Drew Margolin, Research Sample:

Margolin, D. B., & Monge, P. (2013). Conceptual retention in epistemic communities: Examining the role of social structure. In P. Moy (Ed.), *Communication and community* (pp. XX-YY). New York: Hampton Press.

Conceptual Retention in Epistemic Communities: Examining the Role of Social Structure

Drew Margolin & Peter Monge

University of Southern California

Abstract

This paper proposes a theoretical mechanism by which the social structure of an epistemic community will influence the conceptual choices that members of the community make when investigating and communicating novel knowledge claims. It is argued that in structurally cohesive communities members have an incentive to rely on well-established, broadly recognized concepts, while in more fractured communities there are greater incentives for individuals to use more idiosyncratic constructions. The mechanism is tested on data collected from two topics in the study of physics: d-branes and supermassive black holes. Results indicate some support for the argument that communities with greater global cohesion, as measured by a low degree of local social clustering, tend to draw more heavily on concepts that are already established in the community.

Keywords: Community Structure, Co-Evolution, Retention, Epistemic Communities

Conceptual Retention in Epistemic Communities: The Role of Social Structure

Knowledge-producing communities represent unique forms of organized, cooperative activity. In many ways, these *epistemic communities* (Roth & Bourgine, 2005) or *invisible colleges* (Crane, 1969) are the purest example of self-organizing systems of human organization (Ashby, 1965; Contractor, 1994). Research on a new topic can begin as a little-noticed offshoot of an established topic or as an entirely new topic with little or no formal governance structure guiding these efforts (van Raan, 2000). The development of these communities is hardly random or haphazard, however (Chen et al., 2009; Lambiotte & Panzarasa, 2009). While epistemic communities typically lack formal governance structures, organized development has been observed in the community's social (Newman, 2001), semantic (Callon, Courtial, Turner, & Bauin, 1983; van Raan, 2000), and citation structures (de Solla Price, 1965; Small, 2006).

Are the organized features that emerge at these different levels of the community related? For example, does the formation of epistemic community social structure influence the ways in which its members choose conceptual ideas to investigate and communicate new knowledge? While substantial work has been performed in describing and modeling growth and change in the social and semantic features of epistemic communities, little research has provided meaningful theoretical explanations or concrete mechanisms by which these developments might occur (Chen et al., 2009).

This paper argues for one particular theoretical mechanism – namely, that features of a community's collective social structure can encourage or dissuade conceptual *retention* processes (Campbell, 1965). That is, social structures can vary the inducements for members of the community to rely on concepts that are broadly recognized and defined for a large portion of that community. When communities are structurally cohesive, members are rewarded for investigating phenomena and communicating claims in standard, well-recognized terms. By contrast, when a community is loosely structured or fractured into many sub-communities, members have a greater incentive to consider idiosyncratic terms or to invent concepts.

The Organizational Communication Challenge of Knowledge Discovery

Cooperation is of substantial practical importance to the discovery and validation of new knowledge. By pooling knowledge, including methods, discoveries, and insights, members of an epistemic community can increase their exposure to new observations as well as obtain expert critiques and analyses of their ideas (Boyd & Richerson, 1985).

These benefits are not without cost, however. For communities to be sustainable, the efforts to cooperate must be met with cooperation in turn (Boyd & Richerson, 1992; Campbell, 1994). When members of a community make efforts to contribute to the community, they diminish the resources available for their own use. Thus, there must be some mechanism by which the community eventually makes available *at least* these resources to those members that have made contributions, or the community will not survive (Boyd & Richerson, 1992).

For most epistemic communities, this resource is new knowledge. Members contribute efforts, such as reviews, insights, or tests of others' work, and in exchange hope to receive useful new knowledge that stems from the work of those to whom they contribute (Latour & Woolgar, 1986; Sandstrom, 2001). Knowledge has many of the features of a public good, in particular the non-excludability of use for those who have obtained it (Arrow, 1962; Fulk, Flanagin, Kalman, Monge, & Ryan, 1996). This makes it a particularly useful resource for cooperation. The discovery of useful knowledge regarding a phenomenon can lead to more precise measurement or better instrumentation, accelerating the discovery of future phenomena (Latour & Woolgar, 1986). Discoveries in one area also may provide a theoretical framework that helps to more precisely guide studies in related areas, leading to new discoveries (Lakatos, 1970). If members communicate useful new knowledge to the community, and others use this contributed knowledge to discover further knowledge and communicate this to the community, a self-sustaining feedback loop should occur and the community's efforts will be sustained (Contractor, 1994).

There is no guarantee that positive feedback loops will emerge, however. If members communicate claims that are not useful, the resources spent on supporting and testing these claims are largely wasted, and the community will struggle to survive. Knowledge can fail to be useful in several ways. First, claims can lack *relevance* (Kitcher, 1995; Sperber & Wilson, 1986). A new proposition might not be related to the questions, methods, or other concerns of the rest of the community.

A claim also can be *unreliable*. That is, it may not consistently describe phenomena for different community members at the places and times when they wish to apply it. For example, propositions whose predictions cannot be replicated by others are of limited use insofar as other members of the community cannot use them to guide or test their own inferences (Collins, 1985).

A related threat to a community is the production of *untestable* claims (Popper, 2003 [1959]). The testability of a claim is the number of observations, or more broadly, the resources required to obtain the observations necessary to decide whether a claim is reliable. A claim that cannot be easily tested is of limited use to a community, even if it is reliable. The community only recoups its investment after tests have been administered and only in those cases where the tests are passed. Since there is no way to know in advance which claims will be supported by future evidence (Jaynes, 2003), it is difficult to expend resources examining untestable claims in an informed, sustainable way.

A final threat is the *uncertainty* that community members face in deciding whether and how to investigate novel ideas and phenomena (Arrow, 1962; Campbell, 1960). New knowledge is, by definition, somewhat unexpected and unpredictable (March, 1991). For an epistemic community to sustain itself, its members must first invest their own efforts to make discoveries. Since members are unsure of how much must be invested before they will identify a relevant, testable, and ultimately reliable proposition, they require inducements that limit this uncertainty (Pfeffer, 1993).

Conceptual choice

These threats occur along independent dimensions. That is, a proposition that is relevant is not necessarily testable or reliable (Popper, 2003 [1959]). Thus, the set of knowledge propositions that meet all of these criteria is substantially smaller than the set that meet any single one. Many scientific hypotheses do not find consistent support, particularly in the early stages of the development of an idea (Dunbar, 1999). Often many relevant propositions must be tried and tested before a sound, reliable one is discovered (Campbell, 1960). Similarly, the set of reliable knowledge propositions that bear on a specific topic is limited. Of these, only a subset can be tested and so forth.

Thus, members of epistemic communities face a problem of where to direct their resources and attention. The decision can be characterized as a resource *foraging* problem (Pirolli & Card, 1999). In foraging theory, organisms search for food within particular patches where they have reason to believe it will be available. Similarly, members of an epistemic community can search conceptual patches that are relevant, testable, or reliable. Foragers do not search over the entire set of patches to find the richest one. Rather, they search within a patch until they find sufficient resources to sustain their needs or receive information that suggests searching in another patch will be more fruitful.

Community members must balance risks in searching for and assessing knowledge claims (Klahr & Simon, 1999). As investigators searching for new propositions, community members must balance the likelihood of finding a claim that others will find relevant with one that they can test, and then invest the care to select those propositions which turn out to be reliable. As evaluators who judge the acceptability of the claims others bring to the community, members must also weigh this trade-off. Relevant but untestable claims are expensive and risky for the community, but may provide more resources if found to be reliable. Testable but irrelevant claims are inexpensive to examine but may not yield a contribution; thus, the resources spent in examination are wasted.

This tension is partly manifested in the concepts that members use to investigate and communicate their claims. In particular, the extent to which claims are relevant and testable will vary with the extent to which the concepts in which they are articulated are widely recognized and clearly defined (Margolin, 2012). The more widely recognized the concepts, the more relevant the claims to the investigation of others. As Callon et al. (1983) describe, some claims use concepts that are so widely recognized and important to a field that others are compelled to respond to the novel finding. Similarly, the more clearly concepts are defined, the more testable are claims which rely on them (Margolin & Monge, 2012, May). When claims comprise vague concepts for which different individuals recognize different entities, many tests may be required to resolve whether the claims are reliable (Hampton, 2007; Popper, 2003 [1959]).

Concepts tend to cluster together as ideas are frequently considered together (Roth & Cointet, 2010). Authors of knowledge claims can use these clusters to make foraging decisions.

They can choose to communicate claims strictly within the space of concepts that are widely recognized with clear definitions. Doing so increases the chances that the propositions they discover will be relevant and testable, thereby minimizing the risk that their claim will be rejected by others who evaluate their work. In evolutionary theory, the reliance on established components that have worked in the past is described as *retention* (Campbell, 1965; Monge, Heiss, & Margolin, 2008). A species which draws its traits primarily from its parents, with low rates of mutation or change, favors retention over variation (Nanay, 2002).

By relying on retained concepts, however, authors take the risk that no reliable new propositions can be found using relevant concepts and means of testing. In evolution, the new environment that an organism or organization confronts may not fit its established repertoire of traits (Carroll & Hannan, 2000; Kauffman, 1993). Similarly, many novel phenomena do not conform precisely to the conceptual rules and definitions that have been established and disseminated within the community (Johnson, 2003).

In response, authors can loosen these constraints by considering other concepts, such as those known only to them or to close colleagues. By expanding the set of terms and procedures with which to investigate and articulate claims, the chances of identifying reliable patterns increase. These claims, however, will likely only be relevant and testable for the subset of community members that recognize and understand these concepts.

The knowledge claims that the community accepts should reflect these foraging decisions. If community members search for propositions using particular concepts, or restrict the acceptance of claims to those that use particular concepts, these are the concepts that will appear in the community's communication.

Social Structure and Conceptual Choice

These foraging decisions will be influenced by both external phenomena – the subject of the community's inquiry – and social processes. The reliability of a claim is influenced by its ability to describe a stable, underlying phenomenon. Claims that do not match stable phenomena will lead to new searches (Dunbar, 1999; Popper, 2003 [1959]). The phenomenon is silent, however, on *where* (in conceptual space) these searches should be conducted (Lakatos, 1970; Popper, 2003 [1959]). The decision regarding where to look for an explanation, or how to re-organize knowledge in response to a failed prediction, is influenced by social considerations (Klahr & Simon, 1999; Kuhn, 1996 [1962], Latour & Woolgar, 1986).

The social structures of communities influence how information diffuses and how normative rules are interpreted and enforced (Coleman, 1988), each of these can have an impact on the effectiveness of different conceptual foraging strategies.

Links between members of a community can assist in the diffusion of information (Rogers, 2003), which in turn can encourage the accumulation of shared background knowledge (Strang & Meyer, 1993). Shared background knowledge has been shown to influence the ways in which individuals apply concepts to situations (Murphy, 2002; Murphy & Medin, 1985). However, many phenomena and situations are not matched perfectly by only a single concept. When definitions are ambiguous or many concepts might apply, background knowledge is used to judge the relevance of different features of the situation to decide which concepts are most appropriate (Murphy & Medin, 1985; Wisniewski, 1996).

Shared background knowledge allows informally or imprecisely defined concepts to be used with minimal loss of testability. In these circumstances, entities for which there is no known term are more likely to be described and recognized in consistent ways (Clark & Wilkes-Gibbs, 1986). Thus, when individuals in communities share background knowledge, the space of concepts that are relevant and testable is expanded. For example, Uzzi (1997) reports that within structurally cohesive communities, individuals can rely on one another to judge generic terms like how something is “supposed to look” in a consistent manner, even though these terms have no precise definition outside of the community’s context (p. 46).

Cohesive social structures expand the set of concepts that are relevant and improve the testability of claims. Shared background knowledge enables terms that are not formally or unambiguously defined to behave, at least within the community, as though they were. This expansion in the set of concepts makes foraging in this space relatively more attractive. That is, members of these communities have larger sets of terms for which they can be confident that, should they find reliable, novel propositions, others will recognize their relevance and find them reasonably easy to test.

As another outcome of a community’s social structure, norms are also more likely to be enforced within cohesive communities (Burt, 2005; Corten & Buskens, 2010). Within conceptual paradigms in epistemic communities, it is not uncommon for those considered to be deviant in their practices or interpretations to be ostracized (Kuhn, 1996 [1962]; Latour & Woolgar, 1986). Coleman (1988) argues that cohesive social communities provide the threat of consistent punishment by increasing the probability that information regarding normative transgressions is passed along to a large number of community members (Burt, 2005). Cohesive structures also lead to an increased sense of mutual accountability (Uzzi, 1997). Individuals with bonds to others in cohesive communities are less likely to try to gain advantages for themselves at the community’s expense (McGinn & Keros, 2002).

Thus, members of cohesive communities can also expect their claims to be interpreted consistently and with few unanticipated mistakes (Pfeffer, 1993). That is, within cohesive communities where normative enforcement is easily executed, individuals will be discouraged from incorrectly applying concepts, including in tests of novel claims. This reduces the uncertainty faced by investigators. Kuhn (1996 [1962]) argues that when scientists in established paradigms get results inconsistent with the community's expectations, the community often concludes that the scientist is at fault. Authors of new claims can rely on this as a safeguard against careless misinterpretations appearing to reject their work.

By contrast, when social structure is disconnected, authors that forage exclusively in the concepts established in the prior literature are at a disadvantage. Without shared background knowledge, the space of concepts in which reliable and testable claims can be formed is limited. Thus, there is a greater risk that no reliable claim can be found. At the same time, even established concepts may not be interpreted consistently or recognized by all members as relevant. Within distinct fields, the same term can be used with substantially different meanings (Chavalrias & Cointet, 2008; Ghaziani & Ventresca, 2005). Without access to cohesive structures that enforces normative consistency, different individuals may adopt their own definitions of important concepts without others being aware. New ideas often come from outside the dominant paradigm and concepts (Burt, 2005; Sandstrom, 2001). When social structures are fractured, there is little reason to rely on established concepts as little is gained from accepting their restrictive perspective. Members can thus broaden their foraging space in search of reliable novel knowledge.

Signatures of Social Structure and Conceptual Retention

Social and semantic structures can be captured with a myriad of constructs, including the rather abstract idea of *conceptual retention.* What does it mean for individuals to favor the use of established concepts with clear definitions? One way is to consider the extent to which the distributions of concepts used in community communication are consistent over time. When members rely on well-established concepts, they are likely to use dominant concepts, those that have appeared most frequently in the literature in the past (Callon et al., 1983; Steyvers & Tenenbaum, 2005), while relegating infrequently used concepts to more minor roles (Ferrer i Cancho & Sole, 2001). When conceptual definitions are retained, however, dominant concepts will be more constrained in their future use by the fact that they have been so frequently used already in the past (Carroll & Hannan, 2000; Margolin & Monge, 2012, May). By contrast, when conceptual definitions are free to vary or be adjusted to new phenomena, some dominant concepts may take on novel definitions for different members of the community, thereby increasing their usage (Chavalrias & Cointet, 2008; Ghaziani & Ventresca, 2005).

Thus, the similarities of concept usage distributions at future points in time to those at previous points in time suggest that members are favoring established concepts and using them with similar definitions. By contrast, differences in concept-usage distributions from one point in time to another suggest that members are using concepts that were less established in the community in the past and/or are using established concepts with novel definitions.

Structural cohesion can also be captured in different ways. A standard conceptualization is the proportion of authors in the field who can reach one another by virtue of being connected to the largest or *main component* (Moody, 2004). A component of a network is a cluster of nodes which have at least one path of any length connecting each node to each other node (Wasserman & Faust, 1994). The main component of a network is the largest component in the graph.

Figure 1 provides an illustration of two networks with different levels of structural cohesion. These networks show the social network structure for the epistemic community concerning the topic d-branes, a research area in physics, for 1999 through 2001 (Figure 1a) and 2006 through 2008 (Figure 1b). The first network shows a large main component in which many individuals are connected. The main component contains 55% of this network, indicating a relatively cohesive structure. The second network shows a more fractured structure, with the largest component containing only 30% of the total individuals in the network. Several of the smaller components are relatively large and appear to approach the largest one in size.

Large main components have been noted in a variety of studies of epistemic communities. Newman (2001) finds that the portion of the authors contained in the main component is very large—approaching and exceeding 90% for many fields—but is variable over time, a result supported by Velden, Haque, and Lagoze (2010). Guimera, Uzzi, Spiro, and Amaral (2005) find large main component percentages in scientific fields (astronomy), social scientific fields (economics), and artistic communities (the production of Broadway musicals). Moody (2004) finds that the main component percentage for sociology is large, but not nearly as large as for natural science (approximately half of the total authors), and is actually less than would be expected by chance, given network density.

As argued above, it is hypothesized that structural cohesion will be associated with conceptual retention. Concretely, this suggests:

> H1: At a particular point in time (t), the extent to which the distribution in past concept usage (through time t) is similar to the distribution in future concept usage (at t+1) will be positively associated with the percentage of authors in the main component of the community at that point in time (t).

Another way to measure the cohesiveness of the community is by considering the degree to which nodes connect only to those in their local clusters or have ties to nodes in other clusters. The extent of such "local social clustering" can be measured using the clustering coefficient developed by Watts and Strogatz (1998). This coefficient is an inverse measure of cohesion. When local social clustering is high, individuals do not possess many more direct than indirect ties (at a distance of greater than 1). For networks that are low or moderate in density, a high clustering coefficient is indicative of a fractured social structure in which all or most of an individual's ties are to others that reside in the same region of the network.

The networks in Figure 1 also show a disparity in local clustering. The network in Figure 1 (a) shows a clustering coefficient of .713 while the network in Figure 1 (b) shows a clustering coefficient of .793. As argued above, it is hypothesized that structural cohesion will be associated with conceptual retention. Concretely, this suggests:

> H2: At a particular point in time (t), the extent to which the distribution in past concept usage (through time t) is similar to the distribution in future concept usage (at t+1) will be negatively associated with the clustering coefficient among authors in the community at that point in time (t).

Method

Data

Topic identification. It was important to analyze more than one epistemic community in this study. Many studies of epistemic communities focus only on communities that grow in size and quantity of communication (published papers) over time (Börner et al., 2010; Mane & Börner, 2004) which makes it difficult to distinguish effects due to changes in community structures and effects due to the maturity of the field. By including topics that showed both growth and decline in claim production the influence of this co-variation with time is reduced.

Candidate topics were examined for the number of papers published each year as reported by ISI's Web of Science (WoS) database. Aided by the analysis of Small (2006), the topic of "d-branes" was selected. Research on d-branes began in 1995 with the publication of a single paper (Randall, 2005). This research reached its peak in 2000 with the publication of 478 papers, after which production steadily declined through 2010, in which only 167 papers were published (4228 papers in total). A topic that was roughly comparable in size and field was then sought. The topic "supermassive black holes" was selected because it showed consistent growth over a similar time period but did not grow to be substantially larger than d-branes. According to WoS records, research into supermassive black holes began in the 1970s, but the community did not produce years with a substantial number of papers until the 1990s. Beginning in 1991, three papers were published on this topic. Research grew in most years, with 188 papers published on the topic in 2010 (1300 papers in total during the 1991-2010 period).

Data acquisition. For each topic identified, data were acquired via a search in the "topic" field in the WoS query form (Börner et al., 2010). The search results were then refined to include only journal articles. The WoS data were loaded into the Network Workbench (NWB) Tool Version 1.0 (Börner et al., 2010; NWB_Team, 2006). Network Workbench was used to extract co-authorship networks and to count the number of papers in which each concept was used.

Identifying authors. Some authors use their middle initials inconsistently across their publications. This practice leads to ambiguity as the same author may be identified by two different unique names (e.g., Margolin, D. and Margolin, D. B.). In this study, authors who shared a common first initial and last name were considered to be the same individual unless there was evidence to the contrary, that is, if authors had different middle initials, or if three authors sharing the same first initial but two having different middle initials making the case ambiguous (e.g., Margolin, D.; Margolin, D. B.; Margolin, D. S.). This method identified 2192 authors for d-branes and 2803 authors for supermassive black holes over the entirety of the observation period for each.

Identifying concepts. There is no single perfect coding system for textual data as there is no guarantee of a fixed meaning for a particular word (Lakoff, 1987). Since the consistency of conceptual definition is an important variable, all words written in the same way are treated as instances around which definitions can vary to a greater or lesser degree. It is also common to delete from semantic analysis "stop" words (Carley, 1997), or words like "a," "the," and "and," which are highly generic and have no specific, community-based meaning. Concepts were also "stemmed," meaning the suffixes that indicate tense were removed so that only the root concept was considered.

Measures

Co-authorship networks are one-mode author-author networks constructed from two-mode author-paper networks (Chavalrias & Cointet, 2008; Newman, 2001; Wasserman & Faust, 1994). Each co-authorship link between author *i* and author *j* is described by values (*pub*, *t*), where *pub* is the number of papers that the two authors published together in a particular year and *t* is the year in which these publications took place. A co-authorship network at time t is then given by: pub_{ijt}.

Author networks were constructed for rolling three-year windows (times t-2 through time t). The rolling three-year network is given by:

$$\sum_{t-2}^{t} pub_{ijt}$$

for all *i* and *j* at a particular *t*. For example, the rolling network for t = 1999 includes all co-authorship links among authors who published papers together from 1997 through 1999. A 3-year rolling network was chosen to balance the errors that might be introduced by more extreme methods. On the one hand, the influence of social relationships is expected to decay over time (Burt, 2000). Thus, networks that cover a time window that is too long are likely to overstate the meaningful social relationships in the community. On the other hand, co-authorships are just one way to measure meaningful social contact (Crane, 1969). Thus, single year networks are likely to understate the connections in the underlying social community.

Research suggests that the typical half-life of a research front is about three to five years (de Solla Price, 1965; Small, 2006). That is, about half or more of papers lose their relevance within 3-5 years of their publication. In addition, a three-year basis keeps the overlap between the networks at different points in time to a manageable level. Though each network and its structural features are unique results of the particular combination of nodes and links in the network, time windows longer than three years mean that networks will change.

Main component percentage. The main component of a network is the largest component in the graph – the largest cluster of nodes which have some path connecting one another. The proportion of authors in the main component compares the size of the main component to the size of the network as a whole in terms of the number of nodes in each. Thus, this percentage or "main component pct" is calculated as follows:

$$\text{main component pct} = \frac{\text{\# of nodes in main component}}{\text{\# of nodes in entire network}}$$

Local social clustering. In an undirected network such as a co-authorship network, a closed triangle is any set of three nodes i, j, and k where there is at least one co-authorship link *i-j, i-k*, and *j-k*. A connected triple is any set of three i, j, and k where there is at least two links between them, such that all three nodes can reach one another via a path among this set of three nodes (for example: links between *i-j,* and *i-k*, means that k and j are connected, though they do not share a link). The proportion of the connected triples that are closed is then represented by the Watts and Strogatz clustering coefficient:

$$\text{Local social clustering} = \frac{\text{closed triangles}}{\text{connected triples}}$$

Similarity in conceptual distribution. Concepts were coded for the number of papers in which they appeared in each year. The frequency with which a concept appears in published papers is a count variable. To compare distributions of a count variable, it is recommended to use the Poisson distribution (McCullagh & Nelder, 1989).

In this study, the distributions in conceptual usage were compared by considering three measures of a community's past distribution of concept usage: the number of papers in which concepts had been used in the community in the past (1995 through t for d-branes; 1991 through t for supermassive); the square of this number (1995 through t; 1991 through t); and the frequency with which concepts had been used in papers published in the prior year (t).

These measures were used to capture three different aspects of the distribution. The total number of papers in which a concept has been used in the past is a measure of its establishment in the community. The square of a concept's cumulative frequency captures the extent to which concepts that have been used with high frequency may reach their carrying capacity (Carroll & Hannan, 2000; Margolin & Monge, 2012, May; Monge et al., 2008). When a concept has been used very frequently in the past, there may be fewer situations to which the concept easily applies, at least by its original definition. Thus, when community members stick to the original definitions for concepts, as predicted for socially cohesive communities, the squared term will improve the fit of the past distribution to the future distribution. Finally, the frequency in the most recent period (t) is used to capture the influence of recent developments or the emergence of new topics in the field.

For each time t these frequencies were calculated and a Poisson regression was performed on usage at time t+1 with usage at time t as a predictor. The R-squared value (McFadden's estimate) for each regression was then used as a measure of similarity for the year (t) to which it corresponded.

Analysis

All hypotheses were tested using OLS regression with a first-order lag of the dependent variable to control for autocorrelation (Hamilton, 1994), which can bias the estimate of standard errors. The residuals of the regression are then tested for additional auto-correlation (Ostrom, 1990). Results of explained variance then include a reference to the portion of variance explained by the generic, auto-correlation variables captured by the lagged dependent variable and the additional variance explained by the other predictor variables (Monge, Cozzens, & Contractor, 1992).

Auto-correlation of residuals was tested for at lag order 3 using the LaGrange Multiplier F-test. That is, a regression was run on each residual value with the residuals from the previous periods (up to three periods prior) serving as predictors. The F-test examines significance in the regression as a whole against a null hypothesis of no autocorrelation. All regressions were run in Gretl 1.9.4 (Cottrell & Lucchetti, 2011).

Results

Hypothesis 1 predicted that the extent to which the prior distribution of conceptual uses (through time t) predicted the future distribution of conceptual usage (at time t+1) would be associated with the percentage of authors that were in the main component of the social network (at time t). Table 1, model 2, shows the results for research on the topic of d-branes. The bivariate association between the portion of authors in the main component pct (t-2 through t) and conceptual retention (t vs. t+1) is significant and in the predicted direction (B = .28, t = 4.96, p < .001). The model shows a substantial improvement in R-squared over the auto-correlation model (R^2 increase = .35).

Table 1, model 5, shows the results for research on the topic supermassive black holes. Here the association between the portion of authors in the main component pct (t-2 through t) and conceptual retention (t vs. t+1) is not significant nor in the predicted direction (B = -0.03, t = -

0.86, p = .40). The model shows no improvement in R-squared over the auto-correlation model (R^2 increase = .001).

Taken together, these results do not provide support for hypothesis 1. Though the hypothesized effect appears present in the d-branes topic, there is no evidence of its operation for supermassive black holes.

Hypothesis 2 predicted that the extent to which the prior distribution of conceptual uses (through time t) predicted the future distribution of conceptual usage (at time t+1) would be negatively associated with the local clustering coefficient (at time t). Table 1 (model 3) shows the results for research on d-branes. The association between the local clustering coefficient (t-2 through t) and conceptual retention (t vs. t+1) is significant and in the predicted direction (B = -1.05, t = 9.90, $p < .001$). The model shows a substantial improvement in R-squared over the auto-correlation model (R^2 increase = .62). Table 1 (model 6) shows the results for research on supermassive black holes. The association between the proportion of authors in the main component pct (t-2 through t) and conceptual retention (t vs. t+1) is in the direction hypothesized though not statistically significant at the $p < .05$ level (B = -0.47, t = -1.69, p = .114). The model shows no improvement in R-squared over the auto-correlation model (R^2 increase = .003).

Taken together, these results provide partial support for hypothesis 2. The results show the relationship to be in the predicted direction across topics, but the effect sizes and significance are quite different.

Conclusion

This study proposed a mechanism through which the structure of a community's social network can influence the communication choices that members of the community will make. It was argued that a community's social structural cohesiveness would be associated with its conceptual retention. The study found partial support for the relationship between the local clustering coefficient and conceptual retention (hypothesis 2), but did not find consistent support for the relationship between the portion of the authors in the main component and conceptual retention (hypothesis 1). The main component pct did show a significant relationship to conceptual retention in one of the two topics (d-branes), providing some evidence for the hypothesis.

These results suggest two basic conclusions. Despite the small sample sizes, hypothesis 2 is significant for d-branes and is close to significant for supermassive black holes. When sample sizes are small, power is lost, increasing the possibility of Type II error. The fact that the effect was found in two different samples thus suggests that further investigation is warranted (Weber, 2007). In particular, the large effect found for d-branes, in which local clustering explained a large portion of the variance beyond that explained by auto-correlation, is evidence that this hypothesis merits further examination. Because local clustering in the social network is a global property of the network as a whole – which was created from the aggregation and intersection of individual collaboration efforts made by every member of the community – it is difficult to explain how it might correspond to other variables without participating in some direct mechanism. The precise nature of this mechanism is unclear, however, and warrants further research.

Another conclusion stems from the failure of hypothesis 1 to obtain full support: to the extent to which there is an underlying mechanism connecting social structure to conceptual retention, its presence is sensitive to the manner in which community structure is measured. As described above, the proportion of authors in the main component is a somewhat imprecise measure of community structure but it is the one that researchers studying epistemic communities have tended to emphasize, in part because of the ease with which it can be calculated for large networks (Newman, 2001). The evidence in this study calls for the use of more fine-grained measures. In particular, cases in which small sub-networks are unconnected to one another, such that there is no single large component, may show structural changes that are indicative of community cohesion. Similarly, network structures that evolve within large, stable main components may provide useful insights.

As with any research undertaking, this study's findings need to be considered in light of three limitations. The first limitation is the quantity of observations. Even though the data gathered for the study spanned several thousand papers written in two distinct topics in physics, measuring the community at the aggregate reduced these large datasets to two small time series, each with fewer than 20 points in time. Both topics were also drawn from the same general field: physics, so that differences due to institutional factors might be minimized. The conclusions that can be drawn from these data are thus limited until future work is done in this area.

A second limitation is in the narrowness of measures for both the dependent and independent variable. This limitation is a corollary of the first. Because only a small number of observations were available at the community level, it would not have been appropriate to test a large number of distinct measures of structural cohesion nor conceptual retention. This limitation can also be addressed through the collection of data on additional topics.

Finally, there is the limitation inherent in attempting to explain individual or small-group behavior from macro-level observations and mechanisms (Monge & Contractor, 2003). In particular, while the theoretical mechanisms proposed in this study do successfully predict a portion of the evidence observed, these are not the only plausible explanations. It is also possible that intermediate mechanisms could play a more direct causal role. For example, academic conferences which enabled authors to meet one another might have both encouraged structural cohesion and improved the diffusion and acceptance of common terms. Nonetheless, the evidence provided in this study provides some indications of the kinds of intermediate mechanisms which might operate.

Given the study's findings and limitations, further research should proceed broadly along two lines. First, attempts to replicate the present study in a larger and more diverse set of communities would be a substantial contribution. Does structural cohesion predict conceptual retention across epistemic communities, as the theory predicts? If not, are there specific communities in which the effect is strong and others in which it is weak? In addition, are there particular times or qualities in the development of a community in which this theoretical mechanism is strengthened or weakened? These questions could be addressed with both quantitative and qualitative analysis of a larger set of communities.

A second direction for further research is to more closely investigate the mechanism itself. It is possible, for example, that structural cohesion is also the result of conceptual retention. That is, as communities establish as shared vocabulary, members find it easier to collaborate across larger social distances, bringing the community together. In the current study, future retention, controlling for past retention, is predicted by past social structure, so it is unlikely that the observed effect can be explained by this "reversed" phenomenon alone. Nonetheless, further research might attempt to tease apart this distinction.

References

Arrow, K. (1962). Economic welfare and the allocation of resources for invention. In N. B. o. E. Research (Ed.), *The rate and direction of inventive activity: Economic and social factors* (pp. 609-626): UMI.

Ashby, W. R. (1965). *An introduction to cybernetics*. New York: John Wiley.

Börner, K., Huang, W., Linnemeier, M., Duhon, R. J., Phillips, P., Ma, N., . . . Price, M. A. (2010). Rete-netzwerk-red: analyzing and visualizing scholarly networks using the Network Workbench Tool. *Scientometrics, 83*, 863-876. doi: :10.1007/s11192-009-0149-0

Boyd, R., & Richerson, P. (1985). *Culture and the evolutionary process*. Chicago: University of Chicago Press.

Boyd, R., & Richerson, P. J. (1992). Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology, 13*, 171-195. doi: 10.1016/0162-3095(92)90032-Y

Burt, R. S. (2000). Decay functions. *Social Networks, 22*, 1-28.

Burt, R. S. (2005). *Brokerage and closure: An introduction to social capital*. New York:Oxford University Press.

Callon, M., Courtial, J. P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information Sur Les Sciences Sociales, 22*, 191-235.

Campbell, D. T. (1960). Blind variation and selective retention in creative thought as in other knowledge processes. *Psychological Bulletin, 67*(6), 380-400. doi: 10.1037/h0040373

Campbell, D. T. (1965). Variation and selective retention in socio-cultural evolution. In H. R. Barringer, G. I. Blanksten & R. W. Mack (Eds.), *Social change in developing areas: A reinterpretation of evolutionary theory* (pp. 19-48). Cambridge, MA: Schenkman.

Campbell, D. T. (1994). How individual and face-to-face-group selection undermine firm selection in organizational evolution. In J. A. C. Baum & J. V. Singh (Eds.), *Evolutionary dynamics of organizations* (pp. 23-28). New York: Oxford.

Carley, K. M. (1997). Extracting team mental models through textual analysis. *Journal of Organizational Behavior, 18*, 533-558.

Carroll, G. R., & Hannan, M. T. (2000). *The demography of corporations and industries*. Princeton: Princeton University Press.

Chavalrias, D., & Cointet, J. (2008). Bottom-up scientific field detection for dynamical and hierarchical science mapping, methodology and case study. *Scientometrics, 75*, 37-50.

Chen, C., Chen, Y., Horowitz, M., Hou, H., Liu, Z., & Pellegrino, D. (2009). Towards an explanatory and computational theory of scientific discovery. *Journal of Informetrics, 3*, 191-209. doi: 10.1016/j.joi.2009.03.004

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*, 1-39.

Coleman, J. S. (1988). Social capital in the creation of human-capital. *American Journal of Sociology, 94*, S95-S120.

Collins, H. M. (1985). *Changing order: Replication and induction in scientific practice*. Beverly Hills, CA, Sage.

Contractor, N. (1994). Self-organizing systems perspective in the study of organizational communication. In B. Kovačić (Ed.), *New approaches to organizational communication* (pp. 39-66). Albany, NY: SUNY Press.

Corten, R., & Buskens, V. (2010). Co-evolution of conventions and networks: An experimental study. *Social Networks, 32*, 4-15.

Cottrell, A., & Lucchetti, R. (2011). Gretl user's guide: Gnu regression, econometrics, and time-series.

Crane, D. (1969). Social structure in a group of scientists: A test of the "invisible college" hypothesis. *American Sociological Review, 34*, 335-352.

de Solla Price, D. J. (1965). Networks of scientific papers. *Science, 149*(3683), 510-XXX.

Dunbar. (1999). How scientists build models: InVivo science as a window on the scientific mind. In L. Magnani, N. Neressian & P. Thagard (Eds.), *Model-based reasoning in scientific discovery* (pp. 89-98): Pienum Press.

Ferrer i Cancho, R., & Sole, R. V. (2001). The small world of human language. *Proceedings of the Royal Society B-Biological Sciences, 268*, 2261-2265.

Fulk, J., Flanagin, A., Kalman, M., Monge, P. R., & Ryan, T. (1996). Connective and communal public goods in interactive communication systems. *Communication Theory, 61*, 60-87.

Ghaziani, A., & Ventresca, M. J. (2005). Keywords and Cultural Change: Frame Analysis of Business Model Public Talk, 1975–2000. *Sociological Forum, 20*(4), 523-599. doi: 10.1007/s11206-005-9057-0

Guimera, R., Uzzi, B., Spiro, J., & Amaral, L. A. N. (2005). Team assembly mechanisms determine collaboration network structure and team performance. *Science, 308*(5722), 697-702. doi: 10.1126/science.1106340

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Hampton, J. A. (2007). Typicality, graded membership, and vagueness. *Cognitive Science, 31*, 355-383. doi: 10.1080/15326900701326402

Jaynes, E. T. (2003). *Probability theory the logic of science.* Cambridge: Cambridge Univeristy Press.

Johnson, C. V. (2003). *D-Branes*. New York: Cambridge University Press.

Kauffman, S. (1993). *The origins of life: Self-organization and selection in evolution*. New York: Oxford University Press.

Kitcher, P. (1995). *The advancement of science*. New York: Oxford University Press.

Klahr, D., & Simon, H. A. (1999). Studies of scientific discovery: complementary approaches and convergent findings. *Psychological Bulletin, 125*, 524-543.

Kuhn, T. S. (1996 [1962]). *The structure of scientific revolutions* (3rd ed.). Chicago: University of Chicago Press.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91-196). New York: Cambridge University Press.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago University of Chicago Press.

Lambiotte, R., & Panzarasa, P. (2009). Communities, knowledge creation, and information diffusion. *Journal of Informetrics, 3*, 180-190. doi: 10.1016/j.joi.2009.03.007

Latour, B., & Woolgar. (1986). *Laboratory life: The construction of scientific facts*. Ewing, NJ: Princeton University Press.

Mane, K. K., & Börner, K. (2004). Mapping topics and topic bursts in PNAS. *Proceedings of the National Academy of Sciences of the United States of America, 101*(Suppl 1), 5287.

March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science, 2*, 71-87.

Margolin, D., & Monge, P. (2012, May). *Conceptual selection in epistemic communities*. Paper presented at the meeting of the International Communication Association, Phoenix.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd. ed.): Chapman & Hall/CRC.

McGinn, K. L. and A. T. Keros (2002). Improvisation and the logic of exchange in socially embedded transactions. *Administrative Science Quarterly*, 47, 442-473.

Monge, P., & Contractor, N. (2003). *Theories of communication networks*. New York: Oxford University Press.

Monge, P., Cozzens, M. D., & Contractor, N. S. (1992). Communication and motivational predictors of the dynamics of organizational innovation. *Organization Science, 3*, 250-274.

Monge, P., Heiss, B. M., & Margolin, D. B. (2008). Communication Network Evolution in Organizational Communities. *Communication Theory, 18*, 449-477.

Moody, J. (2004). The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociological Review, 69*, 213-238.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Murphy, G. L., & Medin, D. L. (1985). The Role of theories in conceptual coherence. *Psychological Review, 92*(3), 289-316. doi: 10.1037/0033-295X.92.3.289

Nanay, B. (2002). The return of the replicator: What is philosophically significant in a general account of replications and selection. *Biology and Philosophy, 17*, 109-121.

Newman, M. J. (2001). The Structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America, 98*, 404-409.

NWB_Team. (2006). Network Workbench Tool (Version 1.0.0): Northeastern University and University of Michigan. Retrieved from http://nwb.slis.indiana.edu

Ostrom, C. W. J. (1990). *Time series analysis* (Vol. 9). Thousand Oaks, CA: Sage.

Pfeffer, J. (1993). Barriers to the advance of organizational science: Paradigm development as a dependent variable. *The Academy of Management Review, 18*, 599-620.

Pirolli, P., & Card, S. (1999). Information Foraging. *Psychological Review, 106*, 643-675.

Popper, K. R. (2003 [1959]). *The logic of scientific discovery*. New York: Routledge.

Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York: The Free Press.

Roth, C., & Bourgine, P. (2005). Epistemic communities: description and hierarchical organization. *Mathematical Population Studies, 12*, 107-130.

Roth, C., & Cointet, J. (2010). Social and semantic coevolution in knowledge networks. *Social Networks, 32*, 16-29.

Sandstrom, P. E. (2001). Scholarly communication as a socioecological system. *Scientometrics, 51*, 573-605.

Small, H. (2006). Tracking and predicting growth areas in science. *Scientometrics, 68*(3), 595-610. doi: 10.1007/s11192-006-0132-y

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford: Blackwell.

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science, 29*, 41-78. doi: 10.1207/s15516709cog2901_3

Strang, D., & Meyer, J. W. (1993). Institutional conditions for diffusion. *Theory and Society, 22*, 487-511.

Uzzi, B. (1997). Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative Science Quarterly, 42*, 35-67.

van Raan, A. F. J. (2000). On growth, ageing, and fractal differentiation in science. *Scientometrics, 47*, 347-362. doi: 10.1023/A:1005647328460

Velden, T., Haque, A., & Lagoze, C. (2010). A new approach for analyzing patterns of collaboration in co-authorship networks: mesoscopic analysis and interpretation. *Scientometrics*. Retrieved from doi:10.1007/s11192-010-0224-6

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. New York: Cambridge Univ Pr.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*, 440-442.

Weber, R. (2007). Responses to Matsunaga: To adjust or not to adjust alpha in multiple testing: That is the question. Guidelines for alpha adjustment as response to O'Keefe's and Matsunaga's Critiques. *Communication Methods and Measures, 1*, 281-289.

Wisniewski, E. J. (1996). Construal and similarity in conceptual combination. *Journal of Memory and Language, 35*, 434-453. doi: 10.1006/jmla.1996.0024

Table 1. Relationship of Structural Cohesiveness to Strength of Retention

Dependent Variable: Variance in Conceptual Usage Explained by Prior Usage for t = 1997 thru 2009 (n = 13) for d-branes

	Model 1				Model 2				Model 3			
Parameter	B	*t*	*p*	*Sig*	B	*t*	*p*	*Sig*	B	*t*	*p*	*Sig*
Constant	0.43	4.94	<.001	***	0.55	11.68	<.001	***	1.50	16.16	<.001	***
lag)	0.28	1.92	0.08		-0.12	-1.40	0.19		-0.19	-8.16	<.001	***
Main Component Pct (t-2 thru					0.28	4.96	<.001	***				
Local Clustering (t-2 thru t)									-1.05	-9.90	<.001	***
R-squared		0.24				0.55				0.86		
AIC		-54.58				-59.40				-74.57		
LaGrange Multiplier F-test	F(3,8)=	3.67	p = 0.0626		F(3,7)=	0.13	p = 0.938		F(3,7)=	0.22	p = 0.876	

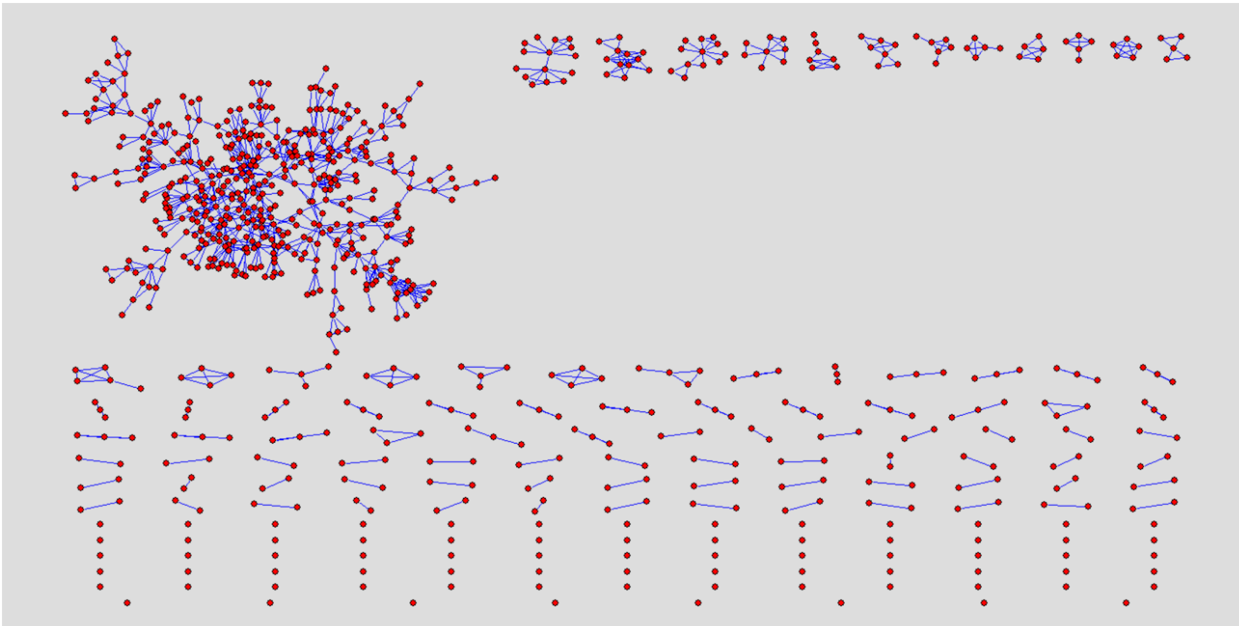
Dependent Variable: Variance in Conceptual Usage Explained by Prior Usage for t = 1993 thru 2009 (n = 17) for supermassive black holes

	Model 4				Model 5				Model 6			
Parameter	B	*t*	*p*	*Sig*	B	*t*	*p*	*Sig*	B	*t*	*p*	*Sig*
Constant	0.05	2.02	0.06		0.05	2.19	0.05	*	0.53	1.81	0.09	
Var Explained (1 year lag)	0.95	21.11	<.001	***	0.98	15.22	<.001	***	0.83	8.91	<.001	***
Main Component Pct (t-2 thru t)					-0.03	-0.86	0.40					
Local Clustering (t-2 thru t)									-0.47	-1.69	0.11	
R-squared		0.95				0.95				0.96		
AIC		-59.98				-58.37				-59.07		
LaGrange Multiplier F-test	(3,12)=	0.58	p = 0.637		(3,11)=	1.53	p = 0.263		(3,11)=	0.27	p = 0.844	

* p < .05, ** p < .01, *** p <.001

Figure 1. Comparing Social Cohesion

(a) More Cohesive Network – Social Structure for d-branes, based on co-authored papers from 1999-2001



(b) More Fractured Network – Social Structure for d-branes, based on co-authored papers from 2006-2008