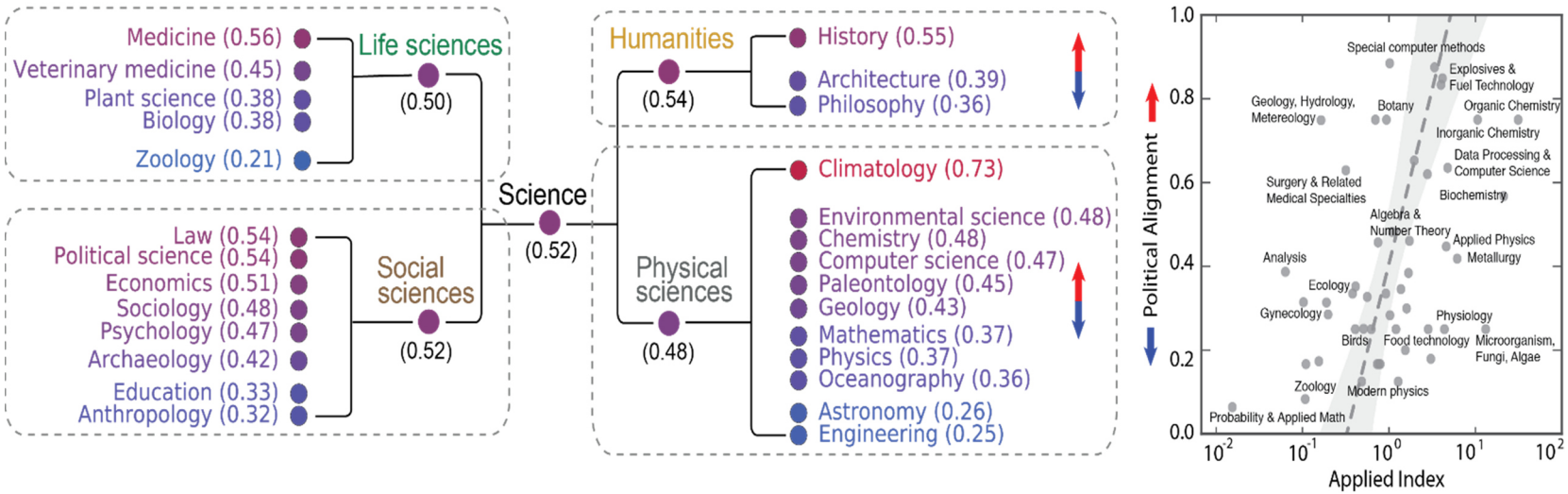

Medicine (0.56)
Veterinary medicine (0.45)
Plant science (0.38)
Biology (0.38)
Zoology (0.21)
Life sciences
(0.50)
Law (0.54)
Political science (0.54)
Economics (0.51)
Sociology (0.48)
Psychology (0.47)
Archaeology (0.42)
Education (0.33)
Anthropology (0.32)
Social sciences
(0.52)
Science
(0.52)
Humanities
(0.54)
History (0.55)
Architecture (0.39)
Philosophy (0·36)
Physical sciences
(0.48)
Climatology (0.73)
Environmental science (0.48)
Chemistry (0.48)
Computer science (0.47)
Paleontology (0.45)
Geology (0.43)
Mathematics (0.37)
Physics (0.37)
Oceanography (0.36)
Astronomy (0.26)
Engineering (0.25)
Political Alignment
Applied Index
1.0
0.8
0.6
0.4
0.2
0.0
10^{-2}
10^{-1}
10^{0}
10^{1}
10^{2}
Special computer methods
Explosives & Fuel Technology
Geology, Hydrology, Metereology
Botany
Organic Chemistry
Inorganic Chemistry
Data Processing & Computer Science
Surgery & Related Medical Specialties
Biochemistry
Algebra & Number Theory
Applied Physics
Metallurgy
Analysis
Ecology
Gynecology
Physiology
Birds
Food technology
Microorganism, Fungi, Algae
Zoology
Modern physics
Probability & Applied Math

Figure 1.3. Comparisons of political alignment across scientific disciplines and sub-disciplines.
Differences in alignment emerge more clearly the more disaggregated the categories. In the disciplinary tree, applied disciplines like medicine, law, and climatology are relatively conservative, while basic science disciplines like zoology, anthropology and philosophy are relatively liberal. This association becomes clearer at the sub-discipline level, revealed in the scatterplot. Conservative alignment tends to increase with the applied index, which measures the ratio of patent to article citations, here shown on disaggregated Dewey Decimal subfield data along with a 90% confidence interval (r=0.43, p=0.002).

Second, we found a significant positive correlation (r=0.43, p=0.002) between political alignment and the relative interest in applied and basic science (Fig. 3B). For example, organic chemistry is the most applied sub-discipline, as measured by patent vs. article citations, and it aligns closely with red books (0.75 on an alignment scale from 0 to 1). This contrasts with a sub-discipline like zoology, which is largely driven by curiosity and basic scientific concerns, and appeals more to those on the left (0.1). This pattern can also be observed in Fig. 3A, which reports co-purchase alignment within the four "schools." Applied disciplines like medicine and law attract readers at the red end compared to other disciplines in their respective schools, while anthropology and mathematics attract readers at the blue end. This mirrors the ideological differences among scientists employed in academy versus industry, as reported in AAAS/Pew surveys from 2009 and 2014 (Kohut *et al.* 2009; Funk *et al.* 2015). A possible interpretation is that scientific puzzles appeal more to the left, while problem-solving appeals more to the right.

A few disciplines, notably paleontology, bridge political divisions and are politically purple because books in these disciplines attract equal interest from both

left and right, which suggests that, whatever our political disagreements, we can at least agree about dinosaurs. However, most purple disciplines (with equal likelihood to have links with red and blue books) do not bridge political divisions. Figure 4 illustrates the internal network structure of seven natural and social science disciplines. Monochromatic clusters, located in different regions of the network, indicate that red and blue political books are linked to different inter-connected clusters of disciplinary books. Simply put, even when the left and right are equally likely to read books in a discipline, they are rarely the same books or even the same topical cluster.

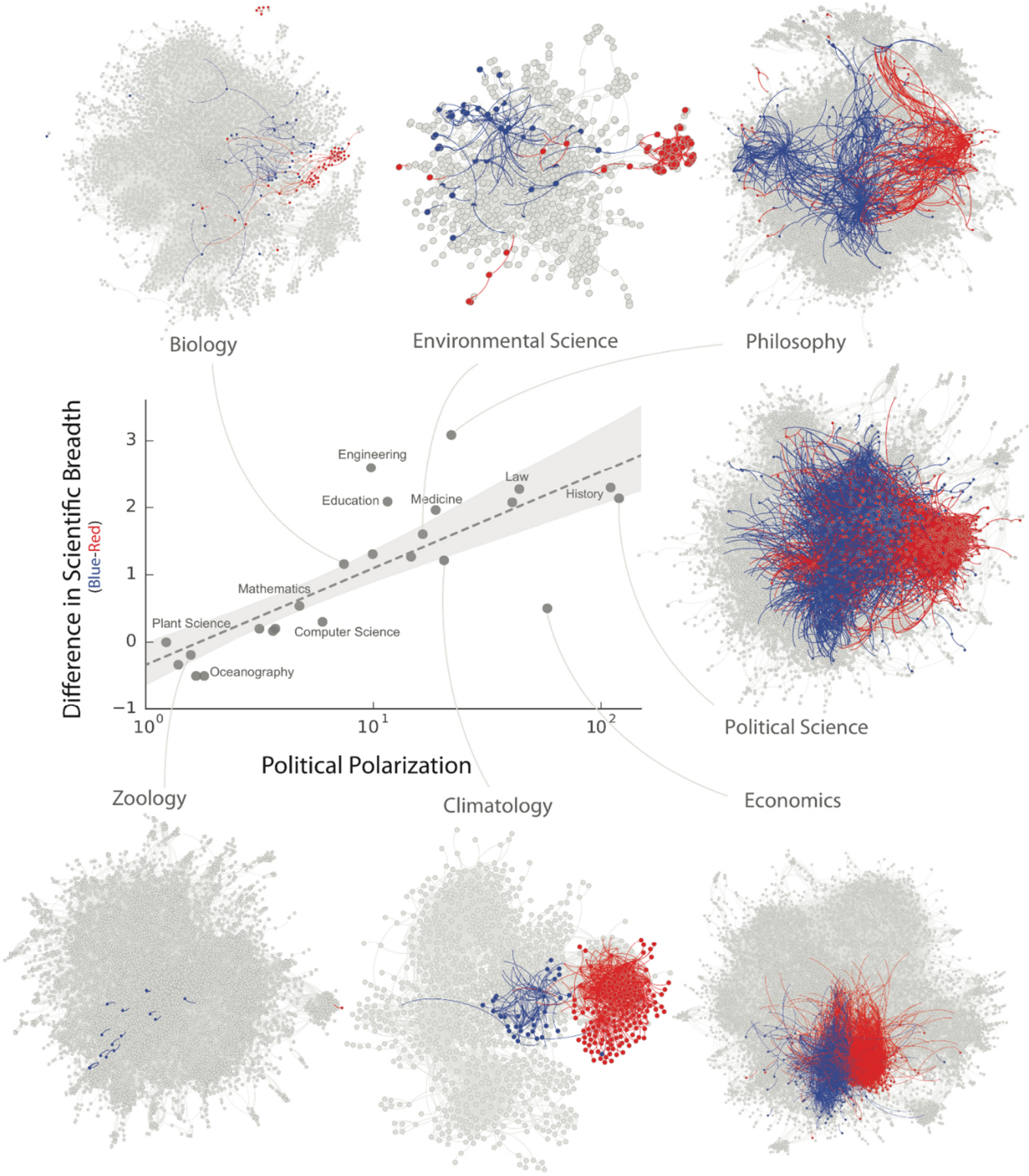


Figure 1.4. Location and polarization of red-linked and blue-linked science books within scientific disciplines.

The network visualizations show the co-purchase networks of books (grey nodes) in seven disciplines, zoology, biology, environmental science, climatology, economics, political science and philosophy, and red and blue political books linked to those disciplines. All disciplines are polarized, with red books linking to fewer disciplinary books, which tend to be

further on the periphery than blue books (see SM). These disciplines are located within the scatterplot where each discipline's polarization is plotted against the difference between the scientific breadths of blue and red books in each discipline. The line represents the estimated relationship, along with its 90% confidence interval (r=0.8, p<0.001). This reveals that the more polarized a discipline, the more likely red books will be co-purchased with fewer science books, which also tend to be on the periphery of that discipline's co-purchase network (see Fig. S4.)

We drilled down further to examine the location of red- and blue-linked books within co-purchase networks at the disciplinary level. Fig. 1.4 shows that blue books link to a larger proportion of scientific titles on average, as indicated by the association between co-purchase polarization and scientific breadth. In disciplines like climatology, environmental science, political science, and medicine, red-linked books tend to cluster on the periphery of the disciplinary network, with blue-linked books closer to the core, i.e. liberals tend to purchase science books that are more likely to be co-purchased by random customers with other books in the discipline, while conservatives tend to purchase books that are likely to be co-purchased (by the average customer) with each other but not with other books in the discipline. The greater centrality of blue-linked books does not appear to be a consequence of academic liberalism. When books by academic publishers are removed, the pattern remains, as do all the other patterns we report (see Fig. S1.5).

These results need to be qualified by inherent limitations in the use of co-purchase links to measure partisan interest in science. First, although half the U.S. population purchases books online, this is not a random sample, which limits the ability to generalize our results to the other half. Second, we do not have individual-level co-purchase data and our imputations from an unknown collaborative filtering algorithm may be biased in ways we cannot measure. Third, co-purchase links can influence purchases that then reinforce new co-purchase links, thereby magnifying estimates of the underlying consumer interest in politics and science. Finally, co-

purchasing patterns can help reveal the distribution of political interests in science but not the underlying causes.

These concerns are mitigated, however, by similarity in the results derived from distinct bookstores, with different purchase patterns and company-specific proprietary algorithms. We replicated the Amazon-based analyses using the Barnes & Noble network and found consistent results across datasets, with political polarization (r=0.97, p<0.001) and alignment (r=0.76, p<0.001) highly correlated across the two websites.

Keeping the limitations of these data in mind, online co-purchases suggest that overall interest in science is equally strong among liberal and conservative readers, but these interests are divergent. We found little support for the view that science is either an apolitical separate sphere that is largely ignored by partisans or a public sphere in which left and right share common scientific interests. Books on science are more likely to be co-purchased with political books compared to novels or books on religion, sports, or the arts, but left and right rarely purchase the same books. Science is polarized at the aggregate level, with liberals attracted to basic science and conservatives attracted to applied, commercial science. Books in a few "purple" disciplines like paleontology are co-purchased by both the left and right, but these "bridging" disciplines also tend to be those with the lowest relevance to political readers. Disciplines with high political relevance, like social science, law, history, and biology, tend to attract politically aligned readers, and even when they attract readers from both left and right, it is not to the same books. Science may not be on the front

lines of the culture wars, but it is not above the battle, nor is it immune to the "echo chambers" that have been widely observed in political discourse. This selective exposure to "convenient truth" limits the capacity for science to inform political debate or temper partisan passions.

APPENDIX

Materials and Methods

Data Collection

We collected metadata for 1,449,525 books from the largest online book retailer, Amazon.com, in spring, 2013. Starting from two seed books, one liberal (Obama's *Dreams from My Father*) and one conservative (Romney's *No Apology*), we scraped all information accessible from the webpage of each book, including descriptive information, reviews, and a list of books that are bought by customers who also bought this book under the "Customers Who Bought This Item Also Bought" section, and then followed the co-purchase mentions to move to other books iteratively. In the end, our crawler obtained the largest strongly connected component in Amazon's directed co-purchase network.

Because every title may have multiple editions or formats, e.g. paperback and hardcover, each of which is associated with a distinct ISBN (International Standard Book Number), we consolidated different editions and formats based on the unique ASIN (Amazon Standard Identification Number) provided in the source code of each book page. After consolidation we ended up with 1,303,504 unique titles in the dataset. In the main text and below we use "books" to refer to the distinct titles and never to the physical objects.

In addition, we collected the science and political books identified in Amazon from Barnes & Noble, in addition to all books from the Barnes & Noble co-purchase mention lists for those books. See below for how science and political books are identified. A brief summary of the two datasets is given in Table S1.1.

Table S1.1. Number of books and co-purchase links in Amazon and Barnes & Noble datasets.

The last row reports the number of books and links present in both datasets. The last column reports the number of co-purchase links in the subgraph of science and political books. Books in Barnes & Noble are a subset of Amazon books, but the two co-purchase networks are quite different when considering the co-purchase links.

	Number of books			*Number of co-purchase links*	
	Total	Science	Politics	Total	Science - Politics
Amazon	1,303,504	428,433	1,256	26,467,385	6,288,423
B & N	439,603	285,942	1,078	3,375,406	2,582,715
Common	439,603	285,942	1,078	664,149	542,578

Political Books

Candidate political books were chosen from three sources. The first source is the "Liberalism & Conservatism" category from Amazon, including 1677 books. The second source of political books is the top 20 Amazon categories that share books with "Liberalism & Conservatism." We identified 1812 books that only belong to these 20 categories. The third source is books written by prominent U.S. political leaders including members of the U.S. Congress since 1993 and major party presidential candidates since 1992, representing 320 books. In total, we prepared 3714 distinct candidate political books for coding.

Every book was read by at least two independent coders, with another independent coder as tiebreaker. The coders' task was to determine whether a book expresses a liberal or conservative ideology or is ideologically indeterminate, based on the information available on the Amazon webpage for that book. To code a book as liberal or conservative, therefore, a book must meet two basic requirements:

1. The book must have political content. That is, the book must express an ideological position on a mainstream political or social issue. This does not mean that the issue has to be the main topic of the book. For example, autobiographies are nominally about a person, but autobiographies of political figures almost always express ideological positions on various political issues. In this case, they should be considered to have political content.
2. The ideology that the book espouses must be consistently liberal or conservative. If central topics of the book appear to express contradictory ideological positions, the book's ideology should be coded as indeterminate.

Two independent coders were hired to code books from the first source and reached agreement on 83.5% of the candidate books. Another two independent coders were hired to code books from the second and the third sources and achieved agreement on 70% of the books despite large diversity among the books. Conflicts were resolved by other independent coders and the authors. In total, we identified 677 conservative ("red") books, 587 liberal ("blue") books, and 2545 indeterminate books, based on Amazon preview text.

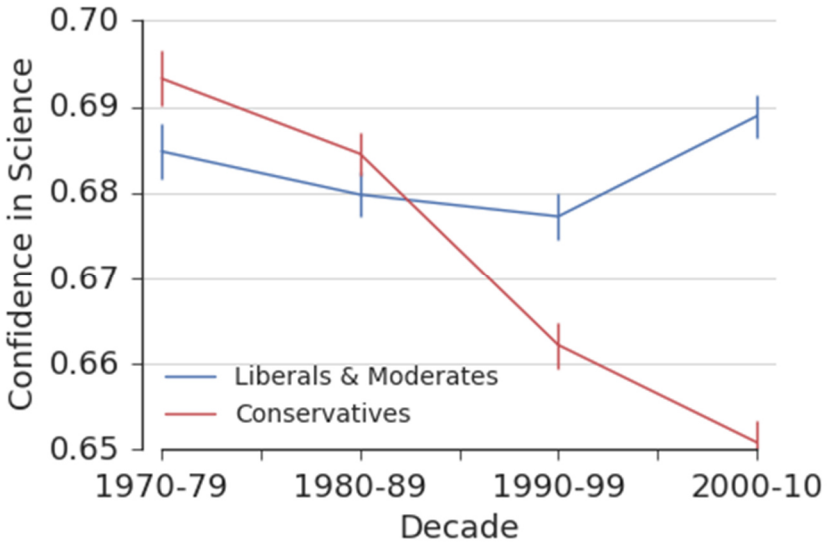


Figure S1.1. Confidence in science among liberals & moderates (blue curve) and conservatives (red curve) since 1970.

Data are from the General Social Survey cumulative file (N=33154). The figure reports decade-specific averages of the responses in each of the 27 years between 1972 and 2010 that the item was surveyed. Prior to the 1990s, conservatives reported higher confidence, although the differences are not statistically significant. After 1990, conservatives report significantly lower confidence in science.

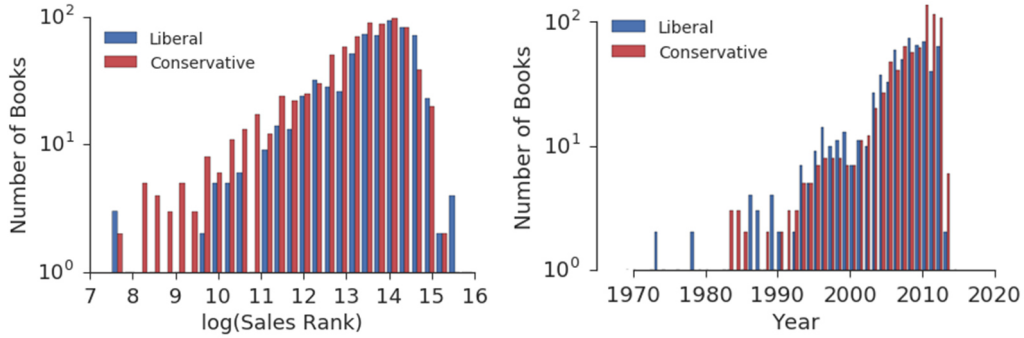


Figure S1.2. Distributions of sales rank and publication year by liberal and conservative books.

Left: Distributions of logarithmic sales ranks from liberal books and conservative books. Rank 1 corresponds to highest sales. Books with missing sales rank are not considered in the plot. The mean logarithmic sales rank is 13.5 for liberal books and 13.1 for conservative books, and median logarithmic sales rank is 13.7 for liberal and 13.4 for conservative. Right: Distributions of publication years for liberal books and conservative books. Mean publication

years of both are 1999. Median publication year is 2007 for liberal books and 2009 for conservative books.

Distributions of sales ranks for liberal and conservative books are shown in Fig. S1.2, together, along with distributions of publication years for the two sets of books. The average logarithmic sales ranks for liberal and conservative books are 13.5 and 13.1, respectively, and the medians are 13.7 and 13.4. The average publication years of both sets of books are 1999, and the median publication year is 2007 for liberal books and 2009 for conservative books. A summary is provided in Table S1.2.

Table S1.2. Sales rank and publication year for political books.

	Number of books	*Logarithmic sales rank*		*Publication year*	
		Mean	Median	Mean	Median
Conservative	677	13.1	13.4	1999	2009
Liberal	587	13.5	13.7	1999	2007

Science Categories

We took the science-related categories in the Library of Congress (LC) and Dewey Decimal (DD) Classification Systems, and reorganized them into a hierarchy of science categories. In summary, we grouped these categories into 27 exclusive high-level topics, corresponding to broadly defined scientific disciplines (e.g. Physics, Chemistry, Medicine, Economics, etc.). These 27 disciplines fall under 4 major scientific "schools" (humanities, physical, life, and social sciences). We then used the LC and DD codes of the books to sort them into disciplines.

An additional 494,278 non-science titles were grouped in four major topics – Arts, Sports, Literature (fiction and poetry), and Religion – as a baseline for assessing co-purchase links between science and politics.

Political Relevance

Political relevance measures how likely books from a given topic will be co-purchased with political books. It is a number between 0 and 1, defined as the probability θ that a co-purchase link from books in a given topic will link that topic with political books (red or blue). More formally, we assume that the number of undirected co-purchase links

X between the topic (e.g. climatology) and political books has a binomial distribution, $X \sim Binomial(K, \theta)$, where K is the total number of links attached to the topic.

A straightforward estimator of this probability θ is the number of co-purchase links between the topic and political books divided by the total number of co-purchase links between the topic and all other topics, X/K. However, this estimator is not appropriate in this application for two reasons. First, a topic might have few links to other topics, which renders this measure of relevance unreliable. For example, if a topic has only one link to other topics and it links to political books, then we have $\theta = X/K = 1$, implying that the topic is extremely relevant, which is dubious since there is only one co-purchase link in total. Although the uncertainty of this estimator due to this small-size effect could be captured by its variance or confidence interval, we need to compare relevance across topics while still taking such uncertainty into consideration, which is not straightforward using this estimator.

Therefore, in order to take into consideration the uncertainty imposed by a small-sample effect and to provide a normalized score across topics, we developed a Bayesian estimate for the probability θ. First, assume a null model in which all the co-purchase links between topics are randomly shuffled while preserving the number of links attached to each topic (cf. configuration model (17)). This "best guess" of the co-purchase network (without knowing the identities of the topics) represents a prior distribution on θ, $\theta \sim Beta(d \cdot k_{political}/m, d(1-k_{political}/m)$, where $k_{political}$ is the total number of links attached to political books and m is the total number of links between topics (including political books). $k_{political}/m$ is the probability of linking to political books in our random null model, and d controls the strength of our prior. The values of θ will depend on d, but we are always interested in the relative size of θ across topics, and thus our results are insensitive to the choice of d. We currently use the average number of links to political books over all disciplines as *d*. After observing the number of links between the topic and political books, we can update our knowledge on θ using Bayes rule:

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}.$$

We can then derive a Bayesian estimate for the probability θ:

$$\theta \sim Beta(d \cdot k_{political}/m, d(1-k_{political}/m),$$

$$X|\theta \sim Binomial(K, \theta),$$

$$\theta | X \sim Beta(X + d \cdot k_{political} / m, K - X + d(1 - k_{political} / m),$$

where the posterior distribution $\theta | X$ combines our initial guess of θ (based on the randomized network) with the actual data X and weights the actual data by the number of co-purchase links attached to the topic.

Finally, the political relevance of the topic is defined as the mean of the posterior distribution of θ:

$$E[\theta | X] = \frac{X + d \cdot k_{political} / m}{K + d}.$$

This model further reflects how the actual observations (X / K) are combined with prior beliefs ($k_{political} / m$) to incorporate uncertainty due to small-sample effects on the point estimator of θ. If the number K of co-purchase links attached to a topic is small, we do not have enough evidence to estimate its relevance and its political relevance will be close to the level in the random null model; if a topic has many co-purchase links, we have greater trust in the estimate.

Political Alignment

Political alignment locates each discipline on the blue-red spectrum. Analogous to political relevance, alignment is defined as the probability that a co-purchase link attached to a book in a given topic is to a red book, conditioned on the link being to a political book (red or blue). Hence, it can be viewed as a measure of how relevant the topic is to red books as opposed to blue books. Note that we restrict our focus to links with political books rather than to all topics. For ease of notation, we also denote this probability θ. One way to estimate this probability θ is to divide the number of links (X_{red}) between the topic and red books by the total number of links (K_p) between this topic and political books, X_{red} / K_p. Similar issues arise as for political relevance, however, and hence we developed an analogous Bayesian model to estimate θ.

First, we assume a null model in which co-purchase links between a given topic and political books are randomly shuffled while preserving the number of links attached to each (cf. configuration model (17)). Let k_{red} and k_{blue} be the total numbers of links attached to red and blue books, respectively. Then the probability of linking to red books in the random null model is $k_{red} / (k_{red} + k_{blue})$, which suggests a prior distribution on θ

,

$$\theta \sim Beta(d \cdot k_{red} / (k_{red} + k_{blue}), d \cdot k_{blue} / (k_{red} + k_{blue})),$$

where d controls the strength of our prior assumption. After observing the number of links (X_{red}) between the topic and red books, and the number of links (X_{blue}) between the topic and blue books, we update our knowledge on θ and obtain the posterior distribution of θ. In summary, the Bayesian model for estimating θ is as follows:

$$\theta \sim Beta(d \cdot k_{red} / (k_{red} + k_{blue}), d \cdot k_{blue} / (k_{red} + k_{blue}))$$

$$X_{red} | \theta \sim Binomial(K_p, \theta)$$

$$\theta | X_{red} \sim Beta(X_{red} + d \cdot k_{red} / (k_{red} + k_{blue}), K_p - X_{red} + d \cdot k_{blue} / (k_{red} + k_{blue}))$$

.

Accordingly, political alignment is calculated as the mean of the posterior distribution:

$$E[\theta | X_{red}] = \frac{X_{red} + d \cdot k_{red} / (k_{red} + k_{blue})}{K_p + d}$$

.

We note one extra step in presenting the alignment score. Naively, one might expect a topic to be right leaning if the topic has a larger probability to link with red books than blue books (i.e., $\theta > 0.5$), but this is not a fair comparison. Because red books have more total co-purchase links than blue books, red books would also possess more co-purchase links in the random null network. Therefore, the fair comparison is to compare θ with $k_{red} / (k_{red} + k_{blue})$, which is the probability of linking to red books in the random null model. To make the presentation of results more accessible and intuitive, we scaled θ linearly so that when $\theta = k_{red} / (k_{red} + k_{blue})$ the rescaled θ would be 0.5. In the main text and in the supplementary materials, θ is always rescaled and 0.5 is the "neutral" point, which accords with intuition. The rescaling is straightforward: θ is rescaled to $0.5 \times \theta / [k_{red} / (k_{red} + k_{blue})]$ when $\theta < k_{red} / (k_{red} + k_{blue})$, and to $0.5 \times [\theta - k_{red} / (k_{red} + k_{blue})] / [1 - k_{red} / (k_{red} + k_{blue})] + 0.5$ otherwise.

Lastly we checked the consistency of our measurements of political alignment by measuring the alignment of all political books and comparing the results with the "ground truth" red and blue hand codings. Specifically, for every political book that has been coded by our coders as blue or red, we pretended that its ideology is unknown and computed its alignment using the same procedure for measuring alignment of scientific topics introduced above. We then classified a book with alignment $\theta > 0.5$ to be red and

otherwise blue. The imputed ideology agrees with human codings for over 96% of the political books. Inspection of the handful of anomalies reveals red books that would be expected to appeal to liberals (e.g. conservative criticisms of religious political influence) and blue books that appeal to the Tea Party (e.g. Alinsky-inspired community organizing).

Political Polarization

Political polarization measures the extent to which interests from conservatives and liberals in a given topic diverge. In other words, even if conservatives and liberals are equally likely to buy books of a given topic, they might buy distinct books. Polarization identifies this possibility. For a given topic, we compute the number of books within the topic linked to both red and blue books, divided by the number of books linked to either (we call this quantity "overlap," denoted by O), compared with its expected value in a null model where links from red and blue books are randomly assigned to books in the topic.

Specifically, for each link between books in a given topic and political books, we shuffle the book in the topic to a randomly chosen topical book, also linked to political books. This results in all politically relevant books from the topic being linked to political books uniformly at random. After all political links are randomized, the fraction of books linked to both red and blue among books linked to either is calculated in the randomized network. 100 such random simulations are carried out to obtain a distribution of the overlap between blue and red (i.e., the fraction of books linked to both blue and red) in the random model. Finally, polarization is measured as the difference between the expected overlap in the random null model and the observed overlap:

$$\frac{E[O]-o}{\sqrt{Var[O]}},$$

If the observed overlap is smaller than what would be expected at random, polarization is positive and increases with the difference between expected and observed overlaps. Polarization of a topic equals zero when red and blue books are co-purchased with books within the topic uniformly at random, but increases as the sets of red- and blue- linked books diverge, indicating red and blue preferences for distinct books.

Applied Index

To quantify the extent to which the political alignment of scientific fields is correlated with their application, we developed an index measuring the extent to which a field is commercially applied.

Among many possible measures of commercial application, a reasonable and tractable one is the degree to which patents build upon knowledge produced by the field. Moreover, with digital patent databases we are able to quantify the contribution to patents made by the fields of science at a comprehensive scale. To that end, we use the US patent database from Google, 1976- 2014, which is the most digitally complete. For each journal we tabulate the number of times that journal was cited by all patents in the patent database. We aggregated these within the Dewey Decimal (DD) and Library of Congress (LOC) categories in which Amazon books are categorized, such that the number of citations received by a field is computed as the sum of citations received by all the journals in that field.

Citations from patents meaningfully reflect gross contributions of commercial relevance to the real world, but they do not fully capture the degree to which a field is applied, because the number of citations is strongly influenced by field size and activity. Imagine a small, focused field that is commercially applied, compared with a large, broad field. Perhaps, the smaller, more focused field receives fewer citations from patents only because it produces fewer total patents, but it is actually more application-oriented because all knowledge in the field is transferred to technology. Therefore, we build a field-level citation network to capture how active or impactful a field is in the scientific space.

We find that the number of citations a subdiscipline (as defined by our classification of science categories) receives from patents is strongly correlated (r=0.8, p<0.001) with the number of citations it receives from other subdisciplines (Fig. S1.3). This correlation reveals that most subdisciplines are proportionally active in both patent and academic domains, and any one of the two types of citations alone is not sufficient to estimate how "pure" or commercially applied the field might be. Accordingly, we constructed a commercial applied index that combines both kinds of citations and measures how much each subdiscipline is cited by patents relative to articles. Specifically, for each subdiscipline i, we denote the number of citations from patents by y_i and the number of citations from articles in other subdisciplines x_i. The expected number $E[Y]$ of citations from patents given the number X of citations from other subdisciplines is modeled $E[Y]=\exp(aX+b)$, which explains the correlation illustrated in Fig. S1.3. Finally, the commercial application index A of a subdiscipline is calculated as the number of citations from patents normalized by its expected number of patent citations given the number of citations from other subdisciplines: $A_i = y_i / \exp(ax_i + b)$.

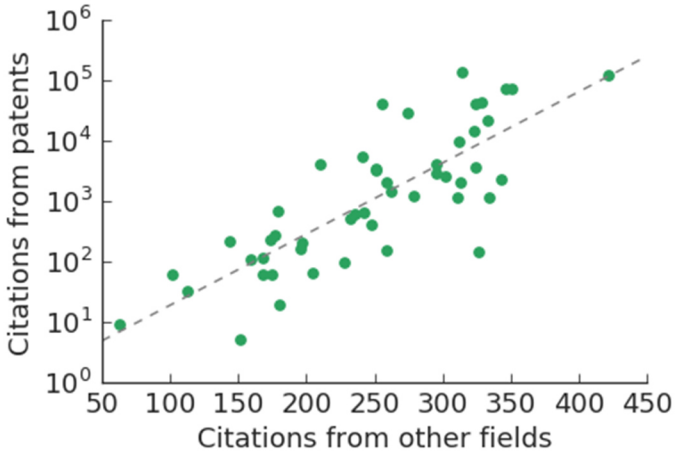


Figure S1.3. Positive correlation between the number of citations a sub-discipline receives from patents and from other sub-disciplines.

A linear regression model $log(Y) = aX + b$ is fitted to data with estimate $a = 0.0279$ (p-value < 0.001).

Scientific Breadth

Scientific breadth measures the breadth of interests in science from conservatives and liberals. For example, if red books link to a narrow subset of books within a discipline, while an equal number of blue books connect to a large and diverse subset of disciplinary books, those purchasing blue books have exposure to a wider range of science books—and likely a wider range of scientific perspectives—than those purchasing red books.

With respect to each discipline, the scientific breadth of conservatives (or liberals) is defined by the number of distinct titles within the discipline linked to red (or blue) books divided by the number of red (or blue) books linked to the discipline.

Centrality of Science Books

The network location of each book with links to red and blue relative to the core or periphery of a disciplinary book network was quantified by assessing the closeness centrality (*22*) of each book with respect to the given disciplinary network. Centralities of blue-linked books were then compared with those of red-linked books within each discipline to assess *core location* of blue and red with respect to the discipline. Note that books at the center of a disciplinary book network are scientifically important—co-present in personal libraries with many other disciplinary books. Books at the periphery are rarely purchased with other disciplinary books.

To compare the difference between centralities of blue- and red-linked books across disciplines, we calculated the widely used *t*-score

$$\frac{E[X]-E[Y]}{\sqrt{S^2(1/n_x+1/n_y)}}$$

for each discipline, where X corresponds to centralities of blue-linked books, Y to centralities of red-linked books, n_x to the number of blue-linked books, and n_y to the number of red-linked books. S^2 is the pooled sample variance of blue- and red-linked books.

The standardized difference (*t*-score) between centralities of blue- and red-linked books is shown in Fig. S1.4. The mean centrality of blue-linked books is larger than that of red-linked books for most disciplines, and as polarization increases, blue-linked books are more central than red-linked books. Exceptions to this pattern include history and economics (see Fig. S1.4).

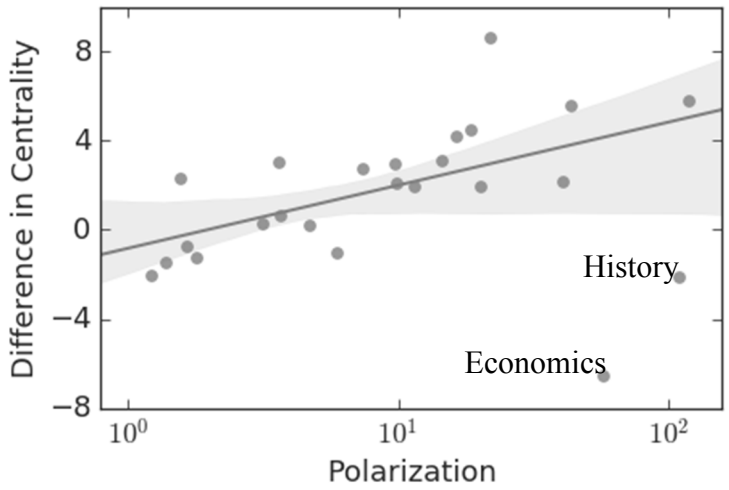


Figure S1.4. Standardized difference between centralities of blue- and red-linked books, by polarization for each discipline.

For each discipline, the plot shows the centrality difference $\frac{E[X]-E[Y]}{\sqrt{S^2(1/n_x+1/n_y)}}$ against polarization, where X corresponds to centralities of blue-linked books, Y to centralities of red-linked books, n_x to the number of blue-linked books, and n_y to the number of red-linked

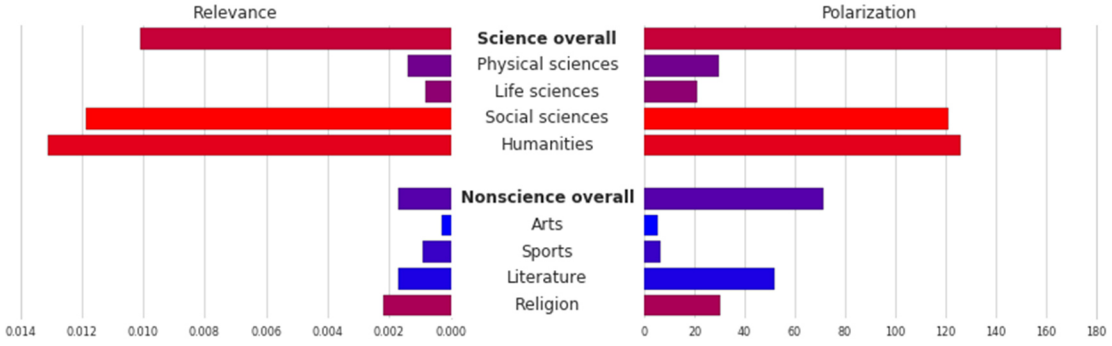
books. S^2 is the pooled sample variance of blue- and red-linked books. A robust linear regression line is shown in the plot with slope 1.2278 (p-value<0.001).

Academic Books

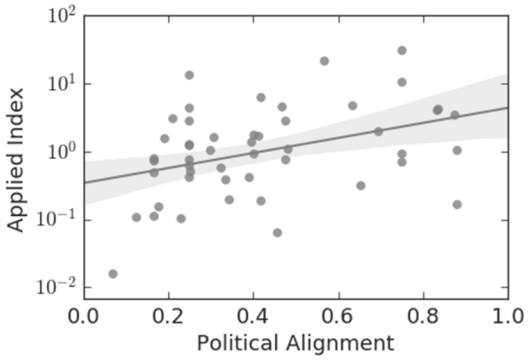
To test the effect of academic books on our findings that liberals have a wider interest in science books within most disciplines, we identified a large set of academic books from our dataset and re-performed our analysis after removing the academic books.

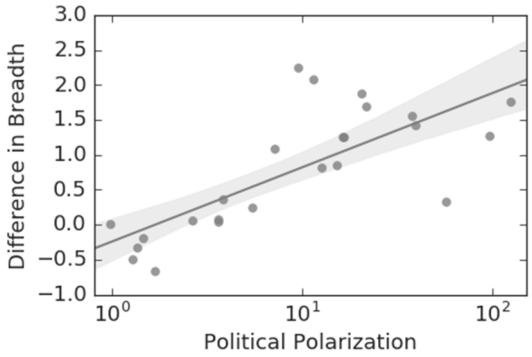
Academic books were identified according to their publishers. First, we compiled a list of academic publishers from the website (services.exeter.ac.uk/bfa/az.htm), to which we added all publishers with "university" or "academic" in their names and manually filtered out publishers that are clearly not academic (e.g., Trump University Press). Finally we classified each book as academic or not by using its publisher as a proxy. In total 136,672 academic books (about 10% of all books) were identified in the dataset.

We performed all the analyses after academic books were removed, and present the results in Fig. S1.5. The results, after removing academic books, remain consistent with the results we report in the main text. Figure S1.5A shows that science is neither an apolitical sphere nor a public sphere. Science is more relevant and polarized than topics outside of science, and is polarized both within and across scientific disciplines. Across fields, there is a significant positive correlation (r=0.39, p=0.005) between political alignment and the commercial applied index of sub-disciplines (Fig. S1.5B), implying that customers for liberal books prefer basic science while conservative customers prefer commercially applied science. Within disciplines, Figure S1.5C plots political polarization by the difference between the average number of disciplinary books linked to a blue book and the average number linked to a red book, for each discipline; and Figure S1.5D reports the difference between closeness centralities of blue-linked and red-linked books for each discipline (cf. Fig. S1.4). Figure S1.5 C and D together reveal that red books are more likely to cluster on the periphery of the disciplinary networks, with blue books linked to a wider variety of science books and blue-linked science books closer to the disciplinary core. Given that academic books are removed, this wider liberal interest in science books does not appear to be a simple consequence of academic liberalism.

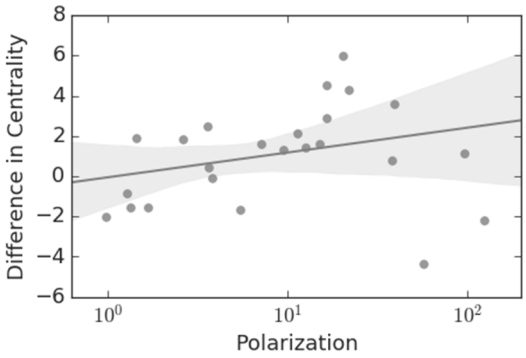


A





B C



D

Figure S1.5. Reproduction of major findings in the main text after academic books are removed.

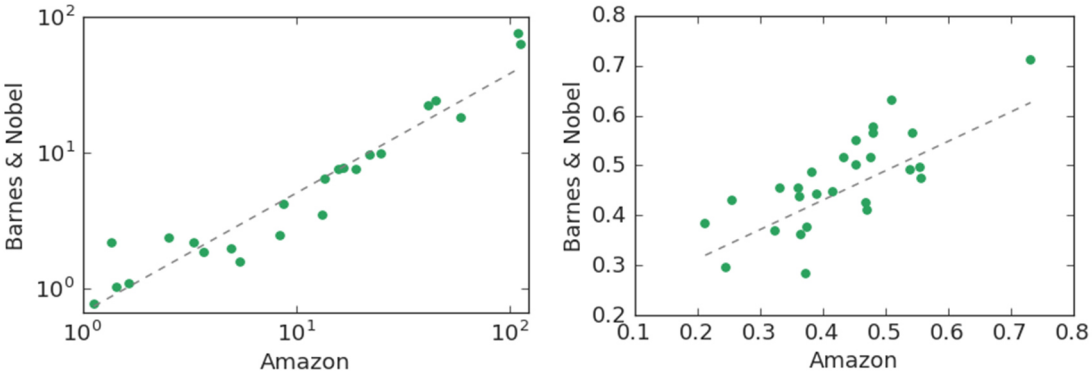
(A) Political relevance and polarization of science topics compared to topics outside of science. (B) Correlation between political alignment and commercial applied index of sub-disciplines (r=0.39, p=0.005). (C) Difference between the average number of science books linked to a blue vs. red book (scientific breadth), by polarization for each discipline. The difference in scientific breadth and *log*(polarization) are highly correlated with r=0.75 and

$p<0.001$. (D) Difference between centralities of blue- and red-linked books for each discipline. A robust regression line is shown as a guide to the eye. A simple linear regression gives slope 1 and p-value 0.003 with two outliers economic and history removed.

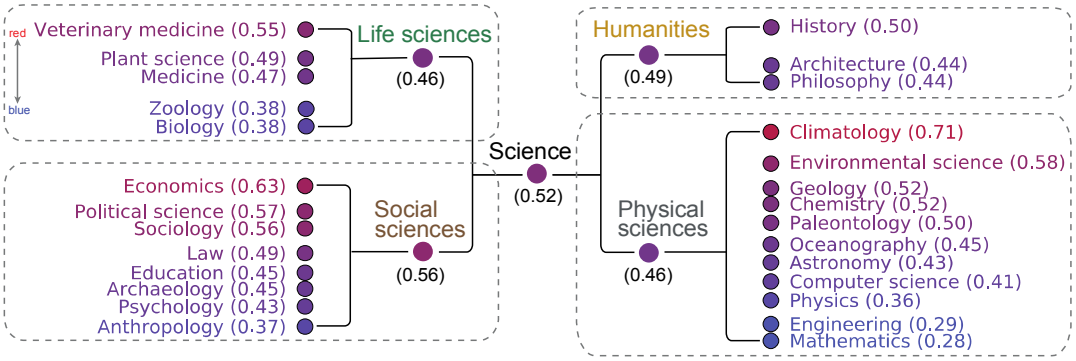
Results from Barnes & Noble

Books in the Barnes & Noble dataset constitute a subset of Amazon political and science books, but the two co-purchase networks are quite different. For example, only 9% of Amazon co-purchase links in the science and politics subgraph are found in the Barnes & Noble network, and only 21% of Barnes & Noble links are found in Amazon. See Table S1 for a brief summary. Nevertheless, the number of political links with each book in Amazon and Barnes & Noble is highly correlated ($r=0.60$, $p<0.001$).

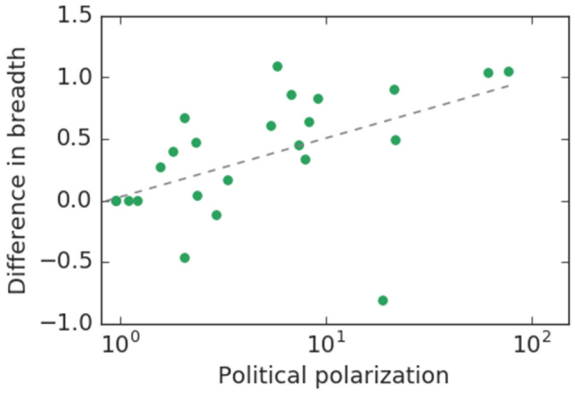
We replicated the Amazon-based analyses using the Barnes & Noble network and found consistent results across datasets. First, we calculated political polarization and alignment of science topics in Barnes & Noble. These measures are compared with those of corresponding topics in Amazon (Fig. S1.6, A and B). The polarization scores of disciplines given by the two networks are nearly identical ($r=0.97$, $p<0.001$); the alignment scores of disciplines are also highly correlated across the two networks ($r=0.76$, $p<0.001$).


Barnes & Nobel
Amazon
Barnes & Nobel
Amazon

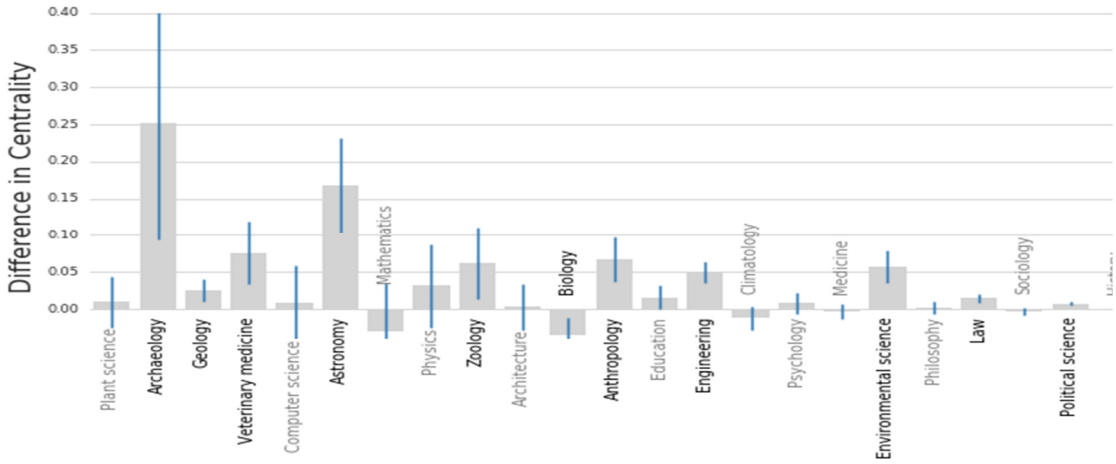
A B


red
blue
Veterinary medicine (0.55)
Plant science (0.49)
Medicine (0.47)
Zoology (0.38)
Biology (0.38)
Life sciences
(0.46)
Economics (0.63)
Political science (0.57)
Sociology (0.56)
Law (0.49)
Education (0.45)
Archaeology (0.45)
Psychology (0.43)
Anthropology (0.37)
Social sciences
(0.56)
Science
(0.52)
Humanities
(0.49)
History (0.50)
Architecture (0.44)
Philosophy (0.44)
Physical sciences
(0.46)
Climatology (0.71)
Environmental science (0.58)
Geology (0.52)
Chemistry (0.52)
Paleontology (0.50)
Oceanography (0.45)
Astronomy (0.43)
Computer science (0.41)
Physics (0.36)
Engineering (0.29)
Mathematics (0.28)

C


Difference in breadth
Political polarization

D



E

Figure S1.6. Results from the Barnes & Nobel dataset.

(A) Political polarization scores of scientific disciplines calculated from the Barnes & Nobel co-purchase network (*Y*) and those from the Amazon network (*X*). The scores from the two dataset are nearly identical (r=0.97, p<0.001). (B) Political alignment of disciplines calculated from the two datasets. The alignment scores are highly correlated across the two networks (r=0.76, p<0.001). (C) Political alignment of disciplines, organized as a tree of science topics. (D) Polarization by the difference between the average number of science books linked to a blue vs. red book, for each discipline. The difference in scientific breadth is significantly correlated with polarization (r=0.54, p=0.005). (E) Difference between mean closeness centralities of blue- and red-linked books for each discipline. Except for biology, mean centrality of blue-linked books is no less, and for half of the disciplines, significantly larger than that of red-linked books.

Alignment of disciplines within the four "schools" in Barnes & Nobel is reported in Fig. S1.6C. There are not enough books in the sub-disciplines from the Barnes & Nobel dataset to statistically test the correlation between alignment and applied index. However, the general pattern still holds: Applied disciplines like veterinary medicine and economics are at the red end of their respective schools, while anthropology and mathematics are most blue.

Finally, across disciplines, blue books link to a larger number of disciplinary books compared to red (Fig. S1.6D), consistent with results obtained with Amazon data. And for most disciplines, science books linked to blue are more central in their respective disciplines in terms of closeness centrality than books linked to red (Fig. S1.6E).

A MEMBER SAVED IS A MEMBER EARNED? THE RECRUITMENT-RETENTION TRADE-OFF AND ORGANIZATIONAL STRATEGIES FOR MEMBERSHIP GROWTH[2]

Abstract

A long line of research documents the essential role of social networks in mediating the recruitment and retention of members in organizations. But organizations also comprise a primary context where people form social ties. We investigate how the network structure that an organization creates among its members influences its ability to grow and reproduce itself. In particular, we propose that two dimensions of organizational strategy influence affiliation dynamics: (1) the extent to which the organization induces social interaction among its members (social encapsulation), and (2) the time and energy that the organization demands of its members. We examine membership dynamics in an ecology where competitor organizations deploying varied strategies vie for the same pool of members. Results show a curvilinear relationship between membership growth and the rate of social encapsulation. Further, we find that an organization's time and energy demand mediates the effect of social encapsulation. For most types of organizations attaining sustained growth requires a balance between open networks (for recruitment) and network closure (for retention).

[2] This is a paper coauthored with Fedor Dokshin, Michael Genkin, and Matthew E. Brashears. Yongren Shi is the first author.

Introduction

Both formal organizations and informal associations face a perplexing question: how to attract and retain members when people only have limited time and energy for participation. A substantial amount of research on organizational recruitment suggests that social network ties between members and outsiders play a crucial role in promoting recruitment and facilitating membership growth in churches (Stark and Bainbridge 1980), religious cults (Snow, Zurcher and Ekland-Olson 1980; Rochford 1982), social movements (McAdam and Paulsen 1993; Walgrave and Wouters 2014), voluntary associations (McPherson and Ranger-Moore 1991), and even terrorist organizations (della Porta 1988). While network ties have been consistently shown to be an important factor in organizational recruitment, less attention has been paid to the role of organizations in shaping those networks (Edwards and McCarthy 2004). As key foci of interaction, organizations mediate tie formation and decay (e.g., Feld 1981; Putnam 2000; McPherson 1983), but we nevertheless know little about how different organizational structures facilitate recruitment and retention.

Although our understanding of recruitment and retention processes has been greatly enhanced by incorporating information about network ties, existing research has also long been criticized for relying on an overly simplistic view of network structure (e.g., Kitts 2000; Gould 2003; McAdam 2003). Network ties to organizational members are generally assumed to have a positive effect on recruitment into the organization (McAdam and Paulsen 1993), and empirical studies tend to use simple counts of such ties to estimate recruitment potential (Gould 1991). However, whether a given tie results in recruitment also depends on the broader organizational

context (McPherson 1983; Kitts 1999). Two facets of organizational contexts are particularly important: level of network closure among members; and the distribution of the behaviors and attitudes that are common in the organizations in the population. First, the positive effect of network ties on recruitment could be negated by the effects of ties to competing organizations, which can drain members from the organization (McPherson, Popielarz and Drobnic 1992; Kitts 2000). As such, networks not only structure an individual's availability for movement participation (McAdam and Paulsen 1993; Snow *et al*. 1980), but also pull existing members out of organizations they had previously joined (McPherson *et al*. 1992). Without accounting for the connections to outsiders, researchers could misunderstand the true impact of network ties on recruitment. Secondly, network ties are more likely to result in recruitment if the potential member is behaviorally or affectively predisposed toward the demands of organizational membership (McAdam 1986, 2003). Conversely, network ties to organizational members may have little impact on individuals who do not share an interest in, or sympathy with, the organization's goals and/or behaviors. A network tie provides an opportunity, but this opportunity likely will not be realized if the behavioral repertoires of the recruit and the organization differ too substantially (McAdam and Paulsen 1993; Lim 2008; Kitts 2000).

We argue that a focus on organizational contexts is necessary to an understanding of recruitment and retention. Specifically, we draw on an ecological perspective of organizations (e.g., McPherson 1983) to develop a dynamic model of organizational recruitment and retention in which organizations with different internal structures compete for a common, finite pool of members. We identify two

dimensions of organizational structure that influence membership dynamics: (1) the extent to which the organization focuses social interaction among members (social encapsulation), and (2) the amount of time and energy that it demands of members (time and energy demand). These two dimensions shape the organization's position in relation to the broader social environment. Higher levels of social encapsulation increase retention by reinforcing commitment among existing members, but may reduce an organization's potential for recruitment. Higher time and energy demand makes it more difficult to recruit new members, but also limits the number of alternative affiliations that the members can hold, thus potentially decreasing competition.

We formalize these ideas in an agent-based model (ABM) and use it to pursue questions at two levels of analysis. First, at the organizational level, we explore how the interaction of the two dimensions influences an organization's fitness within an ecology. Second, at the level of the organizational environment, we explore how adoption of different strategies by competitors influences the overall level of competition and, consequently, the range of successful strategic responses. In other words, the strategies of competitors can significantly alter those that an organization can pursue successfully.

Recruitment and Retention of Members in Organizations

While organizations have a variety of goals, those that do not act to secure their own survival are likely not around long enough for their other goals to matter. Organizations require many different resources for survival (Pfeffer and Salancik

1978), but contributions by human members are most fundamental. Members not only fulfill essential organizational roles, they also provide the means for obtaining all other resources (Scott 1992:171; McPherson 1983; Simon 1976). Simply put, without members an organization ceases to exist. Whether an organization sustains sufficient membership depends on two processes: recruitment and retention. Recruitment is the rate at which new members enter the organization, while retention is the rate at which existing members remain within the organization. Membership grows when the number of new recruits outweighs the number of members lost, and it declines when the organization fails to replace members who exit. Regardless of the organizational domain, the existential challenge facing every organization is to maintain recruitment and retention at levels that enable it to grow or at least not decline in the long run.

Seminal sociological research on recruitment and retention was done in the context of religious movements, cults, and communes (e.g., Lofland and Stark 1965; Kanter 1972; Stark and Bainbridge 1980; Rochford 1982). This research, consisting largely of qualitative case studies, examined the strategies organizations use to attract new members and secure their long-term commitment to the group. A key sociological insight to emerge from this work is that the structure of social ties—between members and non-members as well as among members—is critical for explaining organizational growth and decline (Snow *et al*. 1980; Stark and Bainbridge 1980). The concentration of ties among members reinforces commitment and enhances retention (Kanter 1972; Coser 1974); while social ties to non-members shape an organization's recruitment opportunities (Snow *et al*. 1980; Rochford 1982). For example in her research on American communes, Kanter (1972) observed that geographically isolated

communities were more likely to persist, because they eliminated their members' competing social obligations. Communes that did not eliminate opportunities for outside interaction declined much more quickly. While structural cohesion among members promotes retention, other case studies showed that ties between members and non-members can facilitate recruitment (e.g., Snow *et al*. 1980; Stark and Bainbridge 1980; Rochford 1982).

The case study literature thus brought into focus a general tension between an organization's need to enclose existing members in a cohesive structure, on the one hand, and its need to maintain channels for recruitment, on the other hand. It further suggested that the effectiveness of alternative network configurations for recruitment and retention depends on the surrounding social environment. Examining the Hare Krishna religious movement, Rochford (1982) found that the group maintained an exclusive structure during the period of growing skepticism toward cults in the early 1970s. Closing off members from extra-movement ties protected the group from potential defection. The Hare Krishna opened its ranks, however, when local conditions were favorable—for example in Los Angeles, where devotees' social embeddedness in the local community became an asset for recruitment.

The early case-based literature suggests that the configuration of members' networks influences membership growth, but it emphasizes the contingency of these network effects on the surrounding social environment. Subsequent research on recruitment and retention integrates lessons from this research only selectively. In particular, it is now widely accepted that social network ties serve as channels for recruitment into every type of organization (e.g., McAdam 1986; Granovetter 1995;

Marsden and Gorman 2001; Walgrave and Wouters 2014), but three specific lessons from the case-based literature have largely been ignored. First, network ties can lead to exit as well as recruitment (see Kitts 2000 and Gould 2003 for critical discussions of this point; also see McAdam and Paulsen 1993 and Kitts 1999 for exceptions). Second, organizations have different structures, and therefore manage the boundary maintenance problem differently (Coser 1974; Rochford 1982; Edwards and McCarthy 2004). Dense intragroup networks provide mechanisms for retention, but organizations that adopt such structures isolate themselves from the outside social world, limiting recruitment potential. Likewise, organizations that allow members to maintain many ties to non-members facilitate recruitment, but may lose members due to increased opportunity for exit. Finally, few studies consider that social ties have heterogeneous effects on potential recruits, based on their behavioral and affective disposition toward the organization. A notable exception is McAdam's analysis of Freedom Summer (1986), in which students who were not well committed behaviorally or ideologically were less likely to join the movement organization. However, the issue is likely to apply broadly to many types of organizations, and is currently under studied. An ecological perspective (e.g., McPherson 1983) can help integrate these important elements and improve our understanding of key organizational outcomes.

An Ecological Perspective on Membership

From an ecological perspective, group memberships and other social obligations vie for limited human attention (McPherson 1983; Simmel 1955; McAdam and Paulsen

1993). Given a tie between a member and a non-member, the tie may either result in the recruitment of the non-member into the organization or it may result in the exit of the member from the organization (Gould 2003; Kitts 2000). Social ties that span organizational boundaries facilitate inter-organizational competition and represent both a resource (for recruitment) and a potential liability (for retention).

McPherson (1983) offers a framework—the ecology of affiliation—for thinking about inter-organizational competition for participants. Within this framework an organization's position relative to its competitors is key for its survival. Under the assumption that recruitment into an organization is channeled primarily through network ties, it is possible to define a region in the global social network, called a *niche*, where each organization tends to obtain its resources (i.e., human participation).[3] An organization thrives in a niche that is exclusive, and it may shrink or move if its niche overlaps with many competing organizations. Following this basic model, McPherson and others show in a series of empirical studies that the level of competition within the niche predicts membership growth and decline (McPherson and Ranger-Moore 1991; Popielarz and McPherson 1995; Stern 1999; Rotolo and McPherson 2001; Mark 1998).

McPherson and colleagues find that the competitive pressure within the local environment is critical for understanding organizational survival and growth. But the empirical literature on recruitment and retention also highlights the importance of

[3] McPherson defines the niche specifically as a limited region in a multi-dimensional demographic space (called Blau space) (1983, 2004). Blau space serves as a probabilistic approximation of the global social network, because network ties are more likely to form among similar others (McPherson, Smith-Lovin, and Cook 2001).

internal organizational strategies for membership growth (Kanter 1972; Coser 1974; Stark and Bainbridge 1980; Rochford 1982; Edwards and McCarthy 2004). This implies two questions that existing ecological models do not address: (1) what kinds of strategies help organizations become more competitive within their local environment and (2) how do the multiple differing strategies in the local environment impact niche dynamics?

We focus on two factors that previous literature suggests may influence an organization's competitiveness: (1) social encapsulation of members and (2) the amount of time and energy that organizations demand in exchange for membership. We discuss each factor before turning to a discussion of how local competitive dynamics vary across organizational environments characterized by different types of organizations.

Social Encapsulation

All organizations facilitate the formation of social ties among their members by providing a common context for interaction (Feld 1981). However, the intensity of interaction varies widely across organizations with different internal structures. We use the term *social encapsulation* to refer to the rate at which membership in an organization compels members to create social ties within the group.[4]

Coser's (1974) discussion of "greedy institutions" provides a useful illustration of how organizations encourage network closure. Organizations rely on greedy

[4] This term was popularized in the literature on conversion, especially in the context of cults and communes (Lofland 1978).

institutions, which consist of formal and informal rules, "…to reduce the claims of competing roles and status positions on those they wish to encompass within their boundaries" (Coser 1974, p. 4). Greedy institutions take on many forms, but they all serve to strengthen social bonds to the group, while eliminating the tension associated with multiple embeddings. For example, Kanter (1972) suggests that the seemingly contradictory practices of free love in the Oneida Community and sexual abstinence among the Shakers, both served to discourage members from forming private ties that channel energy and loyalty away from the group and toward the competing institution of marriage. These are of course extreme examples, but all organizations engage in some form of social encapsulation. Professional organizations may host a reception to encourage new and existing members to mingle (Ingram and Morris 2007) and bowling leagues create social bonds among the players over friendly competition (Putnam 2000). While some organizations promote greater closure, others encourage members to maintain network ties with outsiders, which may increase recruitment. For example, Stark and Bainbridge (1980) attribute the growth of the Mormon Church to its members' maintaining open networks. They offer a revealing quote from a church publication, which urges members: "Don't be exclusive" (1980:1387).

Besides the specific measures that organizations take to achieve a desired level of closure, a growing body of research finds that different network structures emerge between members unintentionally, as byproducts of organizational activities and organizational contexts (Burger and Buskens 2009; Small 2009; McFarland *et al*. 2014). For example, Small (2009) examined network formation among adult clients of daycare centers and found that simple organizational rules have unintended

consequences for social closure. By enforcing a strict pickup schedule, for instance, some daycare centers unwittingly encouraged parents to develop relationships. By comparison, parents were much less likely to befriend one another if the pickup time was not standardized.

Thus there are many different features that shape the level of network closure within an organization, not all of which are within the organization's control. We set aside the question of *how* organizations facilitate the creation of different types of networks, and simply observe that organizations induce different levels of network density among their members. Since individuals are often affiliated with multiple organizations, each having its own rate of network rewiring, their personal networks are partially a function of the network pressures that these memberships impose. Moreover, individuals' capacity for maintaining social ties is limited by the time available for social activity and by the cognitive and emotional effort required for maintaining social ties (Brashears 2013; Dunbar 1992; Stiller and Dunbar 2007). Therefore, an increased connection to one set of associates typically comes at the expense of decreased connection to others (Saramäki *et al*. 2014).

Time and Energy Demand

The second organizational characteristic we consider is the demand for time and energy that an organization places on its members. Organizations prescribe, either formally by instituting schedules or through informal rules, a level of commitment that individuals must meet to attain and preserve membership (McAdam 1986; Cress, McPherson, and Rotolo 1997; Iannaccone 1994). The importance of demand for

affiliation dynamics rests on the fact that humans have finite time and energy, thus imposing a constraint on the number and volume of organizational commitments they can maintain (McPherson 1983). This can be treated as a single undifferentiated pool, which is the strategy adopted by most research (e.g., McPherson 1983; McPherson and Rotolo 1996; Cress *et al*. 1997), or it can be divided into blocks that specific organizational types preferentially exploit (Winship 2009). For example, voluntary organizations generally compete for time during non-work hours, which are relatively interchangeable and do not overlap with working hours. In either case, organizations that exploit the same type of people and hours are competing for the same limited pool of time, and those exploiting the same type of people, but in different hours, are not competing (e.g., Kitts 1999).

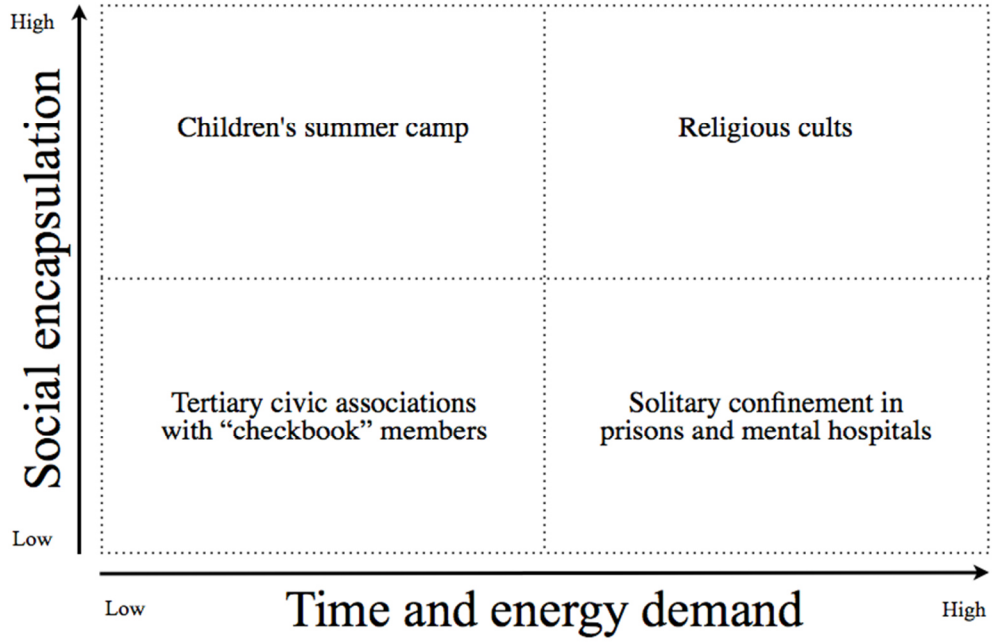


Figure 2.1. Strategy space formed from the two dimensions, social encapsulation rate (Y-axis) and time and energy demand (X-axis).

Organizational Strategies and Membership Growth

Figure 2.1 displays the strategy space with time and energy demand on the X-axis and social encapsulation on the Y-axis. We can define the corners on one diagonal with ideal types of a greedy institution (top right) and an organization that minimally engages members (bottom left). An example of the former is a cult. Cult activities may take up every waking hour of its members, and devoted cult members associate exclusively with fellow members (Coser 1974). In contrast, organizations occupying the bottom left corner neither demand a lot of time and energy nor compel their

members to develop in-group ties. Examples include some contemporary civic organizations, which often facilitate no face-to-face interaction at all and which demand only that members renew their annual dues (Putnam 1995; Skocpol 2003). As Putnam (1995) explains, the ties between members of such civic organizations are "like the bond[s] between any two Red Sox fans (or perhaps any two devoted Honda owners)" (p. 71). All organizations along the bottom of the strategy space foster little social connection between members; the ones in the bottom right corner combine this lack of social connection with extreme levels of demand for time and energy. Solitary monasticism is one example of this extreme, though many empirical examples in this region of the space likely rely on coercion (e.g., sweatshops and solitary confinement in prisons and mental hospitals). Finally, the top left corner includes organizations that achieve high levels of network closure among members while requiring relatively little time and energy. An example is a kids' summer camp, which while lasting only a couple of weeks during a year forges friendships that keep the campers coming back year after year.

The ideal types represented by these four corners reflect an implicit assumption that social encapsulation and energy demand are separate dimensions despite the likelihood of a relationship between them. However, because tie formation requires some investment of time (Marsden and Campbell 2001; Sarämaki *et al*. 2014), in limiting cases, higher social encapsulation will require higher time and energy demand; for example, one must convince members to spend their time in settings that will encourage tie formation with other members. Although we know this relationship is present, the existing literature offers little guidance for specifying its nature.

Moreover, the limit that time imposes on tie formation and maintenance is not constant, but reflects changes in available communication technologies (Mayhew and Levinger 1976; Wellman 2001). We therefore define the two dimensions separately in order to explore the entire strategy space without wedding our model to a link function that may not obtain empirically. Our framework, furthermore, does not imply a difference based on organizational function—i.e., regardless of function, organizations can be located along the two dimensions we define. However, it may be the case that organizations with different functions will concentrate in certain parts of the space (e.g., social clubs will tend to have greater social encapsulation rates than organizations that pursue some instrumental goal). Future studies examining the empirical distribution of organizations in this space will provide useful insights into the nature of time use and organizational life under those specific conditions.[5]

What is the optimal strategy for achieving membership growth and survival? Intuitively, we expect that a higher rate of social encapsulation will help an organization to retain members, but it will also decrease the organization's recruitment capacity. On the other hand, a low rate of social encapsulation will enable members to maintain more ties to nonmembers, facilitating recruitment, but will sacrifice retention capacity. Organizations that balance these countervailing pressures will be best positioned to expand their memberships. A central question, therefore, is where this balance lies, and whether it changes with the types of demands organizations place on members.

[5] We thank an anonymous reviewer for helping us elaborate this point.

With respect to demand, previous research suggests that it is more difficult to recruit members for activities that require greater commitment (Rochford 1982; McAdam 1986). Research within the affiliational ecology tradition further finds that greater demand of time and energy is negatively related to membership duration (i.e., retention) (Cress *et al*. 1997), a result that finds theoretical support in simulation studies (McPherson and Smith-Lovin 2002; Geard and Bullock 2012). We expect therefore that more demanding organizations will struggle to recruit and retain members. However, while high levels of demand impose constraints on survival, we expect that higher rates of social encapsulation might slow decline by improving retention. This expectation follows Coser's (1974) intuition behind greedy institutions. A highly demanding organization will persist longer when its members' social networks present fewer alternatives.

But time is not entirely fungible as temporal allocation is often partitioned via scheduling. Although our primary focus is on the interplay between social encapsulation and organizational demand, we take an initial step to investigate this potentially important relationship. Organizations that occupy different parts of the schedule do not directly compete with each other, but they shape the network structure of their members and may thereby steer members toward specific affiliations in a non-overlapping segment of the schedule. Consistent with this logic, research in social movements has long observed a mutualistic relationship between churches and social movement mobilization (e.g., Morris 1984). The social bonds among congregation members may reinforce commitment to the social movement and vice versa. Thus, allowing for schedule partitioning, we expect that an organization in one segment of

the schedule will develop mutualistic relationships with one or more organizations in other segments, and that the strength of the relationship will be proportionate to the rate of social encapsulation.

Dynamics of Local Competition

The core idea of every ecological perspective is that the local competitive environment constitutes an essential context for organizational growth and survival (Popielarz and Neal 2007). Therefore, a focal organization's membership depends not only on its strategy, but also on the types of competitor organizations within the local environment. McPherson and colleagues measure competition as the extent of overlap between rival organizations and a focal organization's niche (McPherson and Rotolo 1996; McPherson and Smith-Lovin 2002). The more organizations there are vying for the same pool of human resources the greater the competition. This conception, however, ignores the point that organizations respond differently under the same conditions, based on the strategies they employ (e.g., Rochford 1982). Therefore, niche overlaps that appear equivalent in a cross-sectional view, may lead to different levels of constraint on membership growth in the long run, depending on the distribution of strategies among competing organizations.

We define *competitive pressure* broadly as the aggregate constraint that competing organizations place on a focal organization's ability to recruit and retain members (i.e., how hard do other organizations make it for your own to gain, and retain, members) and make two predictions about local competitive dynamics. First, somewhat counter-intuitively, we expect that environments populated by low-

demanding organizations will induce greater competitive pressure than high demanding organizations. The key mechanism is the level of social integration among potential recruits. Individuals in environments characterized by high demanding organizations may be limited to just one, all-encompassing affiliation. While this reduces the volume of available members for other organizations, each highly demanding organization settles in a relatively circumscribed niche and does not induce direct competition between organizations for the same members. In environments populated by low-demanding organizations, by contrast, people can hold multiple affiliations, which provide a basis for direct competition (McAdam and Paulsen 1993). People are also more likely to maintain incipient interests, which make them candidates for recruitment into new organizations.

Second, when multiple affiliations are possible, competitive pressure will depend on neighboring organizations' levels of social encapsulation. Niche overlap is essential to inter-organizational competition, but whether such multiple embeddings undermine a focal organization's cohesion further depends on competitors' ability to churn the personal networks of the organization's members. Some moderate level of social encapsulation should impose the greatest pressure on the focal organization. With either very low or very high levels of social encapsulation, the overlapping competitors are not capable of generating pressure, either because they are too apathetic to create cohesive cores of members, or too greedy in isolating themselves from competition.

Model Framework

To examine the propositions laid out above, we specify and test an agent-based model (ABM) (e.g., Macy and Willer 2002). Previous work provides solid theoretical footing for specifying micro-mechanisms underlying recruitment and retention processes, but how these mechanisms aggregate to give rise to variable patterns of growth and decline at the level of a social system is unclear. Thus, we consider membership dynamics to be emergent of the interactions between individual agents within constraints imposed by their existing organizational affiliations. In this sense, our explanation of membership dynamics follows the structure of "Coleman's boat," which emphasizes that causal relationships between two macro-level phenomena require the specification of a macro-micro-macro causal chain (Coleman 1990). In our case, we are interested in how organizational strategies affect membership outcomes in a competitive ecology. Therefore, we posit that (1) organizations impose specific constraints on the behavior of their members, (2) these constraints influence the personal networks and behavioral commitments of individual agents through network processes, and (3) micro-level network dynamics give rise to specific patterns of membership growth and decline.

General Modeling Approach

We are interested in explaining the effect of two fundamental organizational properties—social encapsulation and time and energy demand—on the growth and

persistence of organizations.[6] Our modeling approach fits within the ecology of affiliation tradition of organizational analysis (McPherson 1983), and specifically shares affinities with two previous models that formalized competitive dynamics among organizations. McPherson and Smith-Lovin (2002) exogenously varied the level of competition in the system and observed the influence it had on membership duration. Geard and Bullock (2010) developed a more complex model, in which competition emerges endogenously through interactions between individual agents, and examined the effect of different organizational demands on several system-level outcomes.

We build on this work in several important ways. First, we introduce social encapsulation as an organization-specific property, allowing us examine how different encapsulation strategies influence retention and recruitment in a competitive ecology. Thus, organizations in our model not only place different demands on their members, but also pursue alternative strategies to either eliminate members' competing obligations, enhance members' ability to recruit, or both. Second, we extend formal approaches to affiliation dynamics by combining existing formal architecture with an explicit model of social influence (DellaPosta, Shi, and Macy 2015). In modeling the spread of organizational behaviors through social networks, we capture the socializing process that typically precedes recruitment (e.g., Stark and Bainbridge 1980; McAdam 1986). In previous models of affiliation dynamics (McPherson and Smith-Lovin 2002;

[6] The focus on organizational outcomes distinguishes our approach from models of team assembly (e.g., Guimera et al. 2005), which seek to understand why particular sets of people come together to accomplish a *particular* task or project.

Geard and Bullock 2010), by contrast, agents with a tie to a group member are equally likely to join the group, irrespective of organization's demand.[7] Further, with our formalization of social influence, we are able to relax the assumption of resource fungibility and examine the effect on resource partitioning, including the fact that organizations that schedule activities at different times are not direct competitors. Overall, our model moves beyond the prevailing image of organizations in the ecology of affiliation tradition as mere containers of members (e.g., McPherson 2004), to entities that fundamentally change who members know and what members do.

Figure 2.2 provides an illustration of a hypothetical social system with two organizations (black and gray) and nine agents (represented by rectangular nodes). Each agent has five time slots, which she fills with different behaviors, including organization-specific behaviors (shaded either black or gray, to match the organizations). We model the spread of membership as a competitive diffusion process, in which network neighbors influence the composition of each other's behavioral profile, and agents gain membership when surpassing an organization's minimum demand and they lose membership when falling below this requirement. In each iteration, members develop ties with co-members at an organization-defined rate (i.e., social encapsulation rate). In the following sections, we formalize this process and describe the simulation algorithm in detail.

[7] While we present results from a model that specifies this socializing process, we tested an alternative specification where membership diffused directly, based on the logic of complex contagion (Centola and Macy 2007). Results were highly similar across the two specifications.

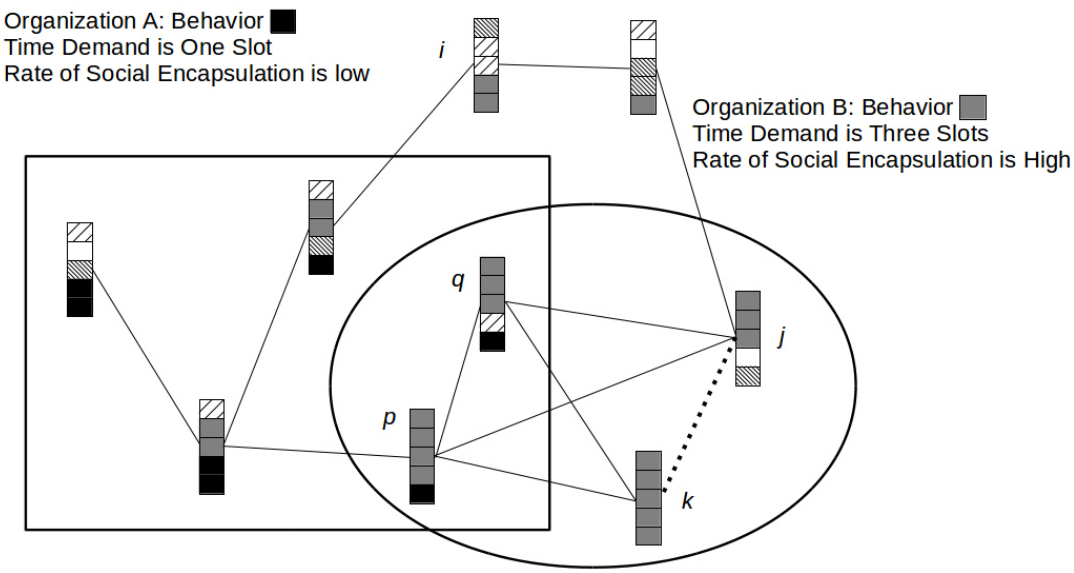


Figure 2.2. A hypothetical system of two organizations with the underlying social network. Cell shape in time slots represents behavioral commitment. Dotted line indicates that the tie is likely to develop in subsequent rounds of the simulation.

Formalizing Ecological Constraints

Individual agents face two constraints on their interaction with others: (1) the limited time and energy allocated to different behaviors, and (2) the limited capacity to form and maintain social ties with others. We model the first ecological constraint, the time and energy budget, as an array of length K, where each cell represents a discrete time slot. An agent can fill the time slots with different organization-specific behaviors, which are considered unique and representative of the type of participation that an organization requires. For example, agents in Figure 2.2 have K=5 slots for behaviors with each pattern or shade representing a different organization-specific behavior. The same organizational behavior can be assigned to multiple slots in the array, indicating

a greater investment in that behavior.[8] In each iteration of the simulation, agents evaluate the behaviors of network neighbors and adopt one behavior based on an urn model (detailed below). The behavior adopted from neighbors is assigned to a randomly chosen time slot, replacing the existing behavior. In successive rounds of the simulation, agents may adopt some behaviors and drop others. Therefore, at any given time, the portion of the total time and energy budget that an agent devotes to a specific organizational behavior signifies her level of participation in that organization.

Each agent also has a limited sociability budget, *T*, which constrains the total amount of attention that she can invest in social ties. We permit tie strength to vary by allowing multiple units of attention to be allocated to a tie over subsequent rounds of a simulation. Investment in social ties is reciprocal—i.e., when A invests one unit of attention in a relationship with agent B, B also invests a unit of her attention in the tie. Any attention units not currently in use are stored in an agent's reservoir and can be allocated in subsequent interactions. When the sociability budget is used up, the focal agent must reduce the strength of an existing tie.[9] By altering the composition and strength of an agent's social ties, it (stochastically) determines which individuals she is able to influence and the intensity of the influence.

---

[8] For simplicity, we refer to slots as containing only a single organizational behavior, but in reality this "behavior" is a complex of related behaviors. For example, someone who devotes one slot to "the Catholic Church" engages in a whole complex of actions related to membership in the Church. Devoting more slots simply means that these behaviors are engaged in for longer periods.

[9] The reallocation of units will not ripple in the chain of connected agents, because when the alter agent reduces a unit of attention due to the ego's retreat, she will put that unit in the spare reservoir, rather than allocate it to one of her social ties.

Interaction and Influence Dynamics

The ABM proceeds iteratively, where each round of the simulation has three steps: (1) a step in which membership status is updated, (2) a tie formation step, and (3) an influence step. How these steps are invoked is determined by the specific rates set in different simulation conditions. At the beginning of each round, all agents update their membership statuses based on their current level of investment in each organizational behavior. If their investment meets or exceeds the minimum demanded by the organization, they are considered members. For instance in Figure 2.2, if agent *i* were to adopt a gray behavior from one of her neighbors, she would meet the minimal demand of organization B and be considered a member at the start of the next round. Next, in the tie formation step, the focal agent allocates units of attention to co-members of each of her organizations at the respective organization-specific social encapsulation rates (see below).

Finally, we specify social influence using an Urn Model (Johnson and Kotz 1977; DellaPosta, Shi and Macy 2015). When the influence mechanism is invoked, the focal agent takes the behaviors from each of his neighbors' time arrays, multiplies these sets of behaviors by the tie strength of each respective alter, and places them all into a pool or "urn". After the urn is filled, the focal agent randomly chooses one behavior and places it into a random slot of his own array (replacing an existing behavior if necessary). The urn model is similar to continuous social influence models (e.g. Friedkin and Johnsen 1999), in that it specifies influence as an incremental adjustment to the behavior/opinion distribution in a weighted network. An advantage of the Urn Model is that it allows distinct behaviors to compete for the limited space in

the time array, which makes it suitable for modeling social influence in a competitive ecology.[10]

Organizational Strategies

Organizational strategies are defined by two dimensions, which are *time and energy demand* and *social encapsulation*. The strategies are fixed for the duration of the simulation. Time and energy demand (*D*) is defined as the minimum time slots that are required by the specific organization for membership and vary from 1 to *K*. In Figure 2.2, organization A demands one slot and organization B demands three slots of behavioral commitment for membership. Only those individuals who meet the minimum membership requirement are considered members and subject to the network constraints imposed by the organization's social encapsulation. Social encapsulation rate (*rE*) defines the organization-specific rate at which members allocate their units of attention to other members. Higher levels of encapsulation indicate a greater tendency for organizational members to allocate attention to other members. For instance, organization B in Figure 2.2 has a relatively high social encapsulation rate, which means that there is a high likelihood that agents *j* and *k* will develop a tie over subsequent iteration (the dotted line).

[10] The behavioral array may remind some readers of the commonly used genetic algorithm in agent-based models (see Chattoe 1998 for a useful discussion). The key difference is in the method of transmission. In genetic algorithms the behavioral profile is based on whether it improves evolutionary fitness, whereas in the Urn model transmission operates through simple social influence.

Simulation Setup

The simulation is set up with a population of 200 agents, each of whom has a time and energy budget of K=20 units and a sociability budget of T=100 units. In the initialization period, 10 organizational behaviors are distributed among agents in a controlled fashion (the details and a robustness analysis of the initialization process can be found in the Online Supplement). After the initialization period, the mechanisms defined by organizational strategies are activated at specified rates.

To summarize, in each iteration, an agent is subject to three distinct processes: (1) membership assignment based on the current behavioral profile; (2) network rewiring based on the rate of social encapsulation for each organizational membership;[11] and (3) social influence from network neighbors.[12] Table 1 provides a step-by-step outline of the algorithm.

Table 2.1. Summary of Simulation Steps

Initial Conditions:
200 actors
Time and energy budget (K) = 20 units
Socialization budget (T) = 100 units
10 organizational behaviors seeded randomly
Initialization period (see the online supplements for details)::
Network rewiring via triadic closure and homophily rules
Allow behaviors to spread until 95% of slots filled
Iteration steps:
Social influence

[11] In a robustness analysis, we include additional network formation mechanisms in this step (homophily and triadic closure). The analysis, presented in the supplementary materials, demonstrates that our results are robust for wide ranges of homophily and triadic closure rates.

[12] The rate of social influence is set at 0.01. A robustness analysis can be found in the supplementary materials.

At each iteration for each agent at a pre-specified influence rate:
Fill urn with behaviors of network neighbors: Randomly pick a behavior from the urn and (re)place it into a random time slot.

Network rewiring
At each iteration for each agent:*
For each organization agent is member of:
Allocate unit of attention to member of organization at rate *rE*
Membership assignment
Re-calculate organizational memberships

Results

We present the results as follows. First, we examine organizational outcomes for a baseline environment to assess how social encapsulation influences organizational growth and survival for organizations at different levels of time and energy demand. Then we look more closely at the composition of membership in four specific conditions, varying the level of social encapsulation in the focal organization. These "virtual case studies" (Baldassarri and Bearman 2007) reveal important qualitative differences between patterns of growth and decline of organizations pursuing different strategies. Next, we assess the influence of different competitive environments on the pattern of organizational growth and decline. By varying competing organizations' time and energy demand (high vs. low) and social encapsulation, we examine how the type of competitor in the environment influences the level of competitive pressure and identify strategies that fare best under particular environmental conditions.

Social Encapsulation and Membership Growth

We first present results from a baseline-environment model where competitor organizations are randomly assigned combinations of time demand and encapsulation

rate from the entire strategy space. We limit the presentation of results to a strategy space with ranges 1 to 14 for time and energy demand and 0 to 2.0 (on logarithmic scale) for social encapsulation rate.[13] Extension to other areas of the strategy space yields converging outcomes. Of primary interest here is the effect of social encapsulation on the rate of membership growth or decline. We additionally examine how this effect changes for organizations that demand different amounts of time and energy.

We terminate each simulation after 20,000 iterations.[14] We examine two outcomes: membership growth and organizational longevity. Membership growth is reported as the log-transformed rate of membership change over the simulation period. Specifically,

$$g = log_{10}\left(mean\left(\frac{S_{t=20,000}}{S_{t=0}}\right)\right)$$

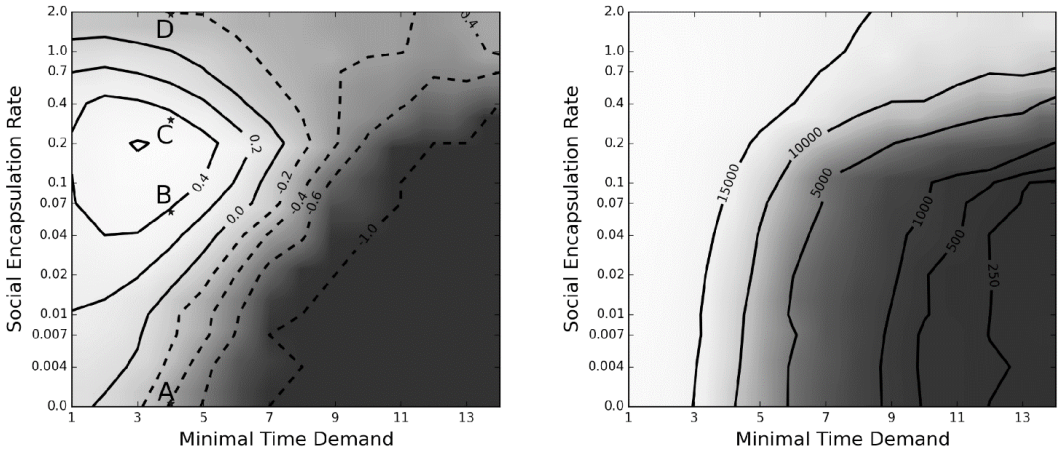
where S_t is the size of membership at iteration t. Many organizations did not survive

[13] A demand of 4 time slots means that the organization requires 20% (4/20) of the total time devoted to the organization-specific behavior. A social encapsulation rate of 2.0 means that in each iteration the agent reallocates two unit of social time (out of 100 units) to co-members.

[14] We settled on this termination time after a close examination of the typical time to a stable outcome. Whether the simulations reach perfect equilibrium is secondary to our question of interest, which is the pattern of membership change over time. The 20,000 iteration span of the simulation can be thought of as representing a period of a few years in natural time. In this span of time, agents may change their affiliations, causing some organizations to grow and others to decline, but this period of time is short enough that the effects of demographic change will be minimal. For every 100-iteration cycle, for example, an agent will be influenced to change one slot of behavior once. Thus, if each 100-iteration cycle represents a week, the simulation lasts about 4 years. In the robustness analyses (presented in supplementary materials), we vary the social influence rate and find consistent results.

the 20,000 iterations of the simulation period; thus, we additionally report the log-transformed longevity of each organizational strategy. For both measures, we report averages of 1,000 runs of the simulation.

Figure 2.3 displays contour plots of strategy outcomes measured as (a) growth and (b) longevity, as an effective way to present complex, three dimensional data. The map surface corresponds to the strategy space, with social encapsulation rate on the Y-axis and time and energy demand on the X-axis. The shade corresponds to the relative outcome (growth or longevity) for an organization occupying that position in the strategy space, with lighter shades indicating higher values. Lines offer an additional guide for the reader. Solid lines circumscribe regions with positive values (growth or longevity) and dashed lines circumscribe negative values (decline). Lines are labeled with their respective growth or longevity values.



(a) (b)

Figure 2.3. Contour Maps of Membership Growth (a) and Organizational Longevity (b). Light shade in (a) indicates membership growth and dark shade indicates decline. Solid contour lines indicate growth with the rates embedded in the lines. Dashed contour lines indicate decline in membership. In (b), lighter shades indicate greater organizational longevity

Our first observation is that organizations that do not rewire the ties of their members to one another (i.e., rE=0, bottom of Figure 2.3a) fared poorly. The contour line distinguishing growth and decline (g=0) bisects the strategy space at the origin, indicating that all non-rewiring strategies resulted in decline. Therefore, while an increased rate of in-group interaction is often considered a natural byproduct of most social groups (Feld 1981), we find that it is also essential to organizational survival. Organizations that did not compel members to socialize with each other invariably lost members over the course of the simulation. In such cases, members invested in closer ties with non-member associates instead, which lead to disengagement from the focal organization.

While the complete absence of social encapsulation is invariably a losing strategy, in general the level of encapsulation that is beneficial for membership growth varies significantly by the amount of time and energy that the organization demands of its members. Concentration of darker shades on the right side of the contour plot in Figure 2.3a indicates that highly demanding organizations have a difficult time achieving positive growth. Regardless of their social encapsulation rate, organizations demanding more than about 40 percent of members' time and energy lose members over the long run.

Results from our model, however, support findings from classic case studies of "greedy" organizations (Coser 1974; Kanter 1972): highly demanding organizations can maintain members' commitment for a long time by monopolizing their social networks and cutting them off from outside contact. Although all demanding organizations have difficulty attaining growth, Figure 2.3b shows that demanding

organizations with higher rates of social encapsulation persist longer than those with lower rates. This effect is reflected in the lighter shades of gray in the upper right corner of the plot, compared to darker shades in the bottom right corner. An important side effect of pursuing a strategy of high social encapsulation, however, is that in cutting itself off from the outside world the organization leaves itself no means for future growth. Ultimately the strategy leads to a prolonged demise, and would be even more likely to do so in a real-world setting where individual members can die, as well as leave the organization.

Social encapsulation is important for less demanding organizations as well, but results in markedly different outcomes. Figure 2.3a shows a curvilinear effect of encapsulation rate on membership growth for organizations that demand less than 35 percent of members' time. The highest and lowest levels of encapsulation result in membership decline, while intermediate levels achieve positive growth. The observed pattern supports the idea that organizations face countervailing pressures from, on the one hand, the need to maintain open networks to enable recruitment and, on the other hand, the need for social encapsulation to enhance retention. Too much investment in either structural property results in decline. Nonetheless, the less demanding organizations can achieve stability and even growth with a relatively wide range of different network encapsulation strategies.

Virtual Case Studies of Four Social Encapsulation Strategies

The contour plots (Figure 2.3) reveal a general pattern of growth and decline for different strategy combinations, but they also mask potentially important differences

in membership dynamics. One of our key questions is how different social encapsulation strategies balance recruitment and retention. For example, an organization can maintain a stable membership either by retaining all existing members or by replacing every member they lose with a new recruit. What does this balance look like in organizations that achieve membership growth and how does it compare to organizations that fail to survive? The question has important substantive implications, because even when they achieve identical growth rates, organizations with high membership turnover (high recruitment relative to retention) will support very different competencies than organizations with low turnover (high retention relative to recruitment).

We now take a closer look at the effects of the social encapsulation rate by examining the network structures and unique retention and recruitment signatures of four specific strategies in "virtual case study" (Baldassarri and Bearman 2007). To control for effects of time demand, we chose four organizations that each demanded 20% (4 slots) of their members' time and energy. We also held the initial conditions identical for the four simulations using fixed random seed. Thus, any difference in outcome should be attributed to the simulation parameters. Within the 20% time demand stratum, we selected (A) an organization that did not rewire members' ties to one another at all (rE=0), (B) an organization that induced members to rewire ties to one another at a low rate of 0.05, (C) an organization with moderate encapsulation rate (rE=0.3), and (D) an organization with high encapsulation rate (rE=2.0). The location of each case within the strategy space is marked with its respective letter in the contour plot of Figure 2.3a. The four cases represent the range of the curvilinear

pattern observed in Figure 2.3a. Two cases that decline in membership (A and D) represent opposite extremes of social encapsulation, while B and C come from the opposite sides of the region with favorable prospects for membership growth. These two pairs of cases provide useful comparisons. While the final outcomes of each pair are similar, the underlying patterns of growth or decline are qualitatively different.

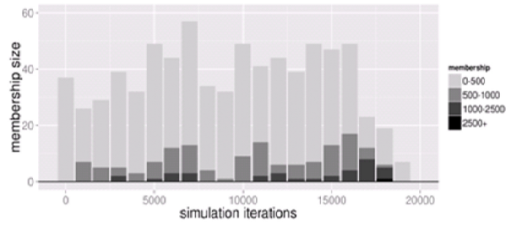
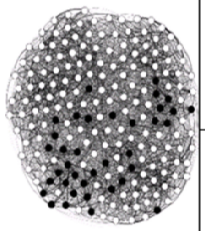
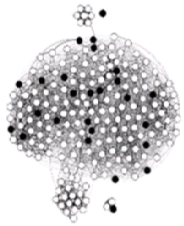
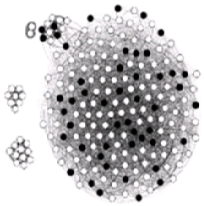
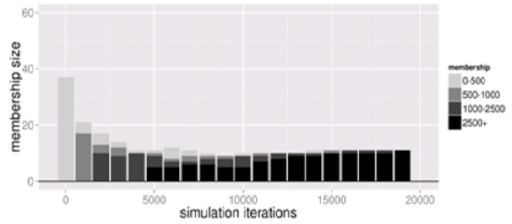
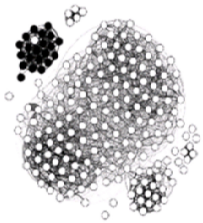
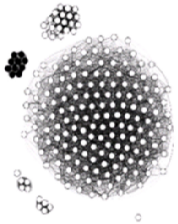
S	Time Series	$t = 0$	$t = 1{,}000$	$t = 10{,}000$
A				
D				

Figure 2.4. Time series plots and network diagrams for two strategies leading to membership decline.
Black dots in network graph diagrams represents members of the focal organization; white dots represent non-members. In time series graph, each column represents a period of 1,000 iterations with its height corresponding to the membership size at the end of the period. The lightest shade in each bar designates the newest recruits (those who only became members during the previous period), a slightly darker shade designates members who joined between 500 and 1,000 iterations before the current period, next darkest shade are those members who joined the organization between 1,000 and 2,500 iterations prior, and the darkest shades are the longest-tenured members, with more than 2,500 iterations of membership.

Let us first compare the two organizations that lose members in the long run (A and D). In Figure 2.4, we present a time-series graph of membership for organizations A and D. Each bar is decomposed to show how long members have held their membership in the organization (see Figure 2.4 caption for details). This decomposition reflects an organization's propensity to recruit and retain. Graphs dominated by light shades of gray indicate organizations with many newcomers and a high level of membership turnover. While graphs with darker shades indicate low turnover.

The time-series graphs indicate that, while organizations A and D finish the simulation run with approximately the same number of members, they decline in very different ways. Organization A maintains high membership turnover. Throughout its lifespan, the majority of its membership is made up of recent recruits (members with tenure of less than 500 iterations). The growth trend is also highly volatile. Organization A increased its initial size by nearly 50%, before declining again to just 18% of its initial size by the end of the simulation. Organization D has a very different composition of members. In contrast to A, organization D has many longtime members and fewer new recruits. The core of long-term members, moreover, appears to protect the organization from high levels of volatility.

To further contrast these two patterns of membership decline, we present three snapshots of the system's networks. The first snapshot (at t=0) represents the initial conditions for both cases. The other two snapshots are from different stages of the simulation: one from an early state (at 1,000 iterations) and one from a comparatively mature state of the system (at 10,000 iterations). In the network diagrams, the black nodes indicate members of the focal organization. Edges between nodes indicate social ties. The network graph in Figure 2.4 reveals that organization A, which does not rewire its members' ties, is not differentiated from the rest of the system. The clustering of members at the beginning of the simulation (t=0) dissipates by iteration 1,000. At iteration 10,000, the E-I index, which defines the proportion of external to internal ties (Krackhardt and Stern 1988), is 0.69.[15] By contrast, members of organization D cluster closely together. By iteration 10,000, they form a close-knit cluster that is completely isolated from the rest of the network (E-I=-1.00).

The network diagrams help explain the pattern observed in the time-series graph. By aggressively rewiring members' ties to each other, organization D is able to protect members from out-recruitment much better than organization A. However, in doing so the organization also inhibits its ability to recruit new members. Organization D slows its decline and survives the simulation period, albeit with a much smaller membership than it started. On the other hand, organization A's members are well connected to non-members in the system. The organization uses these ties to recruit new members, but the ties also make members available for out-recruitment. As there

[15] The E-I index is 1 if all the ties from the organization are directed to non-members and it is -1 if all the ties are direct to co-members.

is no cohesive set of members to serve as a buffer, membership begins to decline.

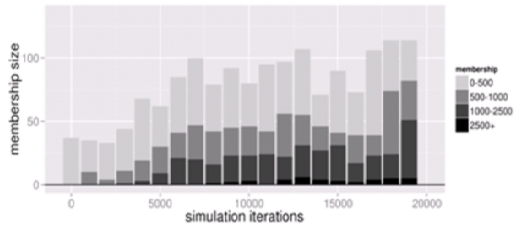
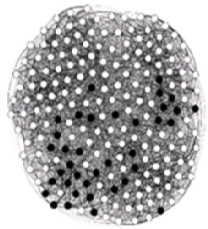
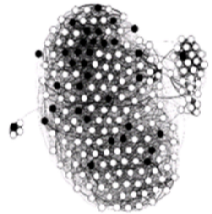
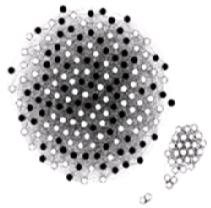
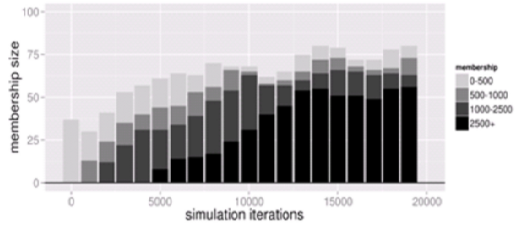
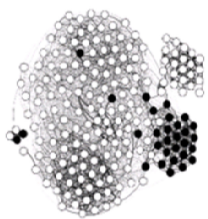
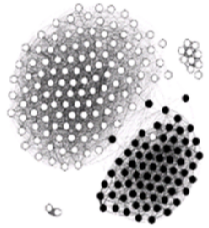
S	Time Series	$t = 0$	$t = 1{,}000$	$t = 10{,}000$
B				
C				

Figure 2.5. Time series plots and network diagrams for two strategies leading to membership growth.
Black dots in network graph diagrams represents members of the focal organization; white dots represent non-members. In time series graph, each column represents a period of 1,000 iterations with its height corresponding to the membership size at the end of the period. The lightest shade in each bar designates the newest recruits, a slightly darker shade designates members who joined between 500 and 1,000 iterations before the current period, next darkest shade are those members who joined the organization between 1,000 and 2,500 iterations prior, and the darkest shades are the longest-tenured members, with more than 2,500 iterations of membership.

The same mechanisms are observed in the two cases of organizational growth (B and C), but these organizations strike a better balance between the structural demands of retention and recruitment. Qualitative differences emerge between these cases as well. In Figure 2.5, we present the membership time-series graphs of the two growing organizations (B and C). The two organizations achieve comparable rates of growth, but whereas the membership of organization B includes many newcomers and few "old timers", organization C's members have much longer tenure on average. Additionally, while organization B's growth comes in bursts, often interrupted by brief periods of decline, little volatility is observed for organization C. Members of organization B are mixed with the rest of the population, an observation illustrated by the network diagrams (E-I=0.39). On the other hand, organization C grows steadily, integrating its recruits into the organizational core and resulting in a significant cleavage between the older and newer members (E-I=-0.80)

Several insights can be derived from the preceding analysis. First, organizations require different network structures to support different membership demands. On balance, social ties that span boundaries of highly demanding organizations serve to facilitate exit and not recruitment. Consequently, for the most

demanding organizations the only effective strategy is the one recommended by Coser (1974): to sever members' ties to the outside world. The trade-off between recruitment and retention is most visible in less demanding organizations, which achieve positive growth only when they strike an appropriate balance between in-group cohesion and out-group ties to potential recruits. But even among organizations that achieve positive growth, different strategies imply contrasting membership structures. Organizations that allow members to maintain a significant share of ties to outsiders have very high membership turnover. By contrast, organizations with high in-group cohesion have very low turnover.

Thus, the level of social encapsulation influences not only growth and survival, but also carries important substantive implications. For example, maintaining a core of committed, long-tenured members makes it easier to preserve institutional memory and maintain a particular organizational culture (Harrison and Carroll 1991). High member turnover, by contrast, is not very amenable to reproducing a stable organizational culture. On the other hand, the constant recycling of members may facilitate greater adaptability and avoid the downsides of institutional inertia.

Competitive Pressure in Local Environment

Next, we consider how environments composed of different types of competitors influence the fitness of focal strategies. Figure 2.6 presents growth contour plots for six environments characterized by different types of competitors. In the top panel, competing organizations are highly demanding of members' time, while in the bottom panel competing organizations demand comparatively little time. Specifically, we

define organizations that demand between 40 and 80 percent of members' time and energy as highly demanding, while organizations in the low-demanding environment require between 5 and 15 percent of individuals' time and energy. A main contrast between these two environments is that low-demanding organizations make it easier for agents to hold multiple affiliations. We additionally vary the level of competitors' social encapsulation rates across environmental conditions. In the left panel competing organizations do not rewire ties (*rE*=0), in the next panel the encapsulation rate is kept moderate (0.01-0.1), and the right panel is extremely high (10.0).

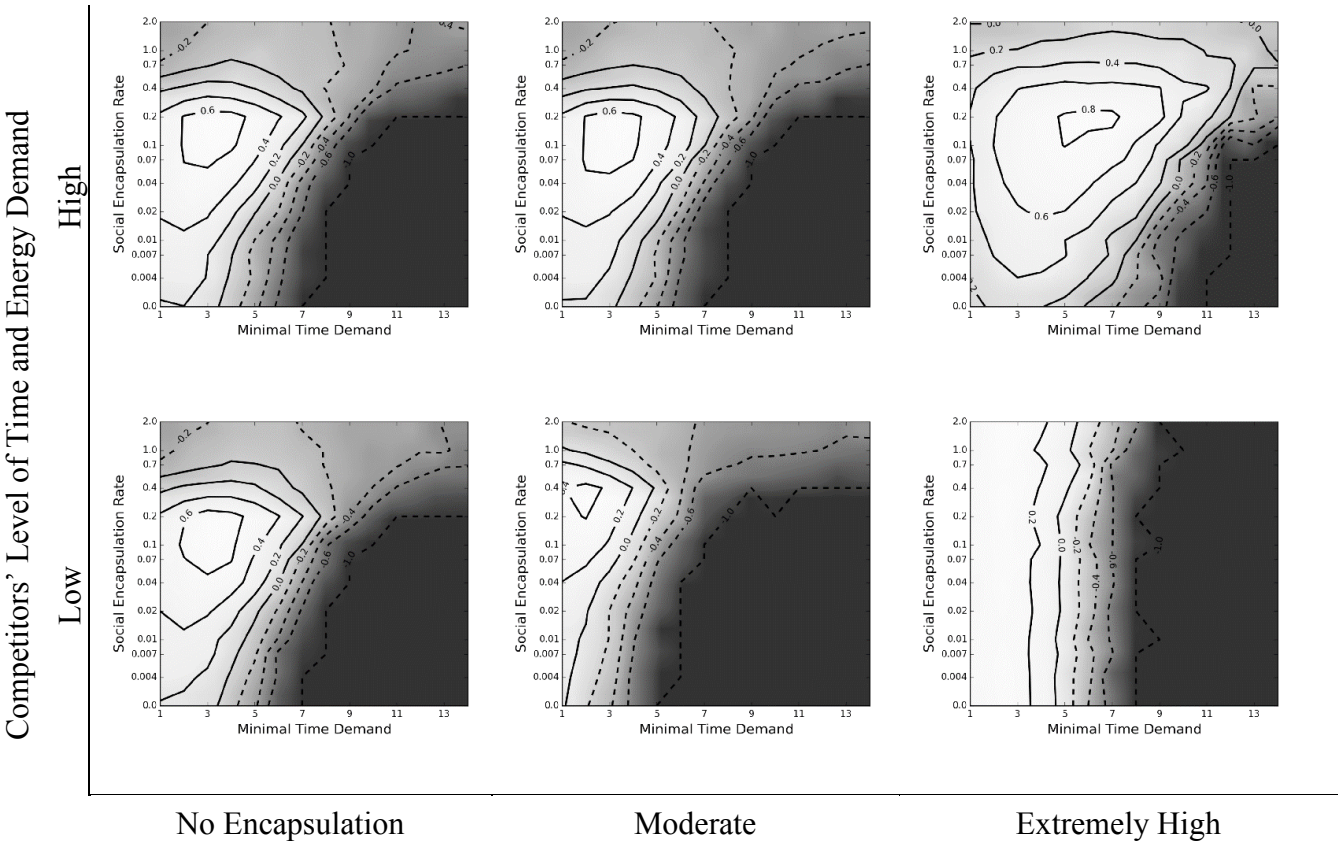


Figure 2.6. Membership growth patterns of the focal organization in competition with organizations of high and low demand and of varying social encapsulation level.

Results confirm the expectation that competitive pressure is generally higher in environments with low-demanding competitors. The comparatively larger areas of the strategy space that lie within the solid contour lines (positive growth) in the top panel indicate that a broad range of focal strategies can achieve growth in an environment rich with highly demanding organizations. In contrast, environments characterized by low-demanding organizations (bottom panel of Figure 2.6) constrain the focal organization's growth strategies to a much smaller area. In other words, when the competition is highly demanding, a focal organization can grow while adopting almost any strategy, but when the competition is less demanding, can only grow if it keeps to a relatively narrow range of strategies. To examine this pattern directly we construct a measure of competitive pressure within different environments by counting the number of strategy combinations that decline under different conditions and dividing it by the total number of strategies in the strategy space. For instance, when every possible strategy leads to decline in membership, competitive pressure is highest—every one of the 13×14 sampled strategy combinations fails ($\frac{13\times14}{13\times14} = 1$). When all strategies survive, competitive pressure is zero ($\frac{0}{13\times14} = 0$). We plot this measure of competitive pressure against social encapsulation for high and low-demanding organizations in Figure 2.7.

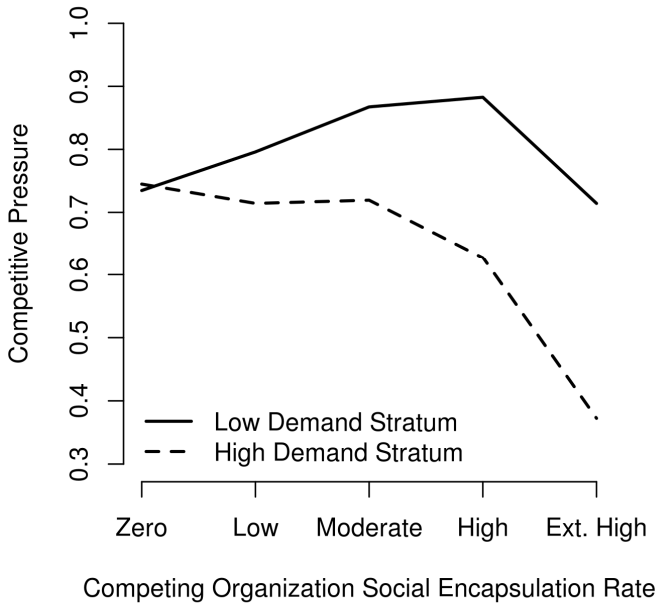


Figure 2.7. Competitive Pressure in High vs. Low Demand Strata.

Across all but one social encapsulation condition, competitive pressure is higher in environments with low-demanding organizations. The only exception is the case where competitors' social encapsulation is set to zero, in which competitive pressure is equal for both low and high demanding conditions. Since passing the membership threshold has no influence on tie formation in this case, the exact threshold value makes no difference for affiliation dynamics. The patterns diverge as the rate of social encapsulation among competitors increases. In the condition of highly demanding competitors, competitive pressure generally declines with higher rewiring rates. At the highest level of rewiring, competitive pressure is at its lowest. In the condition with low-demanding competitors, by contrast, competitive pressure increases steeply with higher social encapsulation rates and then decreases again as

competitors become too self-isolating.

To understand the observed effect of competitors' strategies on the focal organization's growth we need to consider the niche dynamics that emerge in each environmental condition. Multiple affiliations, which are more likely within the low-demanding environment, drastically increase competition. They integrate the social space by creating cross-cutting social ties and enable organizations to compete directly for each other's members.[16] In contrast, in environments characterized by high-demanding competitors, each is likely to hold just one affiliation. Direct competition between organizations is lower in this case as each organization settles into a relatively circumscribed niche.

Competitors' rate of social encapsulation also has an important effect on niche dynamics. In the high-demand condition, where cross-cutting ties are absent, higher rates of social encapsulation lead to greater fragmentation of the social space. Competitive pressure decreases as organizations isolate their members (Figure 2.7, dashed line). Social encapsulation has a different relationship with competitive pressure in the low-demand condition, where multiple affiliations are the norm. The focal organization's members are subject to competing network rewiring pressures from their extra affiliations. These affiliations pull them away from the focal organization, with the force of the pull proportionate to the social encapsulation rate of the competing organization. The graph in Figure 2.7 (solid line) shows that as

[16] An interesting implication is that intense competition among multiple organizations also gives rise to a class of organizational "omnivores" (e.g., Lizardo 2006)—agents whose behavioral profile reflects an interest in multiple organizations but who do not actually meet membership requirements of a single organization.

competitors' rate of social encapsulation increases from zero, competing organizations win these tug-of-war contests more often, thus imposing greater constraint on the focal organization's growth. Competitive pressure declines again, however, when social encapsulation rates reach extremely high levels. In this environment, also presented in the bottom right contour plot of Figure 2.6, competing organizations retreat from competition with the focal organization. Extreme values of social encapsulation eliminate the cross-cutting ties and give rise to a fragmented social space—one in which the focal organization can carve out a larger, less competitive niche.

In general, this analysis reinforces the core idea of ecological approaches to the study of organizations: the fitness of an organizational strategy is always conditional on the local environment. We extend previous ecology of affiliation models (e.g., McPherson 1983) by showing that by adopting different strategies organizations alter the local network structure, thereby creating more or less competitive niches.

Extensions to Other Types of Ecological Relationships

In order to reduce the complexity of an already high-dimensional model, we have thus far followed the convention in the affiliation ecology tradition (e.g., McPherson 1983), assuming that time and energy is an undifferentiated resource. The model based on this assumption describes a broad class of situations in which organizations are competing for the same segment of time or cases when scheduling permits flexibility. However, this is also somewhat restrictive, because different types of organizations may schedule their activities at different times of the day and this partitioning of time may be consequential for affiliation processes (Winship 2009; Young and Lim 2014).

In particular, a differentiated resource base may provide a foundation for non-oppositional types of ecological relationships (e.g., Kitts 1999). Here, we offer an extension of our model that relaxes the assumption that time are fungible resources. An exhaustive exploration of the extended model is beyond the scope of the present article; rather, our intention is simply to draw attention to the possibility of mutualistic as well as competing relationships.

We relax the assumption that time and energy is fungible by modeling competition for time in a bifurcated schedule. This can be thought of as accommodating two classes of organizations; one class of organizations that schedules its activities during the daytime and another class that competes exclusively for individuals' evening hours. The interaction and influence dynamics are equivalent to the above model. The only difference is that we divide the twenty-slot time array into two classes (e.g., daytime and evening), each represented by a ten-slot array. We designate five of the organizations as competitors within one class (e.g., daytime organizations) and five as competitors within the other class (e.g., evening organizations). Each organization's strategies are assigned randomly as in the baseline condition above.

The questions we pose are whether and how the social encapsulation of a daytime focal organization influences its members' affiliations in the evening segment. We expect that, holding everything else in the system constant, increasing the social encapsulation rate of the daytime organization should concentrate its members' affiliations among the evening organizations. We measure this concentration by defining a heterogeneity index of affiliation (*H*). For the set of

members in the focal daytime organization, H is the mean of variances of membership status (1 or 0) in each of the evening organizations, weighted by the size of the overlap[17]. We present the results in Figure 2.8, where the X-axis is the encapsulation rate of the daytime focal organization and the Y-axis is the heterogeneity of evening affiliation at iteration 1,000, averaged over 10,000 replications.

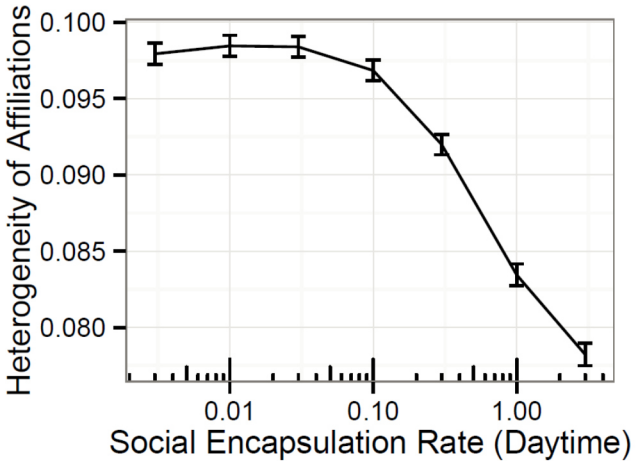


Figure 2.8. Heterogeneity of Evening Affiliations for the members in the focal daytime organization.
Bars indicate a 95% confidence interval.

The results confirm our expectation that as the social encapsulation rate of the focal daytime organization increases, its members' affiliations in the evening become more homogenous. In other words, as daytime social encapsulation becomes more pronounced, the nighttime activities of members converge on a limited number of organizations. More generally, the findings illustrate that when the resource base is partitioned, the social ties formed in an organization occupying one partition will drive

[17] For a formal definition of the index, see the supplementary materials.

the same set of members to share affiliations in a different partition. We presented the partition as reflecting two non-overlapping time segments, but it could alternatively be interpreted as a distinction between different organizational domains, such as religious and non-religious activities (e.g., Chattoe 2006). Our result thus provides theoretical support for studies that identify mutualistic relationships between organizations that occupy alternative domains of social activity. Kitts (1999), for example, observed that members of an environmental organization expressed less commitment if they held affiliations with other environmental groups, but they were more committed if those extra affiliations were with organizations in an unrelated domain. This result also provides an alternative explanation for Feld's (1981) findings that foci are often merged; it may be less that individuals want foci to be connected, than that they are simply constrained by their networks to do so.

Discussion and Conclusion

Social network ties are instrumental for recruiting and retaining members in organizations. Meanwhile, organizations themselves act as the primary contexts of network tie formation (Feld 1981; Putnam 2000). Combining these two insights, this paper examines how different networks form inside of organizations and influence the organizations' ability to recruit and retain members. The results show that basic human constraints—limited time and energy and limited capacity for socialization—induce trade-offs between an organization's abilities to recruit and retain. How organizations manage these trade-offs has important implications for affiliation dynamics and for network processes more generally.

Our research yields several recommendations for organizational design. The results confirm the prediction of the affiliational ecology model that higher participation requirements lead to member exit (Cress, McPherson, and Rotolo 1997). However, we find that a corresponding increase in the rate of social encapsulation can mitigate the losses from increased demand. This result supports Coser's (1974) idea that "greedy institutions", which sever members' ties to the outside world, help secure commitment in extremely demanding organizations (e.g., cults and communes). Although transforming into a greedy organization offers protection from competitors, it does so by sacrificing the prospect of future recruitment, all but guaranteeing an organization's eventual demise. Members of organizations with high levels of network closure eventually die and, lacking outside ties to facilitate recruitment, the organizations face the same fate. Thus, if an organization is going to grow and/or survive in the long run, it ultimately must compete, and that competition will force behavioral demands down, limiting the level of contribution an organization can expect from each member.

To attain a large volume of members quickly, organizations should lower the time and energy required for participation. Our results show that in most environments such organizations can survive and even grow with relatively little network closure. The result helps explain, for example, the documented trend among American civic organizations toward reliance on "checkbook members" who contribute little to the organization's day-to-day activities (Putnam 1995; Skocpol 2003). Existing empirical research suggests that lowering the cost of participation is a conscious and popular strategy for increasing public engagement among advocacy organizations (Walker

2014). New Internet-based advocacy organizations, in particular, have grown by making it possible for members to make only token contributions with a click of a mouse (Lewis *et al*. 2014).

Of course, most organizations require members to contribute more than token amounts of participation. The need for balance between open networks (for recruitment) and network closure (for retention) is most clear in these cases. Whereas token contributions can be sustained with relatively little network closure among members, ensuring commitment to a moderate level of participation requires reinforcement from network neighbors. Thus, while Putnam (2000) famously observed that bowling leagues help to sustain social capital (i.e., a high level of network connectedness), the opposite is true as well: bowling leagues can only survive so long as they develop sufficiently dense networks among their members.

The level of social encapsulation that would be sufficient for survival depends on the surrounding organizational environment. As a general rule, a rate of social encapsulation that is higher than that of one competitors' provides an ecological advantage. However, among organizations that achieve stability or growth, the degree of social encapsulation accounts for substantive variation. For example, organizations that need to develop and maintain complex organizational routines should favor higher levels of social encapsulation, which produces denser networks and thus support longer average membership tenure. In contrast, organizations that value adaptability and wish to limit institutional inertia will prefer lower levels of social encapsulation, because it produces a higher rate of membership turnover. Additionally, while imposing greater social encapsulation than one's competitors tends to provide an

ecological advantage, there is, for any given level of time and energy demand, a cliff beyond which more social encapsulation leads to a lingering organizational death. Thus organizations may be driven to crowd as close as possible to this point-of-no-return, without quite crossing over.

Finally, the initial step we took to relax the assumption that time and energy is an undifferentiated resource suggests a new line of inquiry for research in the affiliational ecology tradition. The small extension described above offers an existence proof of mutualistic relationships between organizations that occupy separate time segments. This has important implications for ecological theory, suggesting that under some circumstances multiple embeddings are not channels of competition, contrary to previous scholarship (but see Kitts 1999 for a notable exception). This points intriguingly towards an explanation for the symbiosis between sustained political movements and religious groups (e.g., the civil rights movement): high levels of network closure obtained within religious organizations supports sustained encapsulation in associated political organizations, even without the need for formal ties (McAdam 1982; Morris 1984). To improve our understanding of these important dynamics, future research should develop conceptions of time and scheduling that better reflect the routines of contemporary organizations (Winship 2009; Young and Lim 2014). Not only do organizations schedule their activities at different times, the flexibility of schedules also differs across organizations. One may expect that more flexible organizations have an ecological advantage due to greater adaptability to competition; although, flexibility may also degrade the effectiveness of social encapsulation, because members are less likely to come into contact (i.e., the

coordination problem).

Like every model, ours rests on multiple simplifying assumptions. Several specific assumptions are worth elaborating both to clarify how model results relate to empirical cases and to highlight fruitful directions for future research. First, although there is likely a relationship between social encapsulation and time and energy demand, we model them as independent parameters. This modeling choice serves an analytical purpose, allowing us to explore the entire strategy space without committing to a poorly specified link function between the two dimensions, and thereby enabling us to examine the mechanisms underlying the ecological dynamics of interest. However, while the results cover all *theoretically* possible outcomes, some of the outcomes are *empirically* unlikely. A relationship between time demand and maximum social encapsulation constrains the empirical strategy space. We expect this relationship is sub-linear, with each additional unit of time yielding decreasing amounts of maximum encapsulation, but additional research is needed to understand the individual and organizational factors that influence this bound. Relatedly, in our model social encapsulation and time and energy demand are assumed to be independent of other organizational characteristics such as size and organizational age. Yet, one may reasonably expect, for example, that social encapsulation is more easily achieved in a small group, such as a book club, than a large group, such as a university. Future research should explore these relationships.

Second, in isolating the variables of interest, our model presents organizations that are more rigid than would be expected in empirical cases. For one, our model fixes organizational strategies for each organization for the duration of the simulation.

While organizations are limited in the extent and pace of their change in important ways (DiMaggio and Powell 1983; Hannan and Freeman 1984), they do adapt to external circumstances (Rochford 1982; Barnett and Carroll 1995) as well as to concerns raised by members and other stakeholders (Hirschman 1970). Additionally, whereas we model membership as a function of a single commitment threshold, many organizations have multiple types of membership. Oliver (1984), for example, contrasts the contributions of core members with token contributions by the majority of social movement participants. Although our simulation does not model such schemes directly, it makes the need for stratified membership levels clear. A highly demanding organizational core may need to be surrounded by a lower demanding layer that serves as a membrane, protecting the core from out-recruitment while simultaneously funneling resources and individuals into the organization. Previous research finds that token participation often leads to more substantial contribution in a variety of organizations (e.g., McAdam 1986). More generally, existing scholarship finds that organizations use multiple types of strategies depending on the organizational environment. Studies of religious movements, for example, find that the groups develop one set of expectations for devotees, while presenting more flexible routines to potential recruits (Snow *et al*. 1980; Rochford 1982).

Finally, we present a structural account of organizational membership and deemphasize alternative factors that are also related with group cohesion. Prior research suggests, for example, that organizations generating greater investment of time and energy from their members would produce higher quality collective goods that make them more attractive to additional members (e.g., Hechter 1987; Iannaccone

1994). Future research may extend the formal architecture we present in this paper to consider this and other factors.

Before closing, we note that our findings at the organizational level can improve our understanding of outcomes at the level of individuals and at the level of the larger social system. As primary contexts of interaction, organizations form a critical meso-level in society (Feld 1981; McPherson 1983; Putnam 2000). But despite their importance for research on both the composition of personal networks (at the micro-level) and properties of global networks (at the macro-level), organizations' role in network formation remains underspecified. For example, previous models of emergent networks consider the role of dyadic mechanisms, such as homophily and reciprocity, as well as neighborhood-based network processes (e.g., Pattison and Robins 2002; Centola 2015), while organizational diversity is either ignored or seen as causally secondary to other network processes (McPherson 2004). Our model shows that basic organizational features can generate substantial variation in global network structure. Thus, we wish to echo recent calls for greater attention to the social contexts of interaction (Entwisle *et al*. 2007; McFarland *et al*. 2014). Organizations also influence micro-level outcomes. As Breiger (1974) observed, the relationships among organizations and individuals are co-constitutive (see also Zweig and Kaufmann 2011; Neal 2014). Thus considering organizational dynamics can offer new insights into key individual level outcomes such as social isolation or its converse, integration. Our research suggests, for example, that lower levels of connection among Americans may, in part, be the result of a decline of organizations that, while effective at creating cohesion, could not compete with organizations that offered more flexible routines

(Putnam 2000; Skocpol 2003).

This study is a significant step toward a better understanding of organizational dynamics more generally. Simulation enabled us to identify general mechanisms of organizational growth and survival. We show, specifically, that social encapsulation and time and energy demand form critical elements of organizational strategy. Our study also opens avenues for future research. While we demonstrate that the broader organizational ecology provides essential context for understanding recruitment and retention processes, future research should build on our initial foray into examining non-competitive relationships. Finally, although the collection of longitudinal affiliation network data is costly and rarely done, we believe that further theoretical development would benefit from empirical research.

APPENDIX

Part 1. Initialization of the Simulation.

The initial distribution of behaviors in the population directly determines the initial size of the membership. If behaviors are randomly distributed in the agents' time slots then high-demanding organizations are least likely to form, and most likely to disband. This is because few agents can accumulate the sufficient volume of behaviors required by organizations for membership. On the other hand, low-demanding organizations can easily gain a large proportion of the population as their members, and easily saturate the population after the simulation starts. Thus, it becomes difficult to conduct a fair comparison of membership growth between organizations with different time and energy demand. We use the following two-step procedure to prevent bias in the results due to the initial behavioral distribution. First, we designed a behavior-propagation method to initialize the behavior distribution in the population so that both high-demanding and low-demanding organizations can maintain a sizeable membership (detailed below). Second, we varied the key parameter in the behavior-propagation method and tested the robustness of our results to varied initial distributions of behaviors.

Initially, 10 organizational behaviors are randomly seeded to 10 randomly chosen agents, while the behavior budgets of the rest of the agents remain empty. In each iteration of the initialization stage, an agent propagates her behavior to a neighbor with a probability proportional to $T(i,j)^z$, where $T(i,j)$ is the tie strength between i and j. $T(i,j)$ is proportional to the similarity of i and j's attributes (as detailed in section 3.2 below) and ranges from 0 to 1, where 0 means the agents do not share a tie and 1 is the strongest possible tie (i.e., both agents invest all their network attention to each other). z is a scaling factor ranging from 0 to infinity.

If z=0, then $T(i,j)^z = 1$, for all pairs (i,j) and i's behaviors are spread uniformly across the population. Under this condition, the initialization procedure generates a distribution of behaviors that is equivalent to random assignment. Highly demanding organizations are unlikely to form under random assignment, however, because few agents will accumulate a sufficient volume of the same behavior to meet these organizations' high membership demand. At the other extreme, when z is very high, the distribution of $T(i,j)^z$ becomes very skewed, which means that behaviors spread disproportionately through very strong ties and become highly concentrated in tight clusters of the social network. However, while this permits high demanding organizations to form, it also generates initial conditions that unfairly advantage low-demanding organizations. In particular, the initial membership size of each low-demanding organization appears small, but because the organizational behavior is highly concentrated among these members, the small initial size masks the relative prevalence of the behavior in the population. When the simulation starts, a simple dilution of the behavior will appear as significant membership growth. To avoid either

of these extremes and to set up a fair comparison, we use a moderate value of z ($z = 6$). This produces a balanced initial distribution of behaviors, allowing high-demanding organizations to form but retaining a significant level of randomness in the distribution of behaviors. The robustness analysis (presented in section 3.4 below) reports a consistent pattern of results for a wide range of z.

Part 2. Mutualistic Relationship of Organizations.

In this variant of the model, we relax the assumption that time and energy is fungible by modeling competition for time in a bifurcated schedule. This can be thought of as accommodating two classes of organizations. For example, one class of organizations that schedules its activities during the daytime and another class that competes exclusively for the individuals' evening hours. Compared to the main model, the difference is that we divide the twenty-slot time array into two classes (e.g., daytime and evening), each represented by a ten slot array. We designate five of the organizations as competitors within one class (e.g., daytime organizations) and five as competitors within the other class (e.g., evening organizations). Initially, we assigned the same evening organization membership for each member in the daytime focal organization. Therefore, when the simulation starts, members who belong to the daytime focal organization also belong to the evening focal organization. For simplicity, we set all the social encapsulation rates of evening organizations to zero, and their time demand to a low level (10%). We vary the rate of social encapsulation of a daytime focal organization and examine the effect this has on the growth of the evening focal organization that members share initially.

For the set of members in the focal daytime organization, H is the mean of variances of membership status (1 or 0) in each of the evening organizations, weighted by the size of the overlap. Denote X_D as the set of members of the focal daytime organization D. The membership status of each $x \in X_D$ in evening organization K is $A_K(x)$, which is 1 when x is a member and 0 when it is not. The size of overlap in membership between the daytime focal organization and an evening organization K is $|X_D \cap X_K|$. Thus the heterogeneity index of evening affiliation is:

$$H = \frac{\sum_K |X_D \cap X_K| \times Var(A_K(X_D))}{\sum_K |X_D \cap X_K|}$$

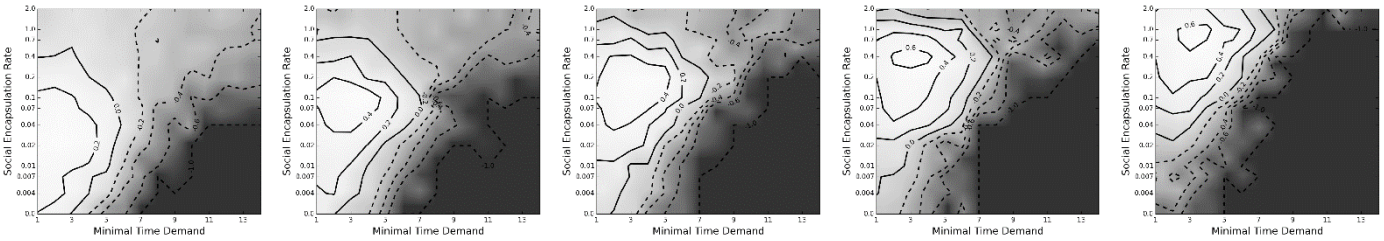
Part 3. Robustness Analysis.

For the robustness analysis, we vary four key parameters in the simulation. These include three parameters that govern interaction dynamics in the model—(1) the rate of social influence (*rI*), (2) the rate of homophilous rewiring (*rH*), (3) the rate of

triadic closure (*rT*), and (4) the scaling factor (z), which is a parameter used for generating the initial conditions (see Part 1 above for details). For each of the parameters, we present a set of growth contour maps that can be compared with the reported results in Figure 2.3a. When testing one of the four parameters, we set the others as: *rI*=0.1, *rH*=0.001, *rT*=0.001, and *z*=6. All the results reported below are averaged over 100 replications.

3.1. Rate of Social Influence.

We vary the rate of social influence, indicating the probability that an agent is influenced by her neighbors at each iteration. Figure S2.2 shows that the curvilinear patterns of growth are persistent even though the range of social influence rate varies by 25-fold between 0.02 and 0.5. However, there are significant changes in contour maps. We observe an upward movement of the growth "sweet spot" in the strategy space as the rate of social influence increases. When the social influence rate is very low at 0.02 (left panel), relatively low levels of social encapsulation can achieve positive membership growth, while high levels of encapsulation produce somewhat less optimal outcomes. This is expected, because when the social influence rate is low, behaviors spread more slowly. Thus, higher rates of social encapsulation close the organization off too much to be able to effectively recruit outsiders. As we increase the social influence rate (right panel), the opposite pattern emerges. The focal organization requires a higher rate of encapsulation to keep the new recruits socializing within the boundary of the organization.



0.02 0.05 0.1 0.2 0.5

Figure S2.1. Social influence.

3.2. Homophily.

To see whether our model is robust to homophily, one of the most researched tie formation mechanisms (see McPherson, Smith-Lovin and Cook 2001), we included it as a tie rewiring mechanism in the robustness analysis. We implement homophilous rewiring by first assigning each agent with two attributes, sampled from a continuous

uniform distribution on the unit interval. The likelihood of agent *i* redirecting her ties to agent *j* is modeled by a McFadden (1974) preference function:

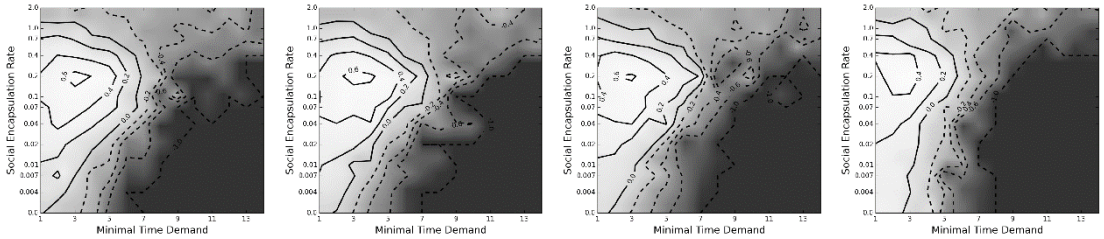
$$p_{ij} = \frac{e^{h \times S(i,j)}}{\sum_k e^{h \times S(i,k)}}$$

where $S(i,j)$ is the similarity between *i* and *j*:

$$S(i,j) = 1 - \sqrt{\frac{\left(Age_i - Age_j\right)^2 + \left(Edu_i - Edu_j\right)^2}{2}}$$

Thus, we add a step in the "Network rewiring" stage of the simulation algorithm (Table 1), wherein each agent, *i*, rewires a tie to a similar agent, *j*, based on the parameter *h*. When *h* is 0, the McFadden preference function reduces to a random rewiring procedure. When *h* approaches infinity, it becomes a deterministic model in which only the most similar pairs of agents can be tied. In the initialization of the simulation, we set *h* at a moderate level (20), and let the network be rewired until the homophily measure reaches equilibrium. After the simulation starts, the homophily mechanism is applied with a specific rate *rHom,* indicating the likelihood that an agent redirects an attention unit to a similar other during the iteration.

We vary the rate of homophilous rewiring (*rHom*) from 0 to 0.1, and the responses in growth rate shown in Figure S2.3 are similar to each other, indicating that the outcome produced by our main model is robust to homophilous rewiring.



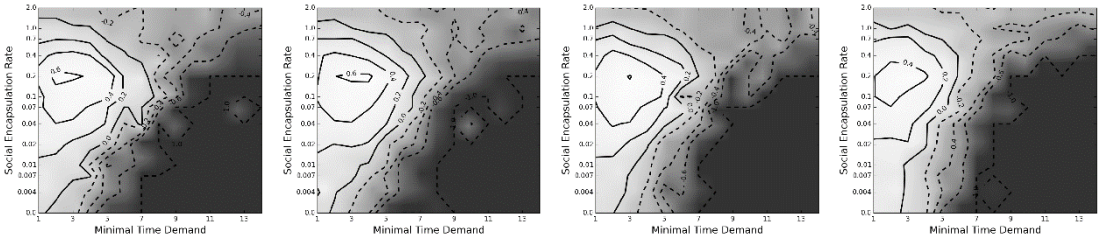
0 0.001 0.01 0.1

Figure S2.2. Homophilous rewiring.

3.3. Triadic Closure.

Triadic Closure operationalized as reallocating one unit of attention between two randomly chosen friends from the agent's ego-centric network. In the robustness analysis, we vary the rate of triadic closure from 0 to 0.1. The simulation results

(Figure S2.4) show a similar pattern of growth to the results. Our analysis confirms that the curvilinear pattern produced by the main model is robust to the rewiring mechanism based on triadic closure.

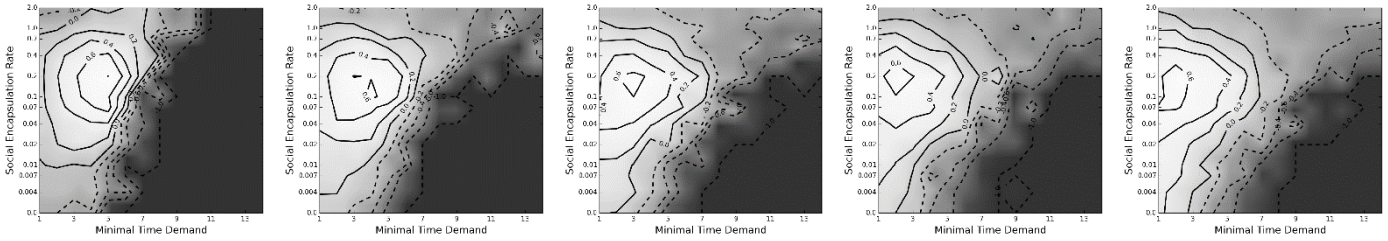


0 0.001 0.01 0.1

Figure S2.3. Triadic closure.

3.4. Scaling Factor z.

Below we report the contour maps for the simulations generated from different initial conditions. We vary the scaling factor z from 2 to 10. We avoid z=0 (random assignment) and the value of z above 10 (which resembles the highly skewed distribution of $T(i,j)^z$ when z approaches infinity). In Figure S2.1, the curvilinear growth patterns are consistent for the scaling factor z varying from 2 to 10. The behaviors of the model under different values of z are also consistent with expectations. When z=2, the contour map (left panel) shows a larger declining area at the high end of the demand dimension, indicating that high-demanding organizations are unlikely to form, and more likely to disband if they do form. When z=10 (right panel), both low- and high-demanding organizations can grow more comfortably (indicated by the high rate of growth at the low end of the demand dimension and the smaller dark area at the high end.)



2 4 6 8 10

Figure S2.4. Scaling factor z.

3.5. Other Parameters.

In addition to these four parameters, we also test other parameters, including

population size (100, 200 and 500), range of time and energy demand (1-10 and 1-20), range of social encapsulation rate (0-1.0 and 0-10.0), individual time budget (10 and 20), and individual attention budget (100 and 200). Patterns in the results are consistent with the main results reported in the paper.

Part 4. Organization Diversity in Different Environments.

We now turn to the question of whether some strategies lead to the maintenance of greater organizational diversity, by measuring the number of organizations that survive until the end of the simulation. Figure S2.5 presents this measure for different combinations of strategies used by organizations in the ecology. The results indicate a monotonic negative relationship between diversity (organizational count) and time and energy demand. The lower the organizations' demand for membership, the more unique organizations survive in the simulation. Social encapsulation interacts with time demand. When the organizations in the simulation use the high encapsulation strategy, the ecology tends to sustain more organizations. This is consistent with the prediction that closed social networks enhance organizational survival.

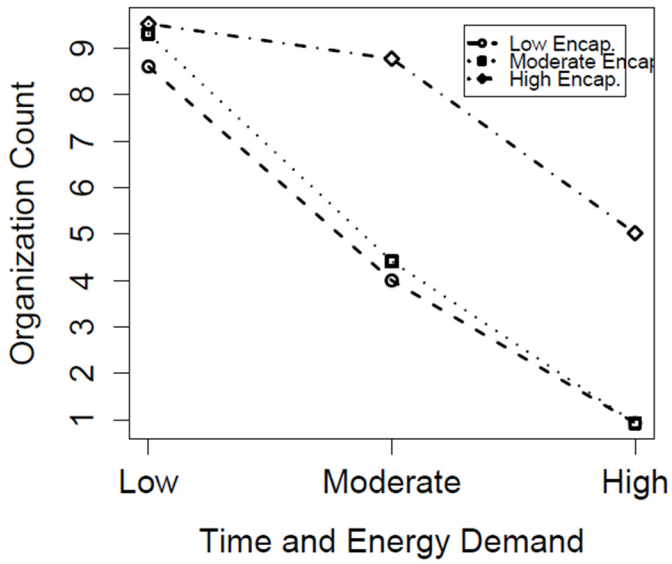


Figure S2.5. Organization Diversity in Different Environments.

MEASURING STRUCTURAL SIMILARITY IN Large ONLINE NETWORKS[18]

Abstract

Structural similarity based on bipartite graphs can be used to detect meaningful communities, but the networks have been tiny compared to massive online networks. Simulation analysis holding underlying similarity constant shows that two widely used measures – Jaccard index and cosine similarity – are robust to increases in network size but biased by the distribution of out-degree. However, an alternative measure, the Standardized Co-incident Ratio (SCR), is unbiased. We apply SCR to members of Congress, musical artists, and professional sports teams to show how massive co-following on Twitter can be used to map meaningful affiliations among cultural entities, even in the absence of direct connections to one another. Our results show how structural similarity can be used to map cultural alignments and demonstrate the potential usefulness of social media data in the study of culture, politics, and organizations across the social and behavioral sciences.

[18] This is a paper coauthored with Michael Macy. Yongren Shi is the first author.

Introduction

Growing availability of social media data provides unprecedented opportunities to use bipartite graphs to reveal unobserved communities among celebrities, organizations, topics, issues, news items, events, books, bands, and videos (Golder and Macy 2014). A bipartite graph consists of two modes, such as celebrities and their followers on Twitter, based on the relationships between members of one mode (e.g. followers) and members of the other (e.g. those followed). These online relationships might include following, liking, commenting, replying, retweeting, purchasing, listening, or watching. For example, co-following on Twitter can be represented as a bipartite network of followers and the friends they follow. The key behavioral assumption is that the greater the similarity between *i* and *j*, the higher the probability that a user who follows *i* will also follow *j*. These similarities and differences among those who are followed can be used to detect meaningful communities, even if the members have no social ties to one another and even if we are unable to measure the attributes of the members. Simply put, instead of classifying the nodes ourselves, based on the analysis of ties and attributes of the nodes, we in effect "crowdsource" the analysis to the millions of users who choose to follow the nodes, based on the assumption that there is meaningful information in those choices that can be extracted using a bipartite graph.

This method need not be limited to communities of people. For example, book co-purchases on Amazon can be used to differentiate clusters of similar content, and co-listening on last.fm can be used to identify musical genres, without having to rely on expert knowledge or observational data about the particular cultural entities.

Bipartite network data from online activities can cover an exceedingly wide variety of topics, ranging from politics to music, from hot-button social issues and events to obscure subcultures.

The digital records of billions of online transactions and communications promise to have a transformative impact on social science. For the past century, quantitative cultural analysis at population scale has relied on opinion surveys. These instruments typically measure subjective and retrospective self-reports based on a set of questions that researchers think are important, not what respondents care about. In contrast, online data are in real time and capture choices that may not be detected in survey responses. For example, consumer preferences involving the expenditure of money may differ from preferences expressed in response to questions in which respondents are minimally invested and about which they feel constrained by perceptions of socially appropriate attitudes. Although surveys can use stratified random samples to generalize results to the underlying population, which is not possible with data from a convenience sample, online data can capture a significant proportion of the entire population, including hard to reach groups such as the very rich and very poor that are susceptible to undercounting in surveys.

Despite the enormous potential applications of bipartite networks to research on public opinion, culture, politics, mass consumption, and social movements, scalable measures of structural similarity have not been developed and validated. Bipartite graphs have been used to reduce complex and often ephemeral cultural preferences to simpler and more durable structures (Mohr 1998). For example, belief graphs have been used to detect cognitive clusters using the associations between individuals and

beliefs (Boutyline and Vaisey, forthcoming; Martin 2002). Bipartite graphs have also been used in content analysis that links words and texts (Tilly 1997; Baldassarri and Diani 2007). However, these applications have rarely involved networks with millions of nodes. This paper examines whether measures of structural similarity that have been widely used for relatively small networks can be applied at Web scale.

The scalability of structural measures developed for small networks cannot be taken for granted. Bipartite networks derived from online activities are not only many orders of magnitude larger, they are also highly skewed in degree distribution. The most popular online celebrities and cultural entities attract tens of millions of followers, purchasers, comments, retweets, and likes, while the vast majority of the entities only attract a handful. In a bipartite network of the kind found in nearly all previous studies, these differences are constrained by the small numbers of nodes. At web scale, these differences could be many orders of magnitude.

In this paper, we use simulation experiments to test the scalability of two widely used continuous measures of structural similarity in bipartite graphs – the Jaccard index (Jaccard 1901) and cosine similarity. Unlike observational data from empirical networks, simulation experiments provide "ground truth" about the similarity of nodes in networks of different size and structure. Holding similarity constant, we varied the size and structure of the simulated network in order to test whether these measures of similarity are biased in ways that could compromise their use with massively large online networks. The results show that Jaccard index and cosine similarity are biased by the highly skewed degree distributions encountered in large networks. We diagnose the source of the bias and propose an alternative

measure, the standardized co-incident ratio (SCR) that avoids the problem by comparing the observed co-follower count to the co-follower count expected in an otherwise identical randomized network. Simulation experiments show that SCR is an unbiased measure.

The second part of the paper applies SCR to data for millions of Twitter co-followers to illustrate the usefulness for discovering meaningful communities in large networks (Garimella and Weber 2014). Observational data cannot be used to assess the external validity of any of structural similarity measures since we have no way to know what attributes attracted the co-followers or how similar are co-followed users on the relevant attributes. Simply put, with observational data, there is no ground truth similarity with which to assess the similarity predicted by alternative structural measures.

Instead, we used observational data to see if meaningful communities could be detected by the one-mode projection of bipartite graphs when applied across several socio-cultural domains, including politics, music, and sports. We analyzed political co-following using a network of 4.9M followers of the US Congress, 58M followers of U.S. music artists, and 11M followers of professional sports teams. These three co-follower graphs have three important properties:

- Members vary widely in the number of followers and co-followers.
- Members are likely to share the same followers because they share an attribute that is relevant to their followers.

- Members have observable cultural attributes that can be used to measure similarities on multiple dimensions and to see how closely these match with the similarities predicted by structural similarity.

In each application, the communities in the projection of an affiliation network corresponded to those that could be expected: Congressional parties, musical genres, and sports leagues.

Measures and Methods

We assess the scalability of two widely used continuous measures of structural equivalence – the Jaccard index and cosine similarity. To preview the results, we find that neither measure is affected by network size, holding the degree distribution constant. However, both measures are affected by differences in degree distribution, holding network size constant. We then show that the bias can be corrected by measuring observed similarity relative to the similarity expected in an otherwise identical randomized network. Remarkably, when measured relative to their expected values, the standardized versions of Jaccard index and cosine similarity converge to the ratio of observed to expected common neighbors, which we refer to as the standardized co-incident ratio (SCR).

2.1 Measures of Structural Similarity

Table 1 is used to illustrate alternative measures of structural similarity between nodes

i and *j* in a bipartite network:

Table 3.1: Cell counts of binary arcs in a bipartite network for constructing continuous pairwise measures of structural similarity.

	Not adjacent to *j*	Adjacent to *j*
Not adjacent to *i*	*a*	*b*
Adjacent to *i*	*c*	*d*

Standardized co-incident ratio (SCR_{ij}) measures the ratio of the observed co-follower count and the expected co-follower count in an otherwise identical randomized network.

$$SCR_{ij} = \frac{d}{E(d)}$$

Cosine similarity (C_{ij}) measures the cosine of the angle between two vectors. For binary vectors, the cosine is the ratio of the common neighbors (d) to the geometric mean of the neighbors of i (or $c+d$ in Table 1) and the neighbors of j (or $b+d$):

$$C_{ij} = \frac{d}{\sqrt{(c+d)(b+d)}}$$

Jaccard index (J_{ij}) measures structural similarity as the proportion of the number of common neighbors that are neighbors of both. As with cosine similarity, the common neighbors are the numerator, which is standardized by the union of the followers instead of the geometric mean:

$$J_{ij} = \frac{d}{b + c + d}$$

Phi Coeffiecient (ϕ_{ij}) is a reduced form of the Pearson product-moment correlation for binary vectors. It is unique in giving equal weight to those who follow neither *i* nor *j* (*a*) as to those who follow both (*d*). Similarity is then reduced by the number who follow one but not the other, which allows the similarity measure to be negative, a second unique property:

$$\phi ij = \frac{(a \times d) \; - \; (b \times \mathrm{c})}{\sqrt{(a + b) \times (c + d) \times (a + c) \times (b + d)}}$$

Phi is related to the chi-squared statistic for a 2×2 contingency table: $\phi^2 = \frac{\chi^2}{\mathrm{N}}$. Expanding the chi-square, we get

$$\phi ij^2 = \frac{1}{N} \sum_{\mathrm{i}} \sum_{\mathrm{j}} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

where *i* and *j* are the row and column indices in the 2×2 matrix, and O_{ij} and E_{ij} are the observed and the expected counts in the cell *ij*. *N* is the matrix sum, where the observed count is influenced by the co-absence count *a* which can be very large at Web scale due to the fact that most nodes attract connections from only a tiny fraction of the population in the network, hence *a* is likely to be very large relative to *d*. For example, in the Congressional co-following network discussed below, the mean value of *a* among 552 members is 4,889,037, which is 36,622 times as large as the mean value of *d* (133.5). In short, the diagonal count swamps the off-diagonal in this

network, as is likely in all large-scale online networks. With $a \gg [b,c]$ and $a \times d \gg b \times c$[19], the $b \times c$ term in phi's numerator becomes negligible, leaving $a \times d$ as the numerator. In the denominator, $a \gg [b,c]$ means $(a+b) \times (a+c)$ is reduced to $a \times a$. Canceling a in both denominator and numerator then reduces phi to cosine similarity. In short, phi converges with cosine as network size increases. Since our interest is focused on scalability to large networks, we omit consideration of phi and focus on Jaccard index and cosine coefficient, the two most widely used measures of structural similarity.

2.2 Testing Scalability with Simulation Experiments

We assess the robustness of Jaccard index and cosine similarity in three experiments. In experiment 1, we vary network size while holding degree distribution constant. In experiments 2 and 3, we vary the in-degree and out-degree distribution[20] while holding network size constant. The experiments use an empirical network drawn from the same Twitter data we use later in the paper to illustrate empirical applications of the similarity measures. The data consists of a set of Twitter accounts of 115 current and recent members of the U.S. Senate and about 1M unique followers.

In experiment 1, we measure structural similarity as we vary the size of the Senate-follower network by randomly sampling followers. The sampled networks

[19] In the Congressional co-follower graph, the mean value of a among all possible pair of Congress members is 198 times as large as the mean value of b or c. The mean value of $a \times d$ is 120 times as large as $b \times c$.
[20] In-degree refers to the number of users who follow a Senator and out-degree refers to the number of Senators followed by a user.

differ in size by orders of magnitude, from 1000 to 1M but not in network structure. In addition to Jaccard index and cosine similarity, we also calculated a standardized version of the co-follower count which we refer to as the standardized co-incident ratio (or SCR). We have no way to know the latent similarity that is captured by co-follower relations, but random sampling insures that there should only be random variation as network size increases. Hence an unbiased measure of structural similarity should not be sensitive to the size of the network. That is what we find for all three measures. Each measure is mean-centered and normed by the standard deviation to provide a common metric. Figure 3.1 reports the mean pairwise structural similarity in each network which is invariant over very large differences in network size.

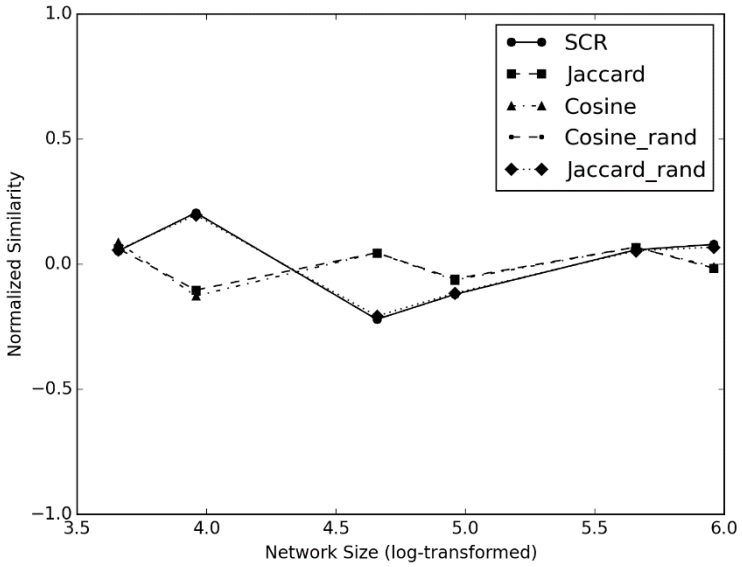


Figure 3.1: Structural similarity measured by SCR, Jaccard and Cosine is unaffected by changes in network size, holding degree distribution constant.
We varied network size by randomly sampling the followers of 115 current and recent U.S. Senators who are widely followed on Twitter. Cosine_rand and Jaccard_rand are also included for comparison.

However, unlike our experiment, network structure does not remain constant as empirical networks increase in size. The upper bound of degree increases with size while the lower bound does not. We therefore need to also test the robustness of similarity measures to differences in the distribution of degree. Experiment 2 is similar to experiment 1 in that we compare networks created by randomly sampling the nodes, but in this study, we vary the degree distribution while holding the size constant. First, we removed high-degree followers, leaving a subset of followers with similar in-degree. Second, we ranked the Senators by in-degree from the followers who remained and randomly removed 50% from among the X% with the highest degree, where X=[50, 55, 60, 75, 100], such that the Gini coefficient ranges from .3 to .8. This procedure holds the distribution of out-degree constant as we manipulate in-degree. Figure 3.2 compares the mean pairwise similarity across networks of identical size whose degree distributions become increasingly skewed as measured by the Gini coefficient. The results show that both the Jaccard index and cosine similarity are unbiased by changes in the distribution of out-degree.

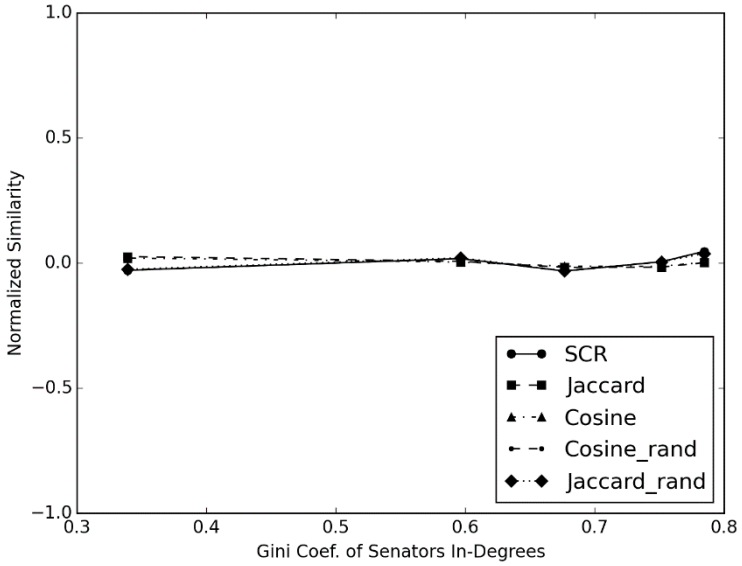


Figure 3.2: Structural similarity is unaffected by changes in the in-degree distribution, holding network size and the out-degree distribution constant.
We varied the in-degree distribution by removing 50% from among the *X*% with the highest degree, where *X*=[50, 55, 60, 75, 100], such that the Gini coefficient ranges from .3 to .8.

Experiment 3 is identical to experiment 2 except that we hold the distribution of in-degree constant as we vary out-degree. First, we removed high-degree Senators, leaving a subset of 66 Senators with similar in-degree. Second, we ranked the followers of the remaining Senators by out-degree and removed 25% (14,600) from among the *X*% with the highest degree, where *X*=[25, 27.5, 30, 37.5, 50, 75, 100], such that the Gini coefficient ranges from .1 to .5. This procedure controls for the possibility that users who follow few Senators prefer those that are most popular. However, it is still possible that those who follow many Senators are less discriminating (Adamic and Adar 2003). Thus, as we remove followers with high out-degree, we should expect to observe either no change in mean pairwise similarity or an increase in similarity as the less discriminating followers with high out-degree are

removed and the degree distribution becomes less skewed. Surprisingly, Figure 3.3 shows the opposite: Both Jaccard index and cosine similarity are *higher* when the Gini coefficient is relatively large, that is, when the distribution of out-degree is *more* highly skewed. The reason is straightforward: When the out-degree distribution is highly skewed, there are more high-degree followers who follow almost every Senator. Because of the prevalence of co-following produced by these high-degree followers, *d* tends to be larger relative to the increase in *a* and *c*, as defined in Table 1. That is why we observe an increase in Jaccard index and cosine similarity as Gini increases.

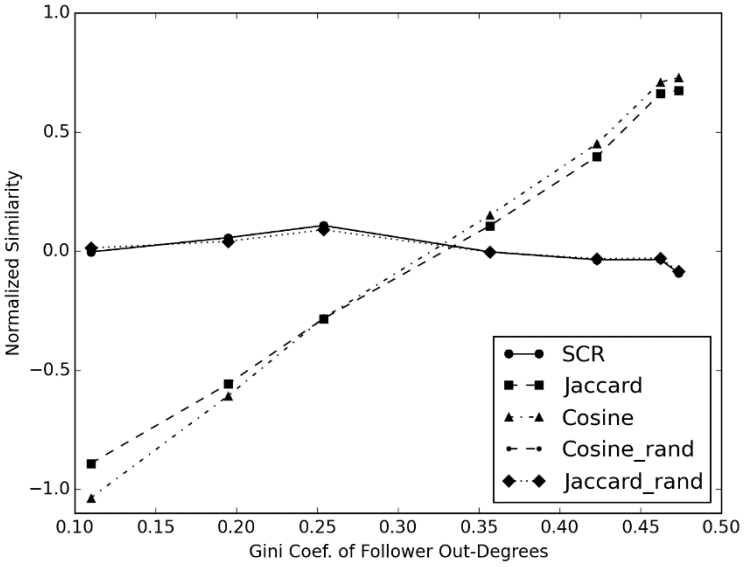


Figure 3.3: Structural similarity is biased by changes in the out-degree distribution, holding network size and the in-degree distribution constant.
The design is identical to that in Figure 3.2 except that we remove high-degree Senators instead of followers, such that the Gini coefficient ranges from .1 to .5.

In contrast, similarity measured by SCR shows no upward trend. SCR has the same numerator as the other two measures – the number of co-followers of each

Senate pair (*d* in Table 1). The difference is the denominator. Instead of comparing the number of co-followers with the number of followers (measured either as the union or the geometric mean), SCR controls for the expected increase in co-following among high-degree followers by taking as a baseline the number of co-followers expected by chance in an otherwise identical network. The randomization procedure preserves the original in-degree and out-degree of every node (Zweig and Kauffman 2011; Gionis, Mielikainen, and Tsaparas 2007; Neal 2014). For a randomly selected Senator *S* with in-degree k_S we sample the k_S nodes from the set of Senate followers based on a probability that is proportional to the followers' degrees. We assign these followers to *S* and update the out-degree of the followers. We repeat the procedure until no followers remain to be assigned.[21] The results in Figure 3.3 show that the expected similarity in a randomized network effectively reduces the bias introduced by large numbers of followers with high out-degree. High-degree followers are also more likely to co-follow in the randomized network, which cancels out the higher observed co-follower count.

As additional confirmation, we also standardized the Jaccard index and cosine similarity by their expected values using the same network randomization procedure as for SCR. We found that both measures converge to SCR when the bias introduced by the inflated co-follower count is removed. (The correlation with SCR is 1.0 for the standardized Jaccard index and .99 for the standardized cosine.) For the remainder of

[21] The last edge to be shuffled is susceptible to being dropped if there is no way to include the edge without altering the degree of the other node. This problem does not arise with any edges prior to the last.

the paper we therefore limit the analyses to SCR, the simpler of the three standardized measures.

Empirical Results

Using the Twitter Public Application Programming Interface (API), we collected three datasets containing all the followers of 552 U.S. Congress members from the 111th, 112th and 113th Congress, 1368 U.S. music artists, and 123 teams from the five major North American sports leagues. We chose these three domains because the entities are widely followed on Twitter and have in common at least one highly salient and identifiable attribute – party membership, musical genre, and league. Nevertheless, in contrast to the simulation experiments, these attributes cannot be used to directly compare structural and cultural similarity because we have no way to know what attributes attracted the attention of followers, or if the same attributes were equally relevant to all followers. For example, we found that co-following in all three domains reflects status similarity, due to the tendency for people to follow publicly visible celebrities. As a consequence, two celebrities might be co-followed even though they have no other attribute in common except their celebrity. In addition, co-following of politicians and sports teams reflects spatial proximity (e.g. a Senator and Representative might be co-followed by residents of the same Congressional district, and a baseball and football team might be co-followed by fans in the same city.) Religion, gender, age, ethnicity, and physical attraction may also have an effect, although we do not have data with which to explore these possible influences.

Although we cannot use empirical observations to verify structural measures of

similarity, co-following can be used to see if we can detect hidden community structure (the extent of clustering in the unipartite projection of the network) and to see if these communities are meaningful (i.e., do they correspond with observable attributes that might be expected to differentiate the nodes). If so, co-following provides a potentially useful way to reveal hidden affiliations among widely followed Twitter users, even in the absence of information about their attributes or the relationships among them.

We looked for community structure based on the modularity index (Q) commonly used in the community detection literature (Newman and Girvan 2004). Modularity measures the extent to which individual nodes are connected with others who share the same community membership. In mathematical form, the modularity of the weighted network can be represented as (Newman 2004):

$$Q = \frac{1}{2w} \sum_i \sum_j \left(w_{ij} - \frac{w_i w_j}{2w} \right) \delta(C_i, C_j)$$

where the edge weight w_{ij} is the structural similarity between two Twitter users i and j, $2w$ is the sum of all edge weights in the graph

$$2w = \Sigma_{ij}(w_{ij}),$$

w_i is the sum of all edge weights attached to i

$$w_i = \Sigma_j(w_{ij}),$$

and C_i indicates i's community membership.

Community memberships were detected using Louvain Modularity (Blondel *et al.* 2008) that finds the partitions that optimize Q. The Kronecker delta function $\delta(C_i, C_j)$ takes the value 1 if nodes i and j are in the same community and 0 otherwise. The

modularity Q of a network could range from -0.5 to 1, where Q=0 means that communities cannot be detected (i.e. each "community" may be nothing more than random clustering); Q>0 means that the sum of the weighted edges within the community exceeds the sum expected in the randomized network.

We assess qualitatively whether the communities are meaningful using network visualizations of the unipartite projections, where pairwise node proximities correspond to structural similarity and nodes are colored based on political party, musical genre, and sports league. The qualitative assessments are accompanied as well by a quantitative measure of the Q modularity when C is assigned based on node attributes.

3.1 Empirical Applications to Politics, Music, and Sports

Using SRC to measure structural similarity, Q modularity for the unipartite projections of the biparatite co-follower networks for Congress members (.1), musical artists (.349), and North American sports teams (.45) are all positive, indicating that there is a non-random optimal partitioning. The results show that co-following can be used to detect hidden community structure in the affiliations among Congress members, musicians, and sports teams.

Having found non-random community structure in all three co-follower networks, we turn to the more important question as to whether these detected communities are substantively meaningful. Put differently, it is not enough to show that the community structure is non-random, we also want to know if the structure reflects intuitively plausible user attributes that might be expected to attract social media followers. To that end, we examine each of the three co-follower networks in

turn.

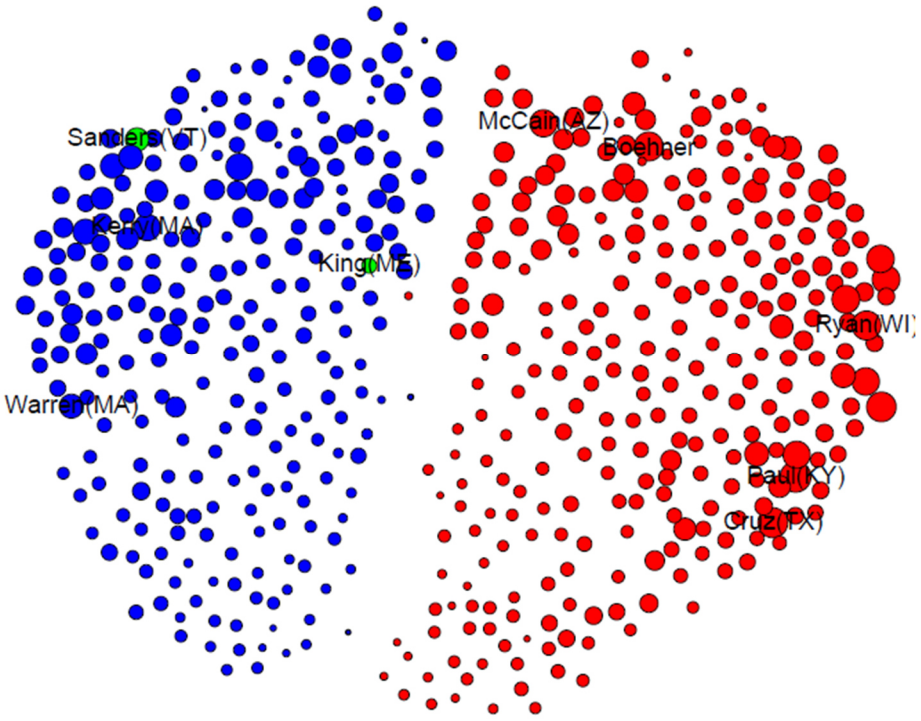
3.2 U.S. Congress

Figure 3.4[22] visualizes the unipartite projected affiliations among 552 widely followed Congress members, where pairwise proximities correspond to structural similarity using SCR, size indicates the relative follower count, and nodes are colored according to party membership as Republican (red), Democrat (blue), or Independent (green). It is immediately evident that Congressional co-following can be used to recover party membership, which suggests we can meaningfully identify the location of individual members in the network of party affiliations derived from structural similarities. Notice that Independents are on the periphery of the blue cluster, with Bernie Sanders most closely affiliated with Democrats who are farthest from Republicans.

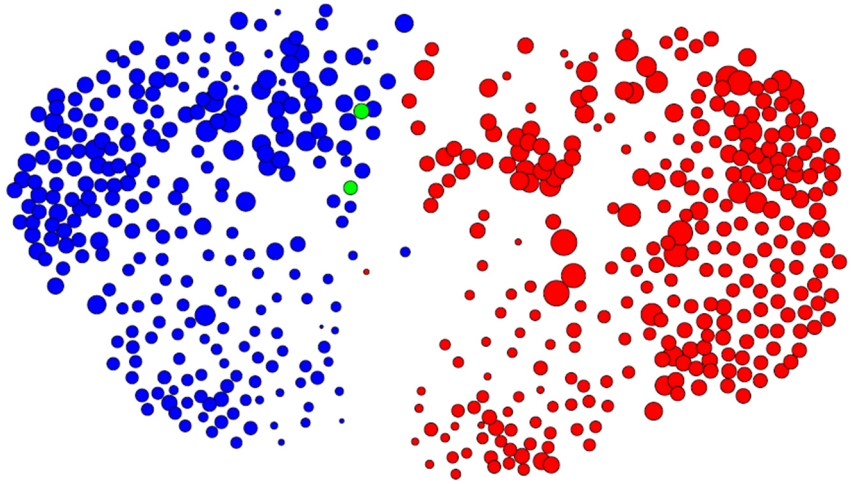
As additional support, the Q modularity based on party membership is Q =.093, compared to Q =.1 using partitions that optimize Q. In other words, party membership is very close to the optimal partitioning. These results are consistent with a related study by Bond and Messing (2015) that used Facebook "likes" of political pages to estimate the ideological alignment of politicians and their supporters. Although they use factorization rather than modularity, they find a strong correlation (.94) between their structural measure of ideology and Poole and Rosenthal's (1991;

[22] The network visualization follows the procedure introduced in Serrano, Boguñá, and Vespignani (2009) for extracting the "backbone" in complex weighted networks, arcs are deleted if the weight is less than one standard deviations above the mean in the distribution of weights in each node's ego networks. The resulting network is directed. Then the tie weights are mean centered by dividing the mean of the weights in the ego's network. This step is only for keeping the network visually compact while it does not change the relative positions of nodes.

2001; Palfrey and Poole 1987) DW-Nominate measure based on roll call voting).



(a)



(b)

Figure 3.4. Network Visualization of U.S. Congress members by Party Affiliation as Republican (red), Democrat (blue), and Independent (green).
3.4a is based on all 5M followers of Congress and 4b is limited to those who follow five or more members. Nodes are sized by follower count.

The partisan polarization evident in Figure 3.4 also speaks to the debate among political scientists as to whether the divisions among political elites reflect divisions within the electorate (Fiorina, Abrams and Pope 2005; DellaPosta, Shi and Macy 2015; Abramowitz 2010; DiMaggio, Evans, and Bryson. 1996; Baldassarri and Gelman 2008). The polarized affiliations in the network visualization are the one-mode projection of the polarized political preferences among millions of citizens in choosing which members of Congress to follow (see also Bond and Messing, 2015, who reach a similar conclusion).

Finally, Figure 3.4 reveals a tendency for the most popular members to be clustered together on the periphery of the network. This indicates that co-following reflects in part degree similarity, i.e., a tendency of those who follow celebrities to a) follow other celebrities and b) only follow other celebrities. For example, we found that the higher a given politician's follower count, the smaller the number of other politicians with whom that member is co-followed. Additional analyses show that SRC is about equally correlated with similar popularity (r=.26) and with similarity in voting record (r=.25) as measured by Poole and Rosenthal (2001).

In addition to party membership and popularity, co-follower ties can also be expected to reflect spatial proximity since that is the basis for Congressional representation. Given the "big sort" documented by Bishop (2009), spatial co-following is likely to reinforce partisan affiliation. Table 2 reports the mean pairwise co-follower similarity as measured by SCR, broken down by same-party and between-party pairings.

Table 3.2. Pairwise SCR Broken Down by State and Party Co-location

	Same Party	Different Party
Same State	0.16	0.09
Different State	0.10	0.06

In addition to party membership and popularity, co-follower ties can also be expected to reflect spatial proximity since that is the basis for Congressional representation. Given the "big sort" documented by Bishop (2008), spatial co-following is likely to reinforce partisan affiliation. Table 2 reports the mean pairwise structural similarity among same-state and between-state Congressional pairings, broken down by same-party and between-party pairings. As expected, people tend to co-follow politicians from the same state, but co-following by geolocation is much weaker than by party affiliation.

3.3 U.S. Musical Artists

Using the Rovi API[23], we collected genre information for every U.S. musical artist with a Twitter account listed on Musicbrainz.com, a publicly created open content music database. We then used the Twitter public API to obtain the 122M followers of the 1368 artists with at least 10,000 followers in the six most popular genres. Figure 3.5 is a network visualization in which proximities correspond to pairwise SCR similarities, circle sizes indicate follower counts, and nodes are colored according to genre: Pop/Rock (purple), Rap (green), R&B (black), Electronic (orange), and Country (red). Music genre produces a high level of modularity (Q=0.167) as evident in the network visualization. Country and Rap are farthest

[23] http://developer.rovicorp.com/io-docs

apart, with R&B closer to Rap and Pop/Rock closer to Country. The two most cohesive communities are Electronic and Country, while the largest communities, Pop/Rock and Rap, are least cohesive.

3.4 North American Professional Sports Teams

We collected 11M unique Twitter followers of 123 North American teams with more than 10,000 followers. Figure 3.6 displays the network visualization, with nodes colored by membership in the five major professional leagues, NFL (green), MLB (blue), NHL (orange), NBA (purple) and MLS (red). The league-based community structure is highly differentiated, as confirmed by much higher modularity (Q=.452) than found in politics and music. While this may seem obvious after the fact (Watts 2011), it is also highly plausible that teams might have been clustered by city since that is the common denominator of the fans of professional sports, while other members of the same league are presumably regarded as rivals by devoted fans. While spatial proximity is by far the best predictor of between-league co-follower affinities, the league remains a far stronger organizing principle of the community structure. Soccer is least popular and the least densely connected to other sports, while basketball, football, and baseball are most popular. However, basketball is similar to soccer in its relative isolation from other sports, leaving baseball and football as the "sister leagues."

Figure 3.5. Network Visualization of Musical Artists by Genre .

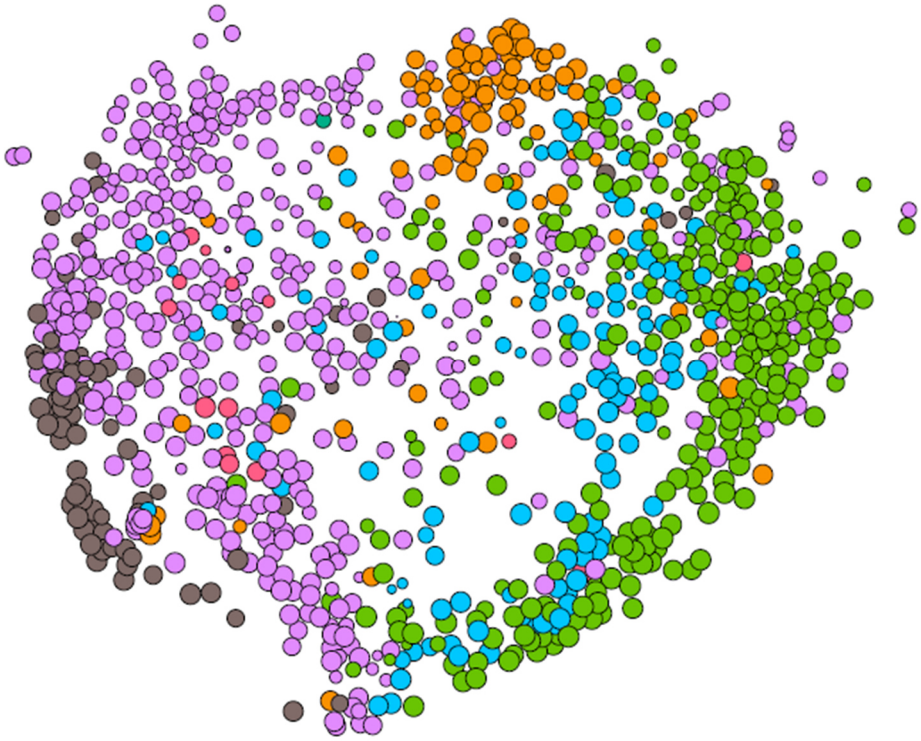
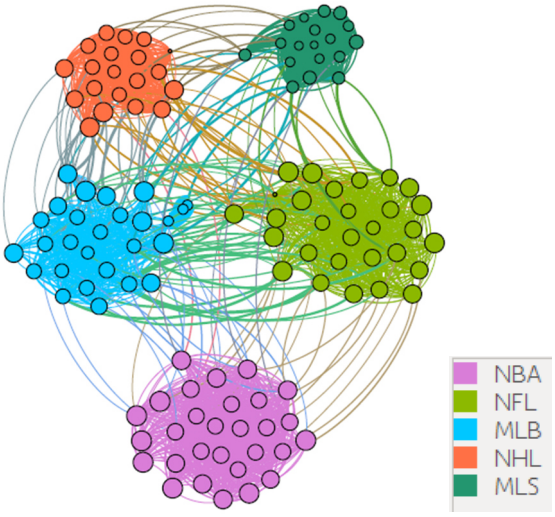


Figure 3.6. Network Visualization of Sports Teams by League.



Discussion and Conclusion

Analysis of the Congressional co-follower network shows that popularity as well political ideology attracts followers. As a consequence, co-following on Twitter exhibits the bandwagon effect of "preferential attachment." Following a celebrity on Twitter is not only motivated by interest in the attributes of the celebrity but also because everyone else is following that person. Two celebrities could have enormously large numbers of co-followers even though they have nothing in common other than their celebrity.

Followers also differ in degree. The more people they follow, the less discriminating they are in their choices, which means that they provide less information in evaluating similarity compared to those who follow selectively. In short, differences in degree pose a challenge to measures of structural similarity: how can we know whether co-following indicates a meaningful cultural affiliation that is not contaminated by a celebrity's popularity or their followers' indiscriminate selection? The answer, we propose, is to take into account the number of co-followers that would be expected by chance, given the distribution of degree. Two widely used measures of structural similarity, Jaccard index and cosine similarity, take into account the number of people that are followed but not the expected number of co-followers. An alternative measure, the standardized co-incident ratio, minimizes the risk of bias due to a highly skewed distribution of degree.

Applications of SCR to co-followers of politicians, musical artists, and sports teams show how social media data can be used to map cultural alignments across any domain that attracts large numbers of followers. This method does not require access

to data about the distribution of cultural attributes or organizational affiliations, nor does it require the existence of, or data about, the social ties among the users that are followed. Where these data are available, modularity analysis can then be used to identify the attributes and/or relationships that map most closely with the pattern of affiliations. Where these data are not available, the affiliations can be used to make inferences about the salient dimensions of cultural alignments.

Bipartite graphs have been long used in social science research to find patterns of associations among culture objects, beliefs and attitudes, and organizational affiliations (Mohr 1998; Martin 2002). Online social media open up vast untapped opportunities to apply these same methods to networks that are many orders of magnitude larger than the hand-coded ethnographic data of an earlier era. Indeed, this method is not limited to co-following on Twitter or to the clustering of individuals. It can be used to map the affiliations among books that are co-purchased on Amazon, songs that are downloaded on last.fm, and social movement hashtags that are diffused over blogs. The possibilities are limited only by our imagination, our access to online data, and the robustness of our models and measures. We hope to have contributed to advances in the latter.

REFERENCES

Abramowitz,AI. 2010. *The disappearing center: Engaged citizens, polarization, and American democracy*. New Haven: Yale University Press.

Adamic, Lada A. and Eytan Adar. 2003. "Friends and neighbors on the Web." *Social Networks* 25:211-230.

Alinsky, S. 2010. *Rules for Radicals*. Knopf Doubleday Publishing Group.

Bishop, Bill. 2008. *The Big Sort: Why the Clustering of Like-Minded America Is Tearing Us Apart.* Boston: Houghton Mifflin.Bond, Robert and Solomon Messing. 2015. "Quantifying Social Media's Political Space: Estimating Ideology from Publicly Revealed Preferences on Facebook." *American Political Science Review* 109:62-78.

Bakshy, E., S. Messing, and L. Adamic. 2015. "Exposure to ideologically diverse news and opinion on Facebook." *Science*, aaa1160.

Baldassarri, Delia and Peter Bearman. 2007. "Dynamics of Political Polarization." *American Sociological Review* 72(5):784-811.

Baldassarri, Delia and Mario Diani. 2007. "The Integrative Power of Civic Networks." *American Journal of Sociology* 113(3): 735-80.

Baldassarri, D. and A. Gelman. 2008. "Partisans Without Constraint: Political Polarization and Trends in American Public Opinion." *American Journal of Sociology*, 114: 408-46.

Barnett, William P., and Glenn R. Carroll. 1995. "Modeling Internal Organizational Change." *Annual Review of Sociology* 21:217-236.

Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte and Etienne Lefebvre. 2008. "Fast unfolding of communities in large networks." *Journal of Statistical Mechanics: Theory and Experiment* (10), P10008 (12pp) doi: 10.1088/1742-5468/2008/10/P10008.

Boutyline, Andrei and Stephen Vaisey. Forthcoming. "Belief Network Analysis: A Relational Approach to Understanding the Structure of Attitudes." *American Journal of Sociology*.

Brashears, Matthew E. 2013. "Humans use Compression Heuristics to Improve the Recall of Social Networks." *Nature Scientific Reports* 3:1513.

Breiger, Ronald L. “The duality of persons and groups.” Social Forces 53:181-190.

Burger, Martjin J., and Vincent Buskens. 2009. “Social Context and Network Formation: An Experimental Study.” *Social Networks* 31(1):63-75.

Centola, Damon. 2015. “The Social Origins of Networks and Diffusion.” *American Journal of Sociology* 120(5):1295-1338.

Centola, Damon and Michael W. Macy. 2007. “Complex Contagions and the Weakness of Long Ties.” *American Journal of Sociology* 113(3):702-734.

Chattoe, Edmund. 1998. “Just How (Un)realistic are Evolutionary Algorithms as Representations of Social Processes?” *Journal of Artificial Societies and Social Simulation* http://jasss.soc.surrey.ac.uk/1/3/2.html.

Chattoe, Edmund. 2006. “Using Simulation to Develop Testable Functionalist Explanations: a Case Study of Church Survival.” *The British Journal of Sociology* 57:379-397.

Coleman, James S. 1990. *Foundations of Social Theory*. Cambridge: Harvard University Press.

Coser, Lewis. 1974. *Greedy Institutions: Patterns of Undivided Commitment*. Free Press.

Cress, Daniel M., Miller J. McPherson and Thomas Rotolo. 1997. “Competition and commitment in voluntary memberships: The paradox of persistence and participation.” *Sociological Perspective* 40:61-79.

della Porta, Donatella. 1988. “Recruitment Processes in Clandestine Political Organizations: Italian Left-Wing Terrorism. *International Social Movement Research* (1):155-169.

Daston, L. and P. Galison, 2007. *Objectivity.* Zone Books ; Distributed by the MIT Press, New York Cambridge, Mass.

DellaPosta, Daniel, Yongren Shi and Michael W. Macy. 2015. “Why do Liberals Drink Lattes?” *American Journal of Sociology* 120(5):1473-1511.

DiMaggio, P., J. Evans, and B. Bryson. 1996. “Have American's Social Attitudes Become More Polarized?” *American Journal of Sociology* 102:690-755.

DiMaggio, Paul J. and Walter W. Powell 1983. “The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields.” *American Sociological Review* 48(2):147-160.

Dunbar, R.I.M. 1992. "Neocortex Size as a Constraint on Group Size in Primates". *Journal of Human Evolution* 22:469–493.

Eakin, E. 2004 "Study Finds a Nation of Polarized Readers." *N. Y. Times.*

Edwards, Bob and John D. McCarthy 2004. "Strategy Matters: The Contingent Value of Social Capital in the Survival of Local Social Movement Organizations." *Social Forces* 83:621-651.

Entwisle, Barbara, Katherine Faust, Ronald R. Rindfuss, and Toshiko Kaneda. 2007. "Networks and Contexts: Variation in the Structure of Social Ties." *American Journal of Sociology* 112:1495-1533.

Farrell, J. 2015. "Corporate funding and ideological polarization about climate change." *Proceedings of the National Academy of Sciences*, 201509433.

Feld, Scott L. 1981. "The Focused Organization of Social Ties." *American Journal of Sociology* 86(5):1015-1035.

Fiorina, M. P. and S. J. Abrams. 2008. "Political Polarization in the American Public." *Annual Review of Political Science* 11: 563–588.

Fiorina, Morris P., Samuel J Abrams, and Jeremy Pope. 2005. *Culture war?* New York: Pearson Longman.

Friedkin, Noah E. and Eugene C. Johnsen 1999. "Social Influence Networks and Opinion Change". *Advances in Group Processes* 16:1-29.

Funk, Cary, Lee Rainie, and Dana Page. 2015. "Public and Scientists' Views on Science and Society" A Pew Research Center Study conducted in collaboration with the American Association for the Advancement of Science (AAAS).

Garimella, Venkata Rama Kiran and Ingmar Weber. 2014. "Co-Following on Twitter." HT '14 Proceedings of the 25th ACM conference on Hypertext and social media: 249-254.

Gauchat, G. 2012. "Politicization of Science in the Public Sphere: A Study of Public Trust in the United States, 1974 to 2010." *American Sociological Review* 77: 167–187.

Geard, N. and S. Bullock. 2010. "Competition and the dynamics of group affiliation." *Advances in Complex Systems* 13:501-517.

Gionis A, Mannila H, Mielikainen T, Tsaparas P. 2007. "Assessing Data Mining Results Via Swap Randomization." ACM Trans. Knowl. Discov. Data 1(3):

14.

Golder, Scott A. and Michael W. Macy. 2014. “Digital Footprints: Opportunities and Challenges for Online Social Research.” *Annual Review of Sociology* 40: 129-152.

Gould, Roger. 1991. “Multiple Networks and Mobilization in the Paris Commune, 1871.” *American Sociological Review* 56:716-729.

Gould, Roger V. 2003. “Why Do Networks Matter? Rationalist and Structuralist Interpretations.” Pp. 233-57 in *Social Movement Networks: Relational Approaches to Collective Action*, edited by Mario Diani and Doug McAdam. New York: Oxford University Press.

Granovetter, Mark. 1995. *Getting a Job: A Study of Contacts and Careers*. Chicago: The University of Chicago Press.

Guimera, Roger, Brian Uzzi, Jarrett Spiro, and Luis A. Nunes Amaral. 2005. “Team Assembly Mechanisms Determine Collaboration Network Structure and Team Performance.” *Science* 308:697-702.

Habermas,J. 1991 *The Structural Transformation of the Public Sphere: An Inquiry Into a Category of Bourgeois Society.* MIT Press.

Hannan, Michael T. and John Freeman. 1984. “Structural inertia and organizational change.” *American Sociological Review* 49:149-64.

Harrison, J.Richard, and Glenn R. Carroll. 1991. “Keeping the faith: A model of cultural transmission in formal organizations.” *Administrative Science Quarter* 36:552-582.

Hechter, Michael. 1987. *Principles of Group Solidarity*. Berkeley, CA: University of California Press.

Hirschman, Albert O. 1970. *Exit, Voice and Loyalty*. Harvard University Press.

Iannaccone, Laurence R. 1994. “Why Strict Churches Are Strong?” *American Journal of Sociology* 99:1180-1211.

Ingram, Paul and Michael W. Morris. 2007. ‘Do People Mix at Mixers? Structure, Homophily, and the “Life of the Party”’ *Administrative Science Quarterly* 52: 558–585.

Jaccard, Paul. 1901. "Étude comparative de la distribution florale dans une portion des Alpes et des Jura", *Bulletin de la Société Vaudoise des Sciences Naturelles* 37:

547–579.

Jelveh, Zubin, Bruce Kogut, and Suresh Naidu. 2014. “Political language in economics.” available at *SSRN 2535453* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2535453.

Johnson, Norman L. and Samuel Kotz. 1977. *Urn Model and Their Application: An Approach to Modern Discrete Probability Theory*. John Wiley&Sons.

Kanter, Rosabeth Moss. 1972. *Commitment and Community: Communes and Utopias in Sociological Perspective*. Harvard University Press.

Kitts, James A. 1999. “Not in our Backyard: Solidarity, Social Networks, and the Ecology of Environmental Mobilization.” *Sociological Inquiry* 69:551-574.

Kitts, James A. 2000. “Mobilizing in Black Boxes: Social Networks and Participation in Social Movement Organizations.” *Mobilization* 5:241-257.

Kohut, A., S. Keeter, C. Doherty, M. Dimock, A. I. Leshner. 2009. “A Survey Conducted in Collaboration with the American Association for the Advancement of Science” Pew Research Center for the People & the Press and the American Association for the Advancement of Science.

Krackhardt, David and Robert N. Stern. 1988. “Informal Networks and Organizational Crises: An Experimental Simulation.” *Social Psychology Quarterly* 51:123-140.

Krebs, Valdis. 1999. “The Social Life of Books: Visualizing Communities of Interest via Purchase Patterns on the WWW.” Available at *OrgNet*: http://www.orgnet.com/booknet.html.

Krebs, Valdis. 2003. “Divided We Stand?” Available at *OrgNet:* http://www.orgnet.com/divided1.html.

Lewis, Kevin, Kurt Gray, and Jens Meierhenrich. 2014. “The Structure of Online Activism.” *Sociological Science* 1:1-9.

Lim, Chaeyoon. 2008. “Social Networks and Political Participation: How Do Networks Matter?” *Social Forces* 87:961-982.

Linden, G., B. Smith, and J. York. 2003. “Amazon.com recommendations: item-to-item collaborative filtering.” *IEEE Internet Comput.* 7: 76–80.

Lizardo, Omar. 2006. “How Cultural Tastes Shape Personal Networks.” *American Sociological Review* 71(5):778-807.

Lofland, John. 1978. "'Becoming a World Saver' Revisited." Pp. 10-23 in J. Richardson (ed.), *Conversion Careers: In and Out of New Religions*. Beverly Hills: Sage.

Lofland, John and Rodney Stark. 1965. "Becoming a World-Saver: A Theory of Conversion to a Deviant Perspective." *American Sociological Review* 30:862-875.

Marsden, P.V., and E.H. Gorman. 2001. "Social networks, Job changes and recruitment." Pp. 467-502 in I. Berg and A. L. Kalleberg (Eds.), *Sourcebook of Labor Markets: Evolving Structures and Processes*. New York: Plenum Press

Martin, John Levi. 2002. "Power, Authority, and the Constraint of Belief Systems." *American Journal of Sociology* 107: 861–904.

Macy, Michael and Robert Willer. 2002. "From Factors to Actors: Computational Sociology and Agent-Based Modeling." *Annual Review of Sociology* 28:143-166.

Mark, Noah. 1998. "Birds of a Feather Sing Together." *Social Forces* 77:453-485.

Maslov, Sergei and Kim Sneppen. 2002. "Specificity and Stability in Topology of Protein Networks." *Science* 296:910-913.

Mayhew, Bruce H. and Roger L. Levinger. 1976. "On the Emergence of Oligarchy in Human Interaction." *American Journal of Sociology* 81:1017-49.

McAdam, Doug. 1982. *Political Process and the Development of Black Insurgency, 1930-1970*. Chicago: University of Chicago Press.

McAdam, Doug. 1986, "Recruitment to High-Risk Activism: The Case of Freedom Summer." *American Journal of Sociology*, Vol. 92, No. 1:64-90.

McAdam, Doug. 2003. "Beyond structural analysis: toward a more dynamic understanding of social movements." in *Social Movement Networks: Relational Approaches to Collective Action*, edited by Mario Diani and Doug McAdam. New York: Oxford University Press.

McAdam, Doug and Ronnelle Paulsen. 1993. "Specifying the Relationship Between Social Ties and Activism." *American Journal of Sociology* 99:640-667.

McFarland, Daniel A., James Moody, David Diehl, Jeffrey A. Smith and Reuben J. Thomas. 2014. "Network Ecology and Adolescent Social Structure." *American Sociological Review* 79:1088-1121.

McPherson, J. Miller. 1983. “An Ecology of Affiliation.” *American Sociological Review* 48:519-532.

McPherson, J. Miller. 2004. “A Blau space Primer: Prolegomenon to an Ecology of Affiliation.” *Industrial and Corporate Change* 13:263-80.

McPherson, J. Miller, Pamela A. Popielarz and Sonja Drobnic. 1992. “Social Networks and Organizational Dynamics.” *American Sociological Review* 57:153-170.

McPherson, J. Miller and James R. Ranger-Moore. 1991. “Evolution on a Dancing Landscape: Organizations and Networks in Dynamic Blau Space.” *Social Forces* 70:19-42.

McPherson, J. Miller and Thomas Rotolo. 1996. "Testing a Dynamic Model of Social Composition: Diversity and Change in Voluntary Groups." *American Sociological Review* 61:179-202.

McPherson, J. Miller and Lynn Smith-Lovin. 1987. “Homophily in Voluntary Organizations: Status Distance and the Composition of Face-to-Face Groups.” *American Sociological Review* 52(3):370-379.

McPherson, Miller and Lynn Smith-Lovin. 2002. “Cohesion and membership duration: Linking groups, relations and individuals in an ecology of affiliation.” *Advances in Group Processes* 19:1-36.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. “Birds of a Feather: Homophily in Social Networks.” *Annual Review of Sociology* 27:415-444.

Massey, Douglas S. and Roger Tourangeau. 2013. “Where do We Go from Here? Nonresponse and Social Measurement.” Ann Am Acad Pol Soc Sci. 645(1): 222–236

Mohr, John W. 1998. “Measuring Meaning Structures.” *Annual Review of Sociology* 24:345-70.

Molloy, M. and B. Reed. 1995 “A critical point for random graphs with a given degree sequence.” *Random Struct. Algorithms*. 6: 161–180.

Morris, Aldon. 1984. *The Origins of the Civil Rights Movement: Black Communities Organizing for Change*, Free Press.

Neal, Zachary. 2014. “The backbone of bipartite projections: Inferring relationships from co-authorship, co-sponsorship, co-attendance and other co-behaviors.”

Social Networks 39:84-97.

Newman, M. E. J. 2004. “Analysis of Weighted Networks.” *Phys. Rev. E* 70, 056131.

Newman, M. E. J. and M. Girvan. 2004. “Finding and evaluating community structure in networks.” *Phys. Rev. E* 69, 026113.

Oliver, Pamela. 1984. “If You Don't Do It, Nobody Else Will: Active and Token Contributors to Local Collective Action.” *American Sociological Review* 49:601-610.

Palfrey, Thomas R. and Keith T. Poole. 1987. “The Relationship Between Information, Ideology, and Voting Behavior.” *American Journal of Political Science* 31:511-530.

Pattison, Philippa and Garry Robins. 2002. “Neighborhood-Based Models for Social Networks.” *Sociological Methodology* 32:301-337.

Pfeffer, J. and G. R. Salancik. 1978. *The External Control of Organizations: A Resource Dependence Perspective*. New York, NY, Harper and Row.

Poole, Keith T. and Howard Rosenthal. 1991. “Patterns of Congressional Voting.” *American Journal of Political Science*. 35(1): 228-278.

Poole, Keith T. and Howard Rosenthal. 2001. *Congress: A Political-Economic History of Roll Call Voting*. Oxford University Press.

Popielarz, Pamela A. and Miller J. McPherson. 1995. “On the Edge Or in Between: Niche Position, Niche Overlap, and the Duration of Voluntary Association Memberships.” *American Journal of Sociology* 101:698-720.

Popielarz, Pamela A. and Zachary P. Neal. 2007. “The Niche as a Theoretical Tool.” *Annual Review of Sociology* 33:65-84.

Putnam, Robert. 1995. "Bowling Alone: America's Declining Social Capital." *Journal of Democracy* 6:65–78.

Putnam, Robert. 2000. *Bowling Alone: The Collapse and Revival of American Community*. Touchstone Books by Simon & Schuster.

Rochford, E.B. 1982. “Recruitment Strategies, Ideology, and Organization in the Hare Krishna Movement.” *Social Problems* 29(4):399-410.

Rotolo, Thomas and Miller J. McPherson. 2001. “The System of Occupations: Modeling Occupations in Sociodemographc Space.” *Social Forces* 79:1095-

1130.

Sabidussi, G. 1966. “The Centrality Index of a Graph.” *Psychometrika*. 31: 581–603.

Saramäki, Jari, E.A. Leicht, Eduardo López, Sam G.B. Roberts, Felix Reed-Tsochas, and Robin I.M. Dunbar. 2014. “Persistence of Social Signatures in Human Communication.” *Proceedings of the National Academy of Sciences* 111:942–947.

Scott, W. Richard. 1992. *Organizations: Rational, Natural, and Open Systems*. (3rd edition). Englewood Cliffs, New Jersey: Prentice Hall.

Serrano, M. A., M. Boguñá, and A. Vespignani. 2009. "Extracting the multiscale backbone of complex weighted networks". *Proceedings of the National Academy of Sciences USA* 106, 6438.

Shapin, S. 1994 *A social history of truth : civility and science in seventeenth-century England.* University of Chicago Press, Chicago.

Shapiro, B. J. 2003. *A Culture of Fact: England, 1550-1720.* Cornell University Press, Ithaca, NY.

Simmel, Georg. 1955. *Conflict and the Web of Group Affiliations*, translated and edited by Kurt Wolff, Glencoe, IL: Free Press.

Simon, Herbert. 1976. *Administrative Behavior*. New York: The Free Press.

Skocpol, Theda. 2003. *Diminished Democracy: From Membership to Management in American Civic Life*. Norman: University of Oklahoma Press.

Small, Mario. 2009. *Unanticipated Gains: Origins of Network Inequality in Everyday Life*. New York: Oxford University Press.

Snow, David, Louis A. Zurcher and Sheldon Ekland-Olson. 1980. “Social networks and social movements: a micro structural approach to differential recruitment.” *American Sociological Review* 45:787-801.

Stark, Rodney and William Sims Bainbridge. 1980. “Networks of Faith: Interpersonal Bonds and Recruitment to Cults and Sects.” *American Journal of Sociology* 85: 1376-1395.

Stark, Rodney and William Sims Bainbridge. 1985. *The Future of Religion*. Berkeley: University of California Press.

Stern, Charlotta. 1999. “The Evolution of Social-Movement Organizations: Niche

Competition in Social Space." *European Sociological Review* 15:91-105.

Stiller, James and R. I. M. Dunbar, 2007. "Perspective-taking and memory capacity predict social network size." *Social Networks* 29:93-104.

Suhay, E. and J. N. Druckman. 2015. "The Politics of Science Political Values and the Production, Communication, and Reception of Scientific Knowledge." *The ANNALS of the American Academy of Political and Social Science* 658: 6–15.

Sutton, G. V. 1997. *Science For A Polite Society: Gender, Culture, And The Demonstration Of Enlightenment.* Westview Press, Boulder, Colorado.

Tilly, Charles. 1997. "Parliamentarization of popular contention in Great Britain, 1758-1834." *Theory and Society* 26(2-3):245-73.

Walgrave, Stefaan, and Ruud Wouters. 2014. "The Missing Link in the Diffusion of Protest: Asking Others." *American Journal of Sociology* 119(6):1670-1709.

Walker, Edward T. 2014. *Grassroots for Hire: Public Affairs Consultants in American Democracy*. New York: Cambridge University Press.

Wellman, Barry. 2001. "Physical Place and Cyberplace: The Rise of Personalized Networking. *International Journal of Urban and Regional Research* 25(2):227-252.

Winship, Christopher. 2009. "Social Interactions, Groups and Scheduling Constraints." In: Hedström P., Bearman P. *The Oxford Handbook of Analytic Sociology*. Oxford University Press.

Yeo, S. K. M. A. Xenos, D. Brossard, and D. A. Scheufele. 2015. "Selecting Our Own Science How Communication Contexts and Individual Traits Shape Information Seeking." *The ANNALS of the American Academy of Political and Social Science* 658: 172–191.

Young, Cristobal and Chaeyoon Lim. 2014. "Time as a Network Good: Evidence from Unemployment and the Standard Workweek." *Sociological Science* 1:10-27.

Zweig, K.A., and M. Kaufmann. 2011. "A systematic approach to the one-mode projection of bipartite graphs." *Social Network Analysis and Mining* 1:187-218.