



# Social centralization and semantic collapse: Hyperbolic embeddings of networks and text<sup>☆</sup>



Linzhuo Li<sup>a</sup>, Lingfei Wu<sup>a,b,c</sup>, James Evans<sup>a,d,\*</sup>

<sup>a</sup> Sociology Department and Knowledge Lab, University of Chicago, 5735 South Ellis Avenue, Chicago, IL 60637 USA

<sup>b</sup> School of Computing and Information, University of Pittsburgh, 4200 Fifth Ave, Pittsburgh, PA 15260 USA

<sup>c</sup> School of Journalism and Communication, Nanjing University, 22 Hankou Rd, Nanjing, 210008 China

<sup>d</sup> Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501 USA

## ARTICLE INFO

### Keywords:

Social networks  
Semantic networks  
Auto-encoders  
Unsupervised learning  
Hyperbolic embedding  
Poincaré disk  
Machine learning  
Science of science

## ABSTRACT

Modern advances in transportation and communication technology from airplanes to the internet alongside global expansions of media, migration, and trade have made the modern world more connected than ever before. But what does this bode for the convergence of global culture? Here we explore the relationship between centralization in social networks and contraction or collapse in the diversity of semantic expressions such as ideas, opinions and tastes. We advance formal examination of this relationship by introducing new methods of manifold learning that allow us to map social networks and semantic combinations into comparable hyperbolic spaces. Hyperbolic representations natively represent both hierarchy and diversity within a system. In a Poincaré disk—a two-dimensional hyperbolic embedding—radius from center traces the position of an actor in a social hierarchy or an idea in a semantic hierarchy. Angle of the disk required to inscribe connected actors or ideas captures their diversity. We illustrate this method by examining the relationship between social centralization and semantic diversity within 21<sup>st</sup> Century physics, empirically demonstrating how dense, centralized collaboration is associated with a reduction in the space of ideas and how these patterns generalize to all modern scholarship and science. We discuss the complex of causes underlying this association, and theorize the dynamic interplay between structural centralization and semantic contraction, arguing that it introduces an essential tension between the supply and demand of difference.

## 1. Introduction

Despite fits of nationalism, temporary tariffs and abortive border restrictions, ‘it’s a small world after all’. Transportation and communication technology from automobiles and airplanes to telephones and the internet have made the modern world more connected than ever before. The march of global media, migration, trade and education across these pathways has gone hand in hand with convergence (Harari, 2014) and polarization (Huntington, 2000) in global culture. The global human population approaches 8 billion, but human languages are going extinct at an accelerating pace. Nearly 50 % of the world’s 7000 languages are on the cusp of being forgotten (Malone, 2005), as global communication becomes centralized around widely used languages like Chinese, Hindi, English, Spanish and Arabic, which reduce the space of ideas that can be natively imagined and shared (Regier, 2008; Whorf & Chase,

\* The authors wish to thank Hauchuan Cui and Shahab Asoodeh for helpful comments and discussion. We acknowledge funding support from AFOSR grants FA9550-15-1-0162 and FA9550-19-1-0354, NSF grant 1158803 and 1829366, and DARPA grant HR00111820006.

<sup>☆</sup> Corresponding author.

E-mail address: [jevans@uchicago.edu](mailto:jevans@uchicago.edu) (E. James).

1956). Music videos from Luis Fonsi's "Despacito" to Psy's "Gangnam Style" are watched billions of times and popular music charts by country manifest little variation, suggesting a global alignment of musical tastes around centralized styles. Similar trends are seen in the consumption of other cultural products. More Nigerians are watching American movies than ever before, and more Ghanans are watching Nigerian films. More books are shared, borrowed and read across more libraries than ever before, leading to a growing global conversation about a collapsing set of ideas.<sup>1</sup>

The same patterns are manifest at individual scales. Classic studies of group conformity have shown that low status social actors often mimic the opinions and behaviors of social elites in order to fit in, shrinking the space of ideas, opinions and expressions (Asch, 1956; Danescu-Niculescu-Mizil, West, Jurafsky, Leskovec, & Potts, 2013). Moreover, 21<sup>st</sup> Century computational studies of communication document linguistic copying to ingratiate and persuade (Danescu-Niculescu-Mizil, Lee, Pang, & Kleinberg, 2012; McFarland, Jurafsky, & Rawlings, 2013), just as sociologists demonstrate how shared cultural tastes influence the selection of new network ties (Lewis & Kaufman, 2018; McPherson, Smith-Lovin, & Cook, 2001). Birds of a feather flock together, just as new social ties channel, correlate and collapse the population of cultural tastes (Rogers, 2010; Rogers & Floyd Shoemaker, 1971; Valente, 1996).

In this paper, we do not attempt to untie the causal knot of society and culture. Increased social connectivity forces semantic forms like language, music and ideas to compete for collective attention, just as increased linguistic and semantic convergence accelerates connection by reducing communicative friction. Instead, here we seek to acknowledge and formalize the deep connection between social centralization and semantic convergence, building on recent attention to the complex duality of social structure and culture (Mohr, 2000; Lee & Martin, 2018; Breiger & Puetz, 2015; Basov & Brennecke, 2017).

We first review recent studies that highlight the relationship between social centralization and semantic diversity. We then introduce a new geometric approach to data analysis that allows us to systematically examine the association between network centralization and the semantic diversity it can sustain in the case of 21<sup>st</sup> Century physics. We map networks of collaborative relationships and co-occurring concepts through manifold learning into comparable, two-dimensional hyperbolic spaces or Poincaré disks. A Poincaré disk can naturally and simultaneously render both hierarchy in networks and diversity in expressions. We illustrate the metrics defined on these spaces with an empirical exploration of shifts in the centralization of collaboration and those in the diversity of topics and ideas investigated. We show how increased density and centralization in coauthorship ties is associated with sharp reductions in the diversity of ideas. Next, we show how these patterns generalize to all modern science and scholarship. We conclude by theorizing the dynamic interplay between social expansion and semantic contraction and consider its implications for multicultural societies and the maintenance of diversity.

## 2. Centralized networks and semantic diversity

Recent research has experimentally and observationally identified a tight relationship between network structure and idea diversity (1991, Carley, 1986; DiMaggio, 1987; Erickson, 1988; Kandel, 1978; Krackhardt & Kilduff, 1990; Krohn, 1986; Mark, 1998). Research on the wisdom of crowds has shown that crowds appear wiser than individual participants in stock markets, political elections, and quiz shows (Surowiecki, 2004), but social centralization undermines crowd wisdom by reducing independence (Lorenz, Rauhut, Schweitzer, & Helbing, 2011). One experiment involving a complex estimation task demonstrated that centralized networks tended to decrease the diversity of crowd opinions away from the truth, while decentralized networks decreased diversity toward it (Becker, Brackbill, & Centola, 2017). Scholars have also examined these patterns naturalistically in the epistemic culture of biomedicine (Cetina, 2009). Our own recent research extracted claims about drug-gene interactions from thousands of papers in the biomedical literature, then aligned them with a massive, robotically-assisted "high-throughput" experiment<sup>2</sup> to evaluate the likelihood of replication (Danchev, Rzhetsky, & Evans, 2019). Centralized communities were much more likely to publish new scientific claims that agreed with their own prior work, but were less likely to replicate.

This relationship between social centralization and diversity collapse was further explored in a field experiment where 733 programmers competed in an algorithm development task under two alternative communication and disclosure regimes (Boudreau & Lakhani, 2015). In the "open" regime, programmers shared intermediate code products and converged rapidly to the most successful recent advances. Under the closed or competitive regime, programmers competed without communication and a much more diverse range of approaches were cultivated.

This negative association between network centralization and system-level diversity may appear at odds with network analyses that demonstrate how the formation or activation of bridging ties in a network, which typically increase overall network centralization, are also associated with increases in the diversity, creativity and ultimate performance of the person or organization possessing them. In Granovetter's classic article "The Strength of Weak Ties" (1973) and his subsequent book, *Finding a Job* (1974), he demonstrated that the renewed activation of weak-ties enabled flows of novel employment information. Similarly, Burt has shown that managers with ties that bridge distinct social clusters within a company tend to generate broader and better strategic ideas (2004), and Stark and Vedres have shown that overlapping entrepreneurial ventures reap performance benefits (2010). Connections, especially when they bridge cognitively diverse groups and design experiences, can be valuable for the generation of creative products ranging from hit video games (de Vaan, Vedres, & Stark, 2015) to research articles (Uzzi, Mukherjee, Stringer, & Jones, 2013).<sup>3</sup>

<sup>1</sup> Based on data from OCLC's WorldCat data up until 2008 across 82,000 global libraries with share catalogs.

<sup>2</sup> NIH-funded LINCS L1000 experiment, which tested approximately 1.4 million drug-gene associations.

<sup>3</sup> Research on cities often focuses on not only the combination of difference, but its creation through specialization, but the equation between cultural difference generation and erosion is never explicated.

What remains unevaluated in these studies is how the increased internal diversity for those with bridging ties influence future system-wide creativity. Studies have shown that the information benefits from bridging structural divides would erode if everyone pursued a bridging strategy (Buskens & van de Rijt, 2008). Nevertheless, the perceived ubiquity of social and cultural barriers has prompted many to advocate greater interdisciplinarity and boundary-crossing creativity (Bauer, 1990; Braun & Schubert, 2003) without considering its cost for the cultural diversity that might limit returns from future expeditions. Most work on connections and creativity focuses on current, individual-level advantages, rather than counting costs for the entire system over longer time scales.

Scholars have begun to consider the limits of systemic creativity that arise from continuous or repeated connections (Burt, 2007). Studies of creative teams reveal the diminishing marginal creativity of repeated collaborations in scientific and scholarly research, product innovation (Guimerà, Uzzi, Spiro, & Amaral, 2005), and Broadway musical creation (Uzzi & Spiro, 2005). March's simulation of organizational learning styles takes this to the global level by exploring the consequences of centralized organizational communication. Through simulation, he shows that when people learn quickly from central organizational dogma, diversity collapses and the organization cannot incorporate novel knowledge of its members because through rapid conformity they have forgotten it (1991). This suggests a mutually constitutive relationship between individual-level innovation (Burt, 2004) and system-level conformity (March, 1991) associated with bridging and centralizing network connections. Centralizing and connecting disconnected societies may enable semantic mixing in the short-term, but through competition and selective cultural extinction, future periods have fewer meanings to mix. The more distant the groups connected, the greater the immediate innovation potential for the socio-semantic "arbitrageurs" that link them, but the larger decrease in system-level diversity that results from putting more semantic possibilities in contact and competition. This suggests that the second law of thermodynamics, where systems spontaneously drift toward greater disorganization or entropy over time, may have a manifestation in the realm of semantics.<sup>4</sup> Meanings in a social system become more *evenly distributed* as social relationships centralize, increasing the common exchange, competition and collective shrinkage of information (see Appendix and Fig. A1).

The correlation between social connection and cultural collapse may not manifest in all contexts. Indeed, it is the centerpiece of classical economics that exchange incentivizes diversity: if I specialize to create more of this and you produce more of that, then through exchange we both get more of what we want. This has been heralded as the principle behind successful "long-tail" Internet retailers like Amazon, who market themselves as able to sell small numbers of vast varieties of products (Anderson, 2006). This could suggest that in the hyperconnected Internet age, people are spreading out rather than concentrating their choices. Detailed research on who buys what, however, demonstrates that the long tail of product variety is extremely thin, with more people than ever before spending more of their money on popular products at the head of the distribution (Elberse, 2008) and simultaneously purchasing novelties in the tail (Goel, Broder, Gabrilovich, & Pang, 2010). As in product markets, it has also been observed that greater urban density is associated with greater innovation and diversity (Bettencourt, Lobo, Dirk Helbing, & West, 2007; Carlino, Chatterjee, & Hunt, 2006), but this fails to account for potential losses in the diversity of rural and other urban areas feeding cities. In short, while it is unclear whether greater social connection always correlates with cultural convergence, empirical examination of potential counter-examples including long-tail product markets and city diversity do not unequivocally argue against it. Moreover, cause and effect among socio-semantic dynamics are also complex and disentangling it is beyond the scope of this article (but see the Appendix and Fig. A10).

In this paper we seek to provide a new theoretical articulation and empirical illustration of the widespread relationship between centralization within a system and the diversity that system may sustain. In order to reformulate our thesis about social centralization and semantic collapse to facilitate its direct empirical assessment, we now explore geometric social and semantic representations that build on social-semantic network analyses, which were first to simultaneously render social actors and semantic expressions (Roth & Cointet, 2010; Saint-Charles & Mongeau, 2009).

### 3. The geometry of social and semantic networks

Why would we want to represent social and semantic networks in a multi-dimensional geometric space? What justifies the complexity of introducing a new paradigm for representing network data? Network scientists historically considered only the topology of social networks, and not their intrinsic geometry. The oldest of us recalls being taught in graduate school that embedding a network in geometry would not help to reduce its complexity because an  $n$ -node network required  $n-1$  Euclidean dimensions to represent. This had actually been disproven in 1984, when the Johnson and Lindenstrauss (1984) demonstrated that networks can be represented in far fewer ( $\leq \log n$ ) dimensions. The rise of autoencoders, or neural network models that embed data from one representation into another, alongside improvements in high performance computing have recently enabled analysts to efficiently and practically describe social networks with a *much lower dimensional* geometric representation (Perozzi, Al-Rfou, & Skiena, 2014; Shuicheng, Xu, Zhang, & Zhang, 2005).

Language and meanings, by contrast, have long been represented within continuous geometric spaces and only more recently tortured into network form. Quasi-continuous meaning spaces have been theorized for a century (Wittgenstein, Anscombe, & Wittgenstein, 1953; Firth & J. R., 1957), and first empirically performed with semantic surveys and factor analysis in the 1950s (Osgood, Suci, & Tannenbaum, 1964). Later approaches used matrix factorization on modest text data (Furnas et al., 1988), and more

<sup>4</sup> While the Second Law of Thermodynamics assumes the system moving toward complete entropy is isolated, even dissipative systems that exchange energy with their environments manifest increased entropy production over time up to some limited entropy (Prigogine & Defay, 1944; I. Prigogine, 1947). This principle has been widely used in explaining chemical, biological and even social phenomena.

recently auto-encoders like Word2Vec (Mikolov, Sutskever, Chen, Corrado, & Dean, 2013), GLoVe (Pennington, Socher, & Manning, 2014) and BERT (Devlin, Chang, Lee, & Toutanova, 2018) on very large-scale text. The resulting word embeddings reveal semantic compositionality (Bolukbasi, Chang, Zou, Saligrama, & Kalai, 2016) and recover widespread and subtle cultural associations (Caliskan, Bryson, & Narayanan, 2017; Garg, Schiebinger, Jurafsky, & James, 2018; Kozlowski, Taddy, & Evans, 2018).

Embedding social and semantic networks into the same low dimensional geometric representation could allow us to directly measure and compare their complex properties. Specifically, we will show that by projecting social and semantic networks onto a hyperbolic space, social centralization and semantic diversity become trivial properties of this new space (Chamberlain, Clough, & Deisenroth, 2017; Krioukov, Papadopoulos, Kitsak, Vahdat, & Boguñá, 2010; Nickel & Kiela, 2017; Papadopoulos, Maksim Kitsak, Ángeles Serrano, & Krioukov, 2012).

### 3.1. The hyperbolic geometry of social and semantic networks

What is hyperbolic geometry? Euclidean or straight-line geometry was constructed to obey Euclid's parallel postulate, which states that within a two-dimensional plane, for any given line  $\ell$  and point  $p$  not on  $\ell$ , there exists exactly one line through  $p$  that does not intersect  $\ell$ . This constructs a Euclidean space characterized by "straight" vectors and "flat" planes. Hyperbolic geometry is not straight, but curved, violating the parallel postulate such that (infinitely) many lines may go through  $p$  without intersecting  $\ell$ . A triangle has 180 degrees in flat Euclidean space, but fewer in hyperbolic space. A mountain pass, coral reef, or crocheted frill furnish examples of negatively curved surfaces that can be rendered with just two hyperbolic dimensions. Fig. 3, panel A presents M.C. Escher's print *Circle Limit IV* and panel B shares the underlying Poincaré disk model, which projects a 2-dimensional hyperbolic space onto a 2-dimensional Euclidean space for visualization (cited and recreated from Dunham et al., 2009). Escher's print alternatively tessellates angels and devils on the disk, *ad infinitum*, revealing how the space produces a "fisheye view" that packs both a single focus and wide context into the same frame. In fact, the image details become exponentially more dense as the radius increases.

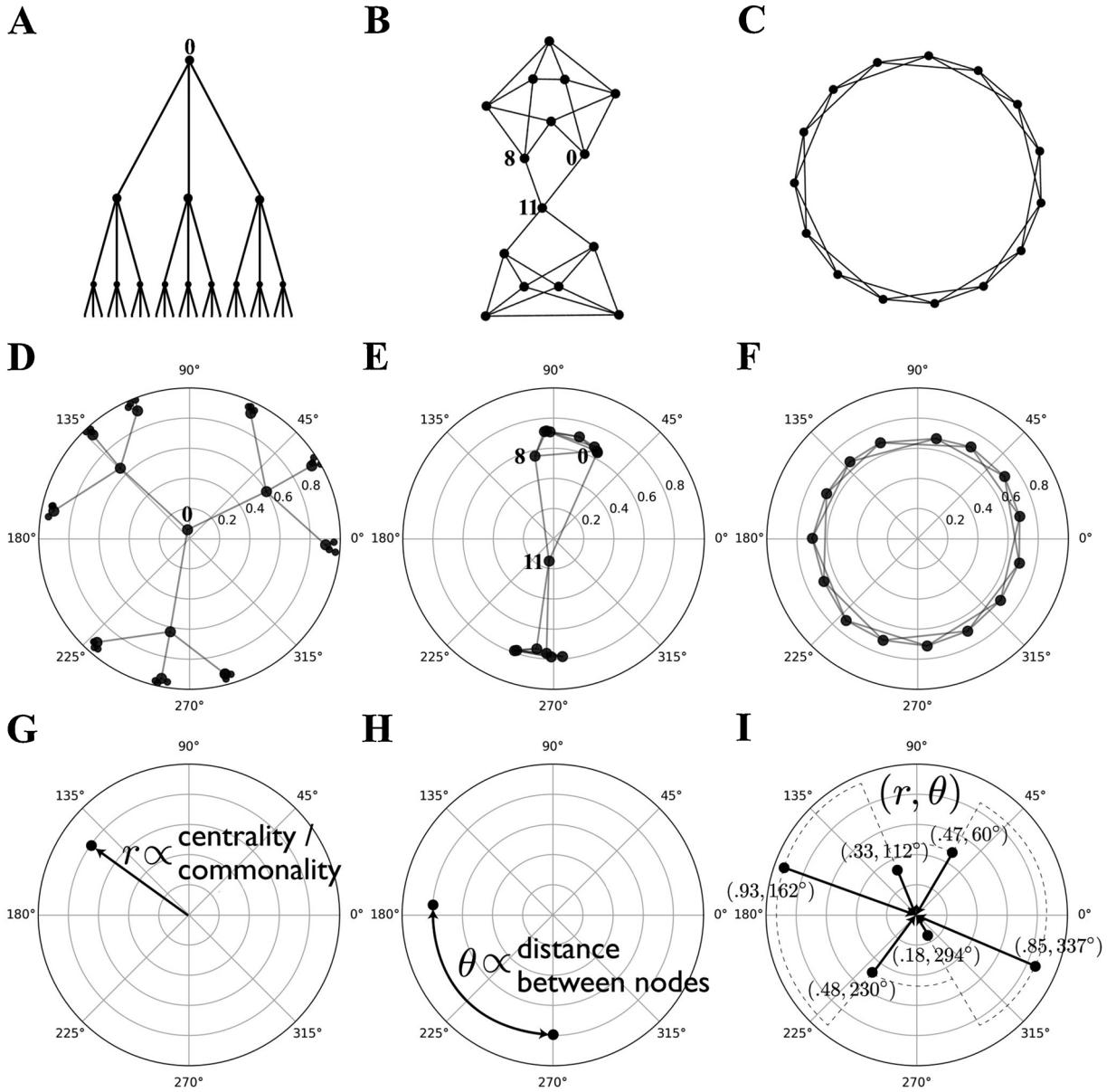
Why model social and semantic networks with a hyperbolic space? Social networks are defined by clusters (Freeman, 1977; Watts & Strogatz, 1998) bridged by network brokers (2009, Burt, 2004; Granovetter, 1973). These two properties are captured in the "small-world" model of Watts and Strogatz (1998). Social networks also manifest diversity in number of connections such that a few social actors possess most connections, while most possess a few. This results in phenomena such as the friendship paradox, where most people's friends have more friends than themselves (Feld, 1991; Zuckerman & Jost, 2001) and has been modeled by a process in which network agents prefer connection to those already popular, resulting in "scale-free" distributions such that any agent in the network has the same relative proportion of those with immediately more and less connections (Barabasi & Albert, 1999; Simon, 1955).<sup>5</sup>

Physicists Papadopoulos and colleagues showed how both "small-world" and "scale-free" properties can be reproduced by translating the process of network linkage into a hyperbolic space such that social actors connect to others within a fixed radius (Krioukov et al., 2010; Papadopoulos et al., 2012). This approach uses the center-periphery asymmetry of hyperbolic space to model the global reach and popularity of central social actors. Actors at the center of a hyperbolic space can reach all other nodes while those at the periphery are exponentially distant from one another and must go through those at center to efficiently reach others (see Fig. 1). A social network embedded in hyperbolic space is characterized by two directions, the "popularity" direction pointing from center to periphery, which captures the network's centralization, and "similarity" along the periphery where nodes are locally clustered, which captures the network's diversity. This interpretation is supported by Chamberlain and colleagues who embedded social networks into hyperbolic space and found that actors with high betweenness (Freeman, 1977) were central in that space, whereas high degree nodes with low betweenness were pushed to the periphery (2017).

Hyperbolic space is also a powerful representation for human languages, where bridging conjunctions like "and" and "but", common verbs like "to be" and "to have", or words with multiple meanings like "book" or "bear", may connect otherwise disconnected words and clusters of meaning. This builds hierarchy in semantics comparable to social networks (Fellbaum, 2005). Hyperbolic geometry is a strong candidate for modeling these bridged, asymmetric structures. Because of the expanded area or "ruffled edge" of hyperbolic space, two concepts can be simultaneously close to the same central, bridging idea but far from one another. As a result, hyperbolic geometry captures the hierarchy common to semantic systems. For example, WordNet is a network of words connected by shared meanings of synonymy. Embedding WordNet in hyperbolic space requires only a few dimensions (2–5) to model the complex relationships between words, far less than the hundreds of dimensions required to model it in Euclidean space (Chamberlain et al., 2017; Fellbaum, 2005; Nickel & Kiela, 2017). Like a social networks, a semantic network embedded in hyperbolic space is characterized by two directions, the "frequency" direction pointing from most to least common words, which captures centralization in the semantic space, and "similarity" along the periphery where meanings are locally clustered, which captures diversity among those meanings.

Geometric embedding models also operationalize and extend key elements of Bourdieu's field theory. Bourdieusian cultural fields offer a model for how individuals, cultural objects, and positions in social structure are located relative to one another in structurally homologous "social spaces," with relations between entities described in terms of "distances" (Bourdieu, 1989; Kozlowski, Taddy, & Evans, 2019). Bourdieu represented these spaces geometrically using correspondence analysis, representing distances between entities and meaningful dimensions through placement in a dimensionalized Euclidean plane (Bourdieu, 1984). The qualities of social

<sup>5</sup> Specifically, the "scale-free" property implies that when scaled by the correct exponent  $\tau$ , the logged distribution of social connections can be represented as a straight line. For example, if  $\tau = -1$ , then if one person had a million Twitter followers, a million people would have one follower.



**Fig. 1.** A simplified illustration of hyperbolic embedding. **A, B, and C** present three graphs designed to have different hierarchies (**A > B > C**). **A** is a tree of branching factor three and across three levels. **C** is a ring lattice in which each node connects to its four nearest neighbors. **B** is created by randomly rewiring edges in **C** while keeping node degrees unchanged and the whole network connected. Specifically, to obtain hierarchical structures we run the rewiring 200 times and update the graph only when (a) it remains fully connected and (b) the standard deviation of betweenness centralities for all nodes in the graph increases. **D, E, and F** show the 2D hyperbolic embedding in Poincaré disks of graphs **A, B, and C**, respectively. **G, H, and I** describe the embedding of hyperbolic space. **G** shows how the radius of the node  $r$ —its distance from center—is proportional to the centrality or commonality of connection with the node in question. **H** shows how the angle between any two nodes  $\theta$  is proportional to the distance between the nodes. **I** illustrates this polar coordinate system for several plotted points.

centralization and semantic diversity, however, manifest structural properties of hierarchy rather than dimension. Recent research demonstrates a powerful identity between social hierarchy in productive communities and organizations, on the one hand, and hierarchy and diversity in knowledge, whereby social structures facilitate the more or less efficient matching of problems to problem solvers (Garicano, 2000). Hyperbolic embeddings are precisely tuned to capture this hierarchy characteristic of social and semantic organization.

But what justifies the complication of embedding social and semantic networks to hyperbolic space? We are about to analyze the case of academic physics, which presents its own illustrative example. Isaac Newton represented the universe in simple, three dimensional Euclidean space, but the process of gravitational attraction between bodies in that space was complex—the product of

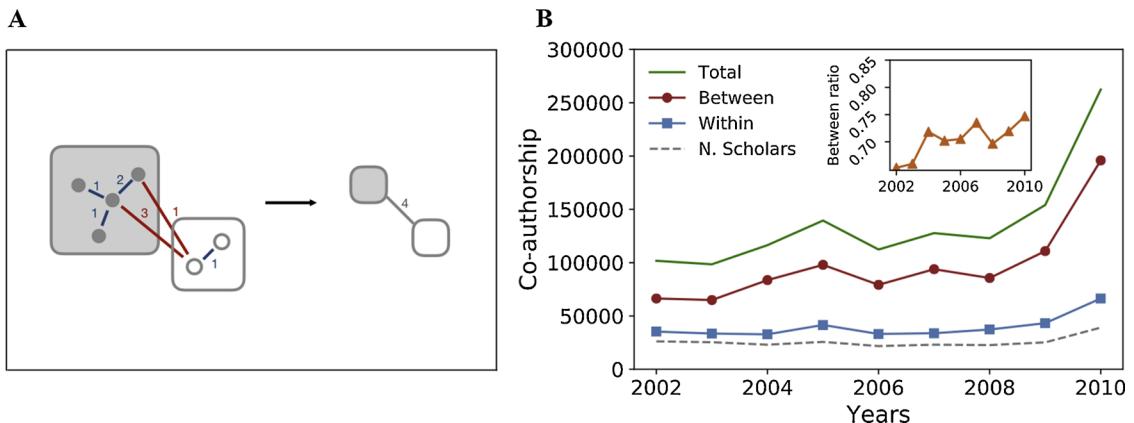
the bodies' masses and the universal gravitational constant divided by the square of their distance. Albert Einstein represented the universe in a more complex, warped space-time, but the process of gravitational attraction was simply the distance between objects in that space. A geometric (and hyperbolic) representation of social and semantic networks may seem like a complicated new addition to the marketplace of cultural analytics. But by efficiently packing the complicated asymmetries of those networks into curved hyperbolic space, the focal properties of centralization and diversity become trivial distance measurements. We embed social and semantic relationships within comparable hyperbolic spaces to directly interrogate the relationship between patterns of social centralization and semantic diversity.

We empirically examine social and semantic entanglement in the context of scientific knowledge production. Many studies have demonstrated that social influence contributes to both inequality and unpredictability in cultural markets (Salganik & Watts, 2008; Strang & Macy, 2001), yet there has been less empirical research on whether or how scientific understanding is driven by the consensus of scientific communities (Kuhn, 1970). As highlighted by Hutchinson (2011), scientism—the view that science has its own logic and the objective natural science is the only legitimate origin of knowledge—remains popular, but provokes culture wars. By showing the entwined social and semantic influences on scientific investigations underscores the social embeddedness of science as a cultural industry.

Specifically, we analyze the link between centralization and diversity across science and scholarship, with a special investigation of the semantic domain of academic physics, as represented by the complete 21<sup>st</sup> Century publications of the American Physical Society (APS), a society organized at Columbia University in 1899 “to advance and diffuse the knowledge of physics”. Insofar as the hyperbolic embedding of socio-semantic structure represents a kind of *physics of sociology*, we felt it might be appropriate to recursively use it to perform a *sociology of physics*. We specifically examine patterns of collaboration between universities in terms of their publishing physicists, and the diversity of physics topics those universities publish. We use published physics rather than a more ephemeral socio-semantic context like chit-chat on Tumblr, Twitter or Facebook largely because it is much more difficult to identify individual persons and trace them persistently in social media settings, even though it may be easier to follow micro-social and semantic interactions in those settings. Inflamed interaction between a flock of tweets is more specific and intense than that between an archive of articles referencing one another. Because academic physicists have a strong incentive to build persistent reputations through publishing within a single area, our data represents a very conservative case from which to examine the relationship between of social centralization and semantic diversity. If we observe that collaboration and topic selection move together among academic physicists, it is likely amplified elsewhere.

We note that while most prior work has focused on collaborations between individuals rather than institutions, the number of ties across institutions closely tracks the total number of ties overall (see Fig. 2). By focusing on coauthorships among institutions, we are able to see how ties that bridge institutional collaboration clusters relate to reductions in topic diversity across the entire system. Moreover, centralization within fields is often discussed in terms of the stable rank order of dominant institutions that integrate the broader field.

Formally, when we embed networks to a Poincaré disk (2D hyperbolic space), the position of each node  $i$  is defined by two parameters in polar coordinates that capture these properties, radius  $r_i$  and angle  $\theta_i$ . The radius quantifies position in the hierarchy: nodes of small radius literally hold a central position. The angle between two nodes  $\theta_i - \theta_j$  quantifies their proximity or structural similarity, and the entropy over all such angles is a direct measure of diversity.



**Fig. 2. The construction of institutional networks.** Panel A shows the co-authorship between five scholars (dots) from two universities (rectangles). Co-authorship within universities is colored in blue and co-authorship between universities is colored in red. Weights represent the number of co-authorships in a given year. When we construct institutional collaborations we keep only the co-authorship between universities. Panel B shows the yearly counts of the number of scholars (the gray, dotted line), the total co-authorship between these scholars (green), which is separated into co-authorship between (red) or within (blue) institutions. Panel B reveals that the number of between coauthorships is proportional to the number of total coauthorships and so highly descriptive of the entire pattern of coauthoring ties.

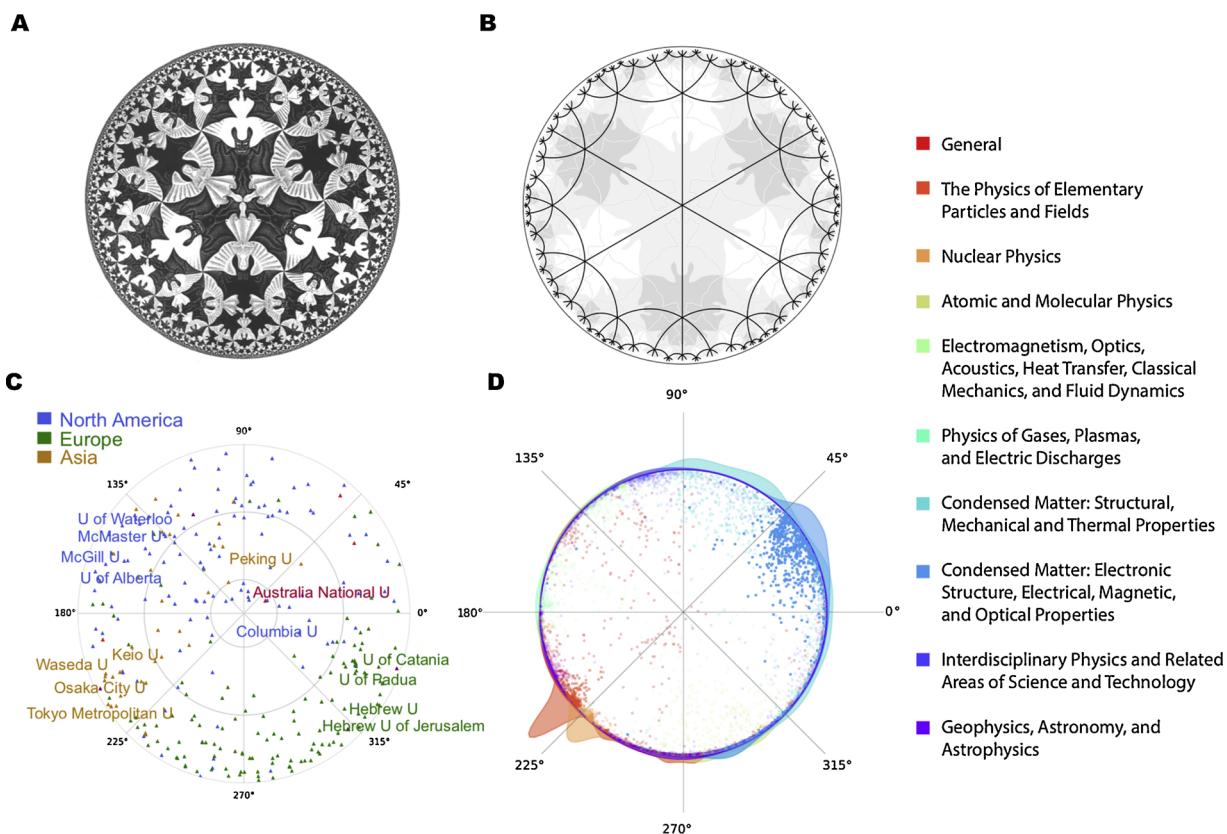
#### 4. Data

We analyze centralization and diversity in science and scholarship using physics articles from the American Physical Society (APS) journals, and then confirm some of these findings with articles archived in the Web of Science (WoS). Because the identification of co-authorship and institutional networks hinges on how author and institution names are disambiguated, we discuss our approach to disambiguation as we introduce our datasets. First, we introduce the WoS data on which our disambiguation approach hinges, and then the APS data that draws on it.

Our WoS dataset contains 43 million journal papers and 615 million citations that span from 1900 to 2014. These papers are published across 15,146 journals. To match the APS data (introduced below), we selected and analyzed 13 million journal articles from our WoS dataset across seven fields during 1994–2013 based on WoS journal classification meta data.

After performing name disambiguation on this data (see Appendix for details), we connected name-disambiguated institutions to name-disambiguated authors in each year, enabling authors to change institutions across years. In total, we obtain 372,495 authors from 28,754 institutions. The most institutionally connected author has 26 institutional affiliations, and 83 % have only a single one.

We weighted author collaboration networks by iterating over all papers for a given year and connecting scholars that collaborated on the same paper. In this way we obtained ten annual networks of author collaboration for 2002 through 2011. Next, we re-normalized the networks by merging scholars from the same institutions into single nodes and then aggregated author collaboration edges to connect institutions (see Fig. 2). Edges between scholars from the same institution are retained as self-loops. To compare institution networks across years, we fixed the nodes by selecting the top 300 institutions, but removed 10 due to missing data and so we analyzed collaborations between the most prolific 290 institutions. After we constructed the ten yearly institution collaboration networks, we embedded each into a 2-dimensional hyperbolic space using the Poincaré disk model (Nickel & Kiela, 2017). As anticipated, we found that the embedding space captured both the hierarchy among institutions and diversity between them, as shown in Fig. 3.



**Fig. 3.** Panel A and B present M.C. Escher's print *Circle Limit IV* and the underlying the Poincaré disk model, recited and recreated from (Dunham et al., 2009). The space becomes exponentially denser as the radius increases. Panel C shows the hyperbolic embedding of the institution collaboration network in 2011 using the the Poincaré disk model. The displayed 290 institutions (triangles) are colored by regions: blue for North America, green for Europe, orange for Asia, red for Australia, and purple for South America. Panel D shows the hyperbolic embedding of the aggregated PACS code co-occurrence network using the Poincaré disk model. The analyzed 5819 PACS codes dots are colored by the 10 subfields to which they belong. Gaussian kernel density estimation is used to visualize the concentration of the angles of PACS codes in each subfield see Fig. A5 for details). To compare across subfields, we also rescale the estimated distributions such that the area covered by the distribution curve is not unity, but proportional to the total number of papers published by 290 institutions in this subfield from 2002 to 2011.

In order to study semantic content in physics, we used the Physics and Astronomy Classification Scheme (PACS), developed by the American Institute of Physics (AIP) and used in the Physical Review journals since 1975 to identify fields and subfields of physics. PACS is a hierarchical partitioning of the full spectrum of topics in physics, astronomy, and related sciences. Each PACS code is a six-digit number, in which the first digit indicates one of ten subfields to which it belongs (see Fig. 3). PACS codes are semantic artifacts—agreed upon expressions of meaning. We constructed the PACS code co-occurrence network by iterating over all 359,395 papers in our dataset and connecting PACS codes that appeared together in the same paper. The network contained 5856 nodes, 244,106 edges, with a total edge weight of 1,124,724. After we constructed this PACS code co-occurrence network, we embedded it into hyperbolic space using the Poincaré disk model, comparable with the institutional collaboration embedding described above. As with the organization network, we found that the representations of PACS codes in a 2D hyperbolic embedding captures both their hierarchy and similarity, as shown in Fig. 3.

In order to extend our validation of the association between social connection and semantic convergence across many fields of science, we used WoS meta-data from 13 million journal articles and journal classifications<sup>6</sup> 1994–2013 to construct network and information theory measures. We were unable to construct hyperbolic spaces for all of these collaborations and topics, and so instead for each field in a given year, we constructed author collaboration networks and counted the average number of collaborators for authors as the measure of social densification in that field. The degree of semantic convergence is measured by the entropy of word distributions. Specifically, we extracted the abstracts of articles and selected all tokenized words that occur twice or more, then after removing stopwords, calculated the entropy of word distributions for each field in a given year. To control for differences in word vocabulary length across fields, we follow the convention of normalizing calculated entropy by its theoretical maximum,  $\text{Log}(N)$ , for which  $N$  is vocabulary length. We examine the association year by year at different levels of resolution, and detailed results are shown in Fig. A7 for 13 fields, in Fig. A8 for more than 200 subfields, and in Fig. A9 for approximately 10,000 journals.

## 5. Methods

We use the Poincaré Embeddings algorithm (<https://github.com/facebookresearch/poincare-embeddings>) developed by Facebook AI scientists Nickel and Kiela (2017) using PyTorch. This algorithm automatically learns hierarchical representations of nodes in networks by training on positive and negative samples by using a shallow, three-level neural network auto-encoder. We embedded our (1) university collaboration network across publications in all time periods (2002–2011), (2) separately for each year, and (3) the co-occurrence network of all PACS codes across publications in all years, using a 2-dimensional Poincaré disk, with a learning rate of 0.5 and a negative sample size of 50.<sup>7</sup> This produced hyperbolic embeddings in which each node  $i$ —a university in the collaboration embedding and a PACS code in the semantic embedding—has radius  $r_i$  and angle  $\theta_i$ . Nodes of small radius hold central positions in the circularly arrayed hierarchy. Angles between two nodes,  $\theta_i - \theta_j$ , quantifies their social or semantic difference (See appendix for the definition of inner product of two nodes in the 2d Poincaré disk).

In order to calculate the position  $\theta_i$  for each university  $i$  in the semantic or PACS code space, we selected the peak or mode of the Gaussian kernel density estimation for all of its weighted PACS codes in a given year (see Fig. A5) from the PACS code positions estimated for all years, 2002–2011 (see Fig. A4). To estimate each university's level in the topical hierarchy, we calculated the hyperbolic distance between all pairs of PACS codes from a university and select the PACS code of the shortest hyperbolic distance to all other codes. This selected PACS code is closest to “the center of mass” in the hyperbolic space and thus represents the diverse topics covered by the university. With this estimation, each university  $i$  has a unique  $r_i$  and  $\theta_i$  inherited from the selected, representative code in both collaboration and topical space for each year.

These preliminaries set us up to directly explore the relationship between social and semantic structure in science and scholarship, and specifically in physics. We first did this by taking each pair of universities and calculating their social and semantic hyperbolic distances at every time point, which generated two time series. We then evaluated these time series pairs in order to explore the complex relationship between social and semantic factors. This suggests that increased institutional contact is associated with a decrease in the diversity of topics explored. We also performed the same analysis between each university and all other universities in the system at every time point to explore the connection between social relationships and the contraction of the entire semantic system.

As mentioned above and detailed in the appendix, we validate our findings with traditional network measures of social density and information theoretic measures of semantic diversity. Nevertheless, we argue that our hyperbolic metrics are conceptually superior. Consider three plausible measures to capture the spread and size of topics in semantic space: 1) the span of vocabulary; 2) that span, weighted by topic frequency—or entropy; and 3) that span, weighted by frequency and accounting for semantic overlapping between words—what our embedding approach accomplishes by representing words as weights on orthogonal dimensions of meaning. Entropy measure can only be an unbiased measure of semantic diversity if all keywords are independent from one another and represent orthogonal dimensions of meaning. We chose hyperbolic embeddings for our measure as they model semantic overlapping and frequencies directly as angles and radius, respectively.

<sup>6</sup>The journal classifications are used to map articles into subfields and fields of science.

<sup>7</sup>We specified a batch size of 30, which limits the size of examples from data on which the algorithm trains. We trained the model on 300 sequential epochs, of which the first 20 were “burn-in” epochs used only to initialize the model. The model was evaluated every 5 epochs, and every time gets updated if it has lower loss against the validation data. The co-occurrence of collaboration edges in our network are undirected and so we set the parameter symmetrize to be true. In addition, the epsilon parameter is set be no smaller than 1e-5 such that the maximum hyperbolic radius is 1-1e-5.

## 6. Findings

### 6.1. The social and semantic geometry of physics

In order to summarize the global social and semantic structure of 21<sup>st</sup> Century physics, we calculated and graphed the radius and angle in embedded social space for each of the 290 top physics publishing universities and in semantic space for thousands of PACS codes. We found that in both social and semantic disks, the radius systematically quantifies the institutional “hierarchy” of university physics departments in the inter-university collaboration network, and angles quantify the research diversity between them. As demonstrated in Fig. 3C, high-ranking institutions from various global regions occupy the central area in the social space, including Columbia University from North America, Peking University from Asia, and Australian National University from Australia. Geological groups emerge and cluster together on the periphery, including institutions from Canada (University of Waterloo, McMaster University, McGill University, and the University of Alberta), Japan (Kelo University, Waseda University, Osaka City University, and Tokyo Metropolitan University), Israel (Hebrew University and Hebrew University of Jerusalem), and Italy (the Universities of Catania and Radua).

PACS codes group together on the periphery, forming and indicating subfields in physics (Fig. 3D). The most clustered PACS codes include “The Physics of Elementary Particles and Fields” (red) and “Nuclear Physics” (orange); which nontrivially overlap in the lower left quadrant of the Poincaré disk. The tall peaks of density distribution centered on these fields suggest that their codes are nearly always used with others from the same field in research articles. They rarely combine with other field codes. Situated at the antipodes of the disk; in the upper right; are “Condensed Matter: Structural; Mechanical and Thermal Properties” (light blue) and “Condensed Matter: Electronic Structure; Electrical; Magnetic; and Optical Properties;” (blue) which each have a higher  $\theta$ ; and are substantially more spread. Nuclear physics and the physics of elementary particles focus on individual particles and their properties. Condensed matter physics focuses on interactions between particles; often in the context of large; naturalistic materials. It may not do too much violence to the epistemic cultures of physics to say that particle physics is the *economics* of the physical sciences, flush with cash for large-scale experiments; focused on the essential monads of existence in artificial isolation and furnished with a sense of superiority that their work is more fundamental and important than their neighbors (Fourcade, Ollion, & Algan, 2015).<sup>8</sup>

Condensed matter physics might then be viewed as the *sociology* of the physical sciences, fixed on social interactions between particles in complex, overlapping physical fields, often embedded in the rich context of materials and phenomena that can be macroscopically experienced by ordinary humans, like sandpiles (Jaeger & Nagel, 1992), coffee rings (Deegan et al., 1997) and the physics of food (Poon, 2002). Consider the Ising model from statistical mechanics that sources from “soft matter” concerns and models the emergent property of phase transitions in ferromagnetic materials as atomic spins arranged in a graph. The graph arrangement allows each spin to interact with its neighbors. This model is frequently used by social network analysts and physicists to model human social interaction (Klemm, Eguíluz, Toral, & Miguel, 2003; Wasserman & Pattison, 1996). Insofar as physics increasingly tends to cleave between a focus on particles or interactions (see Fig. 4C), we leave it to the reader to decide which represents the angel and which the devil in Fig. 3A.

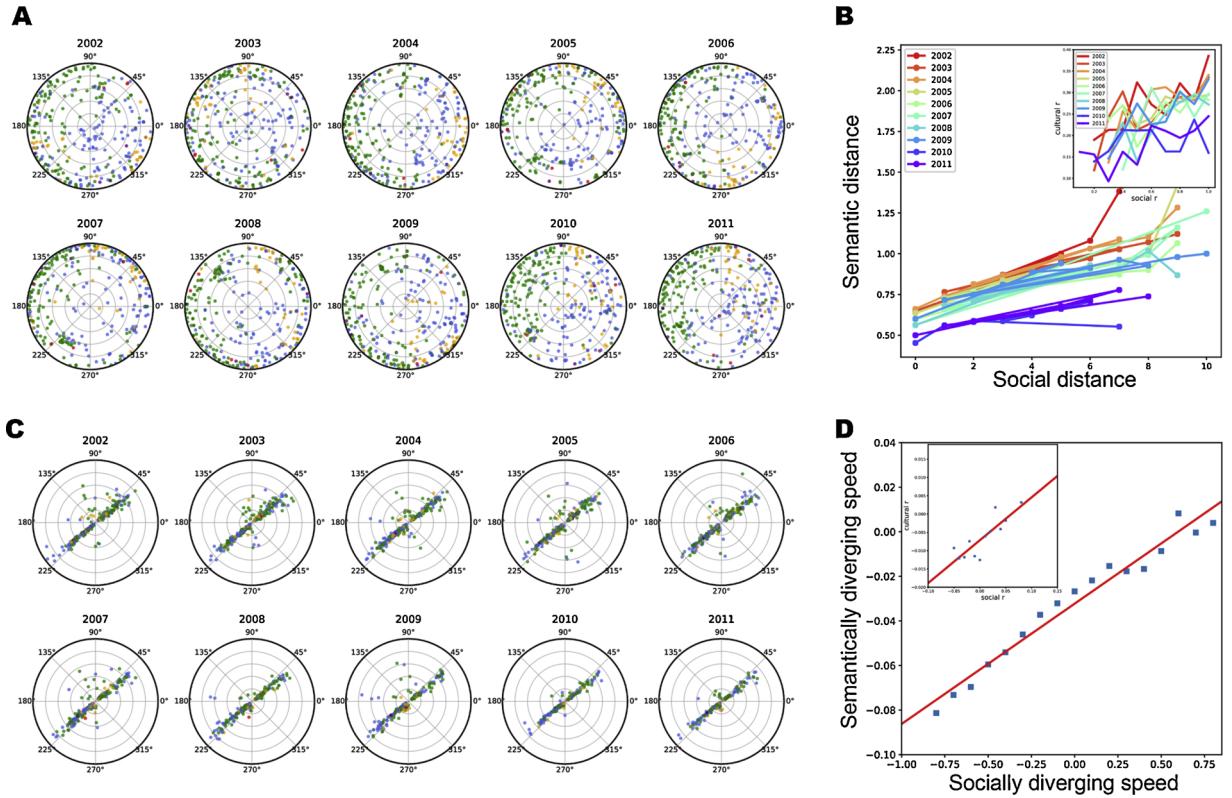
With a special hubris, spokespeople in physics have claimed their science as the fundamental center (Young & Freedman, 2013) or even the only meaningful science—in the words of Ernest Rutherford: “all science is either physics or stamp collecting.” This work highlights that 21<sup>st</sup> Century physics is not a unified front, but rather a growing polarization between particle and condensed-matter physics, which originates from two fundamentally different perspectives on the decompositionality of our universe (P. W. Anderson, 1972). These debates have become cause and consequence of polarized communities and led to real-world dramas, including the cancellation of U.S. Superconducting Super Collider (SSC) in 1993 after \$2 billion had been spent, and the 2016 debate and cancellation on the ambitious plan of Circular Electron–Positron Collider (CEPC) in China with an early stage budget of 30-billion-yuan (\$4.3-billion). In this way, our investigation into the polarization in physics may provide insight into similar patterns in politics, ideology, and other domains.

Plasma physics, fluid dynamics and classical mechanics lie off of the fundamental particle / soft matter axis, but have also been much less intensively published in 21<sup>st</sup> Century APS publications.<sup>9</sup> The most spread field is not a field at all, but rather “Interdisciplinary Physics and Related Areas of Science and Technology” (deep blue), which diffusely spreads around the entire system and whose codes are frequently connected to those from other fields. That these codes are not often close to the center of the disk suggests that rather than “Interdisciplinary”, they are more likely “Multidisciplinary” codes that interact with other specific fields at the periphery, but do not integrate those fields into the broader system of physics. The field codes closest to the center are those from “General” (red), which is not a category at all, but rather was derived from the first two decimal places of each PACS code and so, by definition, connects different subfields of physics by forming the trunk of their hierarchical tree.

We map changes in the social hierarchy and centralization among university physics collaborations over time, from 2002 until 2011 in Fig. 4A, where universities are colored by region. Here we see strong and consistent geographical clustering in collaboration

<sup>8</sup> Wolfgang Pauli denigrated the physics of solids as “Schmutzphysik” or dirt physics, while Murray Gell-Mann jabbed that “solid state physics” be replaced with “squalid state physics” (Montgomery & Largent, 2015 p. 195).

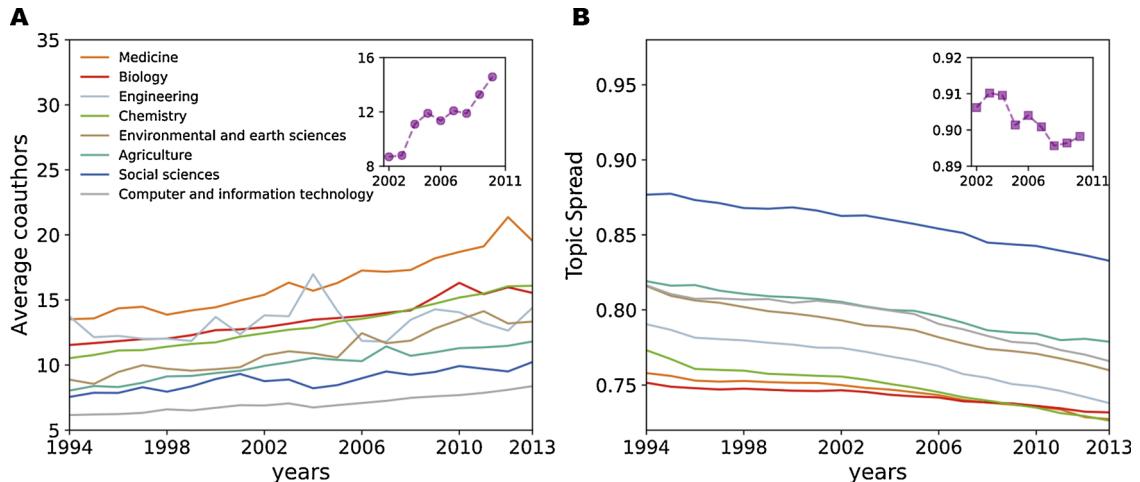
<sup>9</sup> Some fields like plasma physics have suffered in the U.S. and Europe from encumbrances for the international collaboration, ITER, the International Thermonuclear Experimental Reactor.



**Fig. 4.** (A) The hyperbolic embedding of annual university physics collaboration networks (2002–2011). We embed each annual collaboration network into a hyperbolic space and project them on Poincaré disks, such that each institution  $i$  has two parameters in the hyperbolic social space, angles  $\theta_{si}$  and radius  $r_{si}$ . The universities (dots) are colored by region, blue for North America, green for Europe, orange for Asia, red for Australia, and purple for South America. (C) Annual projection of institutions in the hyperbolic space of physics topics. We embed the PACS code co-occurrence network into Poincaré disks as shown in Fig. 3D, and then calculate the position of each university in a given year from the position of their PACS codes, such that each institution  $i$  has two parameters in the culture space, angles and radius  $r_{ci}$ . The manner by which we aggregate the PACS code parameters is illustrated in Fig. A5. (B) The correlation between social and semantic spaces. We calculate the hyperbolic distance between every pair of universities in both Social and Semantic space as  $C_{ij}$  and  $S_{ij}$  for all pairs of institutions in each year, then plot  $C_{ij}$  against  $S_{ij}$  across years (data is binned). The inset shows the correlation between social and semantic hyperbolic distances at the university level, i.e., we calculate the average hyperbolic social and semantic distances from one institution to all the other institutions  $1/n \sum_{j=1}^n S_{ij}$  and  $1/n \sum_{j=1}^n C_{ij}$ , and plot them against one another. (D) The correlation between the social and semantic time series, social distance  $S_{ij}(t)$  and semantic distance  $C_{ij}(t)$ . For each pair of institutions, we use an OLS regression to fit the slopes  $\beta_{sj}$  (social diverging speed) of  $S_{ij}(t)$  against year  $t$  and  $\beta_{cj}$  (semantic diverging speed) of  $C_{ij}(t)$  against years  $t$ . The large, positive slopes imply that two institutions are moving toward and away from each other in the corresponding spaces. We find that  $\beta_{sj}$  and  $\beta_{cj}$  are positively correlated, with a Pearson correlation coefficient equal to 0.98 ( $P\text{-value} < 0.001$ ). Similarly, in the inset we show that the slopes of the time series  $r_{si}(t)$  and  $r_{ci}(t)$  are also correlated, with a Pearson correlation coefficient equals 0.85 ( $P\text{-value} < 0.001$ ).

over time, with North American, European, and Asian universities tending to collaborate with one another. Australian and South American Universities are dispersed as satellites of geographically diverse, dominant universities from Asia and the West. Within the hierarchy, North American institutions remain the most central collaborators throughout, with the lowest average radius  $r$ , but Asian universities experienced the greatest change in  $r$ , migrating rapidly from the periphery toward the collaboration network's center and resulting in greater centralization in the global collaboration system.

We also chart changes in the semantic position of physics universities over the same time period in Fig. 4C, with universities colored by region, as in Fig. 3C, but plotted against the subfield locations represented in Fig. 3D. It is interesting that the representative positions of most university physics departments cluster near the subfield of "Condensed Matter Physics" (45°) or "The Physics of Elementary Particles and Fields" (225°). Moreover, the field of university physics departments tends to collapse toward polarization between the physics of particles and of complex materials over time. In 2002, a scattering of European and a few U.S. universities have a wider mixture of PACS codes, focusing on plasma or fluid physics, but by 2011, they tend to represent some mixture of particle or systems physics. North American Universities appear most likely to specialize in particle or systems physics, while European and Asian Universities represent a much more even mix of the two polar branches.



**Fig. 5. The densification of collaboration networks and the convergence of topics and wording.** We selected and analyzed 13,178,911 journal articles from our WoS dataset across seven fields during 1994–2013. **A.** The average number of coauthors increases over time for a majority of fields. The fields are based on WoS meta data of journal classifications. To count the average number of coauthors, we sum over the number of papers, weighted by number of coauthors in the paper, for each name-disambiguated scholar and average over all scholars. The inset shows the increase of coauthors for APS data. **B.** The spread (entropy) of words from article abstracts decreases over time for a majority of fields. We select 3,510,489 words that occur twice or more in the abstract to represent a research article (commonly used words are removed). To control the word vocabulary length difference across years we normalize the calculated entropy by its theoretical maximum  $\text{Log}(N)$ , in which  $N$  is the vocabulary length. The inset shows the decrease in entropy of the PACS codes.

## 6.2. Social centralization and semantic contraction

When we compare the social and semantic hyperbolic distances between every pair of universities over all years, binned, we see a very strong positive association, as graphed in Fig. 4B. This suggests that social and semantic distances are tightly bound together across all years. We also see that the range of distances is much smaller in the social than the semantic world. Nevertheless, the range of the two distances are very different. Topical distances vary between .4 and 1.4, while social distances move from .5–8. This difference reflects the deep topical inertia in academic physics, while patterns of collaboration can shift dramatically from year to year. We also see that the semantic space of university physics is dramatically contracting over time, nearly halving for the least connected institutions between 2002 and 2011.

Next, we explore the association of collaboration and focus over time. For each pair of institutions, we estimated an OLS regression to identify the slope representing change in social distance over time. The coefficient for time,  $\beta_{sij}$ , should be interpreted as the speed of social convergence or divergence between any two institutions  $i$  and  $j$  across the 2002–2011 time series. Then we separately estimated the slope of semantic change over time, where the coefficient on time,  $\beta_{cij}$ , reflects the speed of semantic convergence or divergence between that pair of institutions. The Pearson correlation coefficient between the two diverging speeds,  $\beta_{sij}$  and  $\beta_{cij}$ , is 0.98 ( $P$ -value < 0.001), providing strong evidence that in the context of academic physics, collaboration is associated with topical convergence (see Fig. 4D). In summary, not only are university physics collaborations moving with their evolving interests and expertise, but they are moving the same way over time.

In appendix Fig. 5, we validate the relationship between social centralization and semantic contraction with standard network and information theory measures. In the inset of the right panel, we see that when an institution's clustering coefficient in the network goes up, their entropy over PACS codes goes down. This suggests with greater density comes less (individual) diversity. We note that the hyperbolic measures between pairs of universities and between each university and all others provides more dispositive and descriptive information regarding social centralization and contraction. The intellectual burden of projecting social and semantic networks to a geometric representation that may be unfamiliar to cultural analysts is directness of the measures of social centralization and semantic diversity that result.

Finally, we sought to explore the association between centralized collaboration and the convergence of topics beyond physics, but disambiguating authors and assembling or constructing a terminology for all fields was beyond the scope of this investigation. We relied on the established relationship between enlarging teams and social centralization across science (Danchev et al., 2019; Wuchty, Jones, & Uzzi, 2007). We selected and analyzed 13 million journal articles from our WoS dataset across seven fields between 1994 and 2013. Based on the WoS data that codes journals by field, the average number of coauthors increases over time for all fields, which corresponds to the pattern we find for physics (APS) as shown in the inset. This corresponds to a decreasing spread or entropy over the words used in article abstracts for those same fields, which also corresponds to decreased entropy among PACS codes across APS articles. Insofar as greater collaboration is a critical component of greater centralization in APS, we see that greater collaborative

centralization in the form of enlarging and overlapping teams is linked with a reduction in the distribution of topics discussed and investigated.

While the trend of decreasing topic spread and increasing collaboration are observed for both physics and beyond overtime, we also found that similar patterns exist cross-sectionally for almost all science disciplines in all years investigated. In appendix Fig. A7 we show that for the 13 major fields in science, topic spread is lower for those fields with greater author collaboration. OLS regressions show that slopes for different years are stable between -0.01 and -0.03 and significant for most years. The associations remain consistent at higher resolution for more than 200 subfields (Fig. A8) and more than 10,000 journals (Fig. A9). These findings add further evidence that social connections and semantic convergence are closely associated, insensitive to which method we use.

## 7. Discussion

We live in an age of social hyper-connectivity and centralization. From the rise of social media to the drop in transportation costs that facilitate social gathering, meeting, and mixing, there are more opportunities to connect in a growing and increasingly centralized hub of relationships than ever before. In the sociological and organizational literature, the process of bridging social (Burt, 2009) and cultural holes (Lizardo, 2014; Pachucki & Breiger, 2010) has been linked with a wide range of individual, organizational and city-level benefits through increased access to cultural information, and opportunities for strategic and creative recombination. Individual benefits incentivize strategic networking, enabling entrepreneurs to engage in arbitrage across communities and markets (Burt, 2009). Scientists and engineers similarly strive to place their ideas and inventions at the center of complex networks in order to cement their continuing influence and importance (Latour, 1987). All of these forces reduce the diversity of ideas, symbols and other semantic forms, ultimately placing them in competition and resulting in rising rates of cultural extinction. With growing attention to the duality of social and cultural structures (Mohr, 2000; Lee & Martin, 2018; Breiger & Puetz, 2015; Basov & Brennecke, 2017), we argue for the importance of attending to this cycle of social exploration and cultural exploitation that characterizes processes of colonialism, globalization, and capitalism, resulting in cultural creativity, convergence, and sometimes collapse. The process is manifest in the extinction of indigenous languages, practices and artifacts. Moreover, the causal arrow goes in both directions. Shared global languages, opinions and tastes facilitate more social connection and centralization, which recursively put more semantic forms in competition and at risk of abandonment.

In this paper, we demonstrated that social and semantic worlds could be informatively represented with hyperbolic geometry to simultaneously reveal the complex relationship between social centralization and semantic convergence. We argue that the burden of learning this novel representation is justified by the importance of existential and practical benefits for understanding and preserving cultural diversity, which modernity threatens. Hyperbolic embeddings reveal centralization and diversity as simple and direct measurement of the space in which social and semantic networks are embedded.

We then rendered the university collaboration network of 21<sup>st</sup> Century physics traced by the corpus of the American Physical Society within a two dimensional Poincaré disk. We comparably represented the semantic field of physics. We found that social connection among academic physics departments is becoming more centralized and dense over time, and that universities tend to increasingly sponsor research polarized between particle and nuclear physics, on the one hand, and condensed matter and plasma physics, on the other. We ran the same analyses for each university in their collaborations with all other universities, finding that greater external collaboration is consistently associated with future convergence to the most popular topics in the broader field. These findings were supported by a standard analysis of network density and information theoretic diversity. We also demonstrated that increased and enlarged teaming across all areas of science and scholarship, a driving force of collaborative centralization, is associated with a similar drop in the diversity of words used within those fields.

There are several limitations of our analysis. First, our case of modern physics may not be perceived as representing the fluidity and spontaneity of other semantic domains. Nevertheless, we argue that it represents a conservative case, given the strong rewards for topical consistency in academic promotion. Second, our short time frame, 2002–2011, does not allow us to observe long-term change and influence between social and semantic spheres. Third, by calculating statistics like percentages over many estimated models (see Fig. A10), we implicitly avoid taking into account the social and semantic dependencies between cases. Hyperbolic distances between university pairs in the collaboration and PACS code networks were constructed in context of all other collaborations and published PACS codes and so their measurements are naturally dependent upon one another. Despite these limitations and complexities, we believe that our illustration demonstrates many ways in which hyperbolic embeddings can make visible and shed light on the complex association between social centralization and semantic diversity so characteristic of modernity.

The great irony is that modern tastes for diverse music, food, clothing and ideas may themselves limit the sustainability of diversity. The tight relationship between social centralization and semantic contraction poses a paradox for modern, multicultural societies that value diversity. Extensive research in social, behavioral and organizational science documents the largely positive effect social and cultural diversity exerts on the collective production of information, goods and services (Joshi & Roh, 2009; Page, 2008). Individuals from culturally distinct groups embody diverse cultural perspectives and cognitive resources that combine to produce solutions and designs that outperform those from homogeneous groups (Hong & Page, 2004; Mannix & Neale, 2005; Nielsen, 2012; Woolley, Chabris, Pentland, Hashmi, & Malone, 2010). Collaborations between inventors from distinct social groups result in more creative patents (Fleming, Mingo, & Chen, 2007), scientific teams representing distinct disciplines produce more highly cited papers (Wuchty et al., 2007), gender diversity broadens the questions scientists ask (Nielsen, 2012), and political diversity leads to improved

political pages on Wikipedia (Shi, Teplitskiy, Duede, & Evans, 2017). Beyond performance, some enjoy a taste for difference or heterophily (Lazarsfeld et al., 1954). But if and when semantic differences fuse through social centralization, they become less different. And when differences erode, so do the performance benefits that come from them. Social centralization consumes semantic diversity.

Only by considering the relationship between society and meaning on both short and long time scales can we clearly see the essential tension between the enjoyment and protection of semantic diversity. Semantic differences can be generated through isolation and specialization, but as with indigenous languages like Wichita, Mandan and Yurok, or megafauna like the Tasmanian tiger, Chinese Baiji dolphin, or North American heath hen, all of which irreversibly disappeared within the last hundred years, it is easier to extinguish semantic and biological forms than to evolve new ones. We highlight the power of hyperbolic embedding in making diversity and centralization visible, and argue for greater incorporation of long-run cultural dynamics into theories of social centralization, integration and change to insure long-run benefits from sustainable diversity.

## Appendix A

As we will show below, a Poincaré disk can accurately represent a branching tree structure, and higher dimensional hyperbolic spaces can capture hierarchies with multiple heritage, such as directed acyclic graphs. As a result, negative curvature and diverging, branching structure of hyperbolic space enables us to model the hierarchy and sparse, bridging structures common to many realistic social and semantic networks in far fewer dimensions than a comparable number of Euclidean dimensions

### *Physics, Hyperbolic Geometry and Structural Representation*

It is notable that hyperbolic geometry was connected with physics from inception. Nikolai Lobachevsky, one of hyperbolic geometry's 19th Century discoverers, argued that the universe might not be flat, but rather hyperbolic (Bonola, 1955), and Einstein's general relativity intensified interest in the curvature of the universe by positing that mass and energy bend spacetime. In general relativity, mass and energy can be used to determine the universe's curvature ( $\Omega$ ) as the average density of the universe divided by the mass energy required for it to be flat. Despite recent estimates that put  $\Omega$  at approximately 0 (Biron, 2015), spacetime in the vicinity of massive bodies such as stars, solar systems and galaxies is positively curved, and space between massive bodies, in dark energy dominated areas,<sup>10</sup> is negatively curved (Krioukov et al., 2012).

Statistical physicists, computer and network scientists first applied hyperbolic geometry as a compact representation of social networks (Krioukov et al., 2010; Papadopoulos et al., 2012) and semantic hierarchies (Chamberlain et al., 2017; Nickel & Kiela, 2017). A Poincaré disk can represent a branching tree structure without distortion, and higher dimensional hyperbolic spaces can perfectly capture hierarchies with multiple heritage, such as directed acyclic graphs. As a result, negative curvature and diverging, branching structure of hyperbolic space enables us to model the hierarchy and sparse, bridging structures common to many realistic social and semantic networks (Nickel & Kiela, 2017).

### *Euclidean geometry v.s. Hyperbolic geometry*

*Euclidean geometry* was constructed to obey Euclid's fifth postulate, which states that within a two-dimensional plane, for any given line  $\ell$  and point  $p$  not on  $\ell$ , there exists exactly one line through  $p$  that does not intersect  $\ell$ . The result is a space dimensionalized by "straight" vectors and "flat" planes having zero curvature, such that a drawn triangle's angles add up to  $180^\circ$  and the Pythagorean theorem holds.<sup>11</sup> Curved geometries including both elliptical and hyperbolic geometries violate Euclid's fifth. A 2-dimensional elliptical geometry, for example, has positive curvature and can describe the surface of the earth, such that all lines through  $p$  intersect  $\ell$ , and any two lines perpendicular to  $\ell$  intersect at a "pole". The sum of a triangle's interior angles on an elliptical surface is always greater than  $180^\circ$ .<sup>12</sup>

Hyperbolic space, by contrast, has negative curvature, such that infinitely many lines may go through  $p$  without intersecting  $\ell$ . A saddle, mountain pass, coral reef, or crocheted frill furnish examples of negatively curved surfaces. A triangle etched upon them will have angles that sum to less than  $180^\circ$ . As two intersecting lines diverge at a constant rate in Euclidean geometry (e.g., the Pythagorean theorem), and converge in elliptical geometry, they diverge exponentially in hyperbolic geometry such that there is "more space" in a 2-dimensional hyperbolic disc than a Euclidean circle or elliptical globe. In Euclidean geometry, the circumference of a circle with radius  $r$  equals  $2\pi r$ , but in hyperbolic geometry it is always more and elliptical geometry always less. Fig. 3, panel A presents M.C. Escher's print *Circle Limit IV* and panel B shares the underlying Poincaré disk model, the projection of a 2-dimensional hyperbolic space onto a 2-dimensional Euclidean space, cited and recreated from (Dunham et al., 2009). Escher's print alternatively tessellates angels and devils on the disk, *ad infinitum*, and reveals how the space becomes exponentially more dense as the radius increases.

<sup>10</sup> As dark energy is constant over the universe, regions with little mass are proportionally "dominated" by dark energy.

<sup>11</sup> The Pythagorean theorem states that the square of the hypotenuse of a right triangle (the side opposite the right angle) is equal to the sum of the squares of the other two sides.

<sup>12</sup> Consider a triangle etched onto earth's surface with one side stretching along a quarter of the equator and the other two sides meeting at the South Pole. All three angles will equal approximately  $90^\circ$ , summing to  $270^\circ$ .

The inner products in Poincaré disk and Euclidean space are defined as follows. For two points  $x = (r_x, \theta_x)$ ,  $y = (r_y, \theta_y)$  in the Poincaré disk using polar coordinates, the hyperbolic inner product of  $x, y$  is  $\langle x, y \rangle = \|x\| \|y\| \cos(\theta_x - \theta_y) = 4 \operatorname{arctanh} r_x \operatorname{arctanh} r_y \cos(\theta_x - \theta_y)$ . In a Euclidean space, the inner product of  $x, y$  is defined as  $\langle x, y \rangle = \|x\| \|y\| \cos(\theta_x - \theta_y) = r_x r_y \cos(\theta_x - \theta_y)$ .

### *The Information Theory of Cultural Collapse*

Shannon's information theory has recently been used in cultural models of meaning (Martin & Lee, 2018) and furnishes a formal model of communication and information (Shannon & Weaver, 1963) that illuminates the theoretical and observed relationship between social connection and cultural contraction. Communication plays a central role in information theory and can be equated with social connections or pre-existing "information channels" required to exchange messages. Information, in Shannon's theory, is modeled as message passed between sender  $X$  and receiver  $Y$  that resolve the receiver's uncertainty about the sender's communicated meaning. Messages may express and transmit cultural forms including tastes, expectations, ideas, behaviors and artifacts (Carley, 1991, 1986; DiMaggio, 1987; Erickson, 1988; Kandel, 1978; Krackhardt & Kilduff, 1990; Krohn, 1986). Information  $I$  is equivalent to the surprise of experiencing a given message or cultural communication, such that an improbable or unexpected message has more information than an anticipated one and is experienced by the receiver with greater surprise. Information, then, is the news quality of a message. If senders and receivers of information mingle through repeated back-and-forth communication, however, the *mutual information*  $I(X;Y)$  of their messages—the information common to both  $X$  and  $Y$ , modeled as two random variables—increases. This increase in mutual information necessitates a drop in *joint entropy*  $H(X,Y)$ —the average information produced by both actors communicating together (see Fig. A1, top panel). This relationship assumes that individuals exhibit a conserved number of cultural forms and that what forms they possess, they express (Leydesdorff & Ivanova, 2014).

The pattern for two communicating actors generalizes to an arbitrary number of communicators linked via network. In larger networks, the inverse relationship holds between the multivariate mutual information and multivariate joint entropy, with some slight technical amendment. In the case of many variables—or individuals in society—the multivariate mutual information can be positive or negative, and so for the identity to remain precisely equivalent in a larger network and that of a single dyad, the joint entropy shrinks linearly with growth of the union between (1) the multivariate mutual information (e.g.,  $I(X;Y;Z)$  in a three-person network; see lower left panel in Fig. A1), and (2) all conditional mutual information measures (e.g.,  $I(X;Y|Z)$ ,  $I(X;Z|Y)$ ,  $I(Y;Z|X)$  in a three-person network; see lower left panel in Fig. A1). As a result, when increased channels of communication are unleashed by a denser pattern of social relationships with an arbitrary number of individuals, the cultural information passed between them shrinks in direct proportion to their shared or mutual information.

This information theoretic rendering of the relationship between structure and culture reinforces the mutually constitutive relationship between individual-level innovation unleashed by novel social connections and communication (Burt, 2004) and the system-level diversity depressed by it (March, 1991). Connecting disconnected societies enables cultural mixing in the short-term, but through competition and selective cultural extinction, future periods have fewer cultural forms to mix. The more socially and culturally distant the groups connected, the greater the immediate innovation potential for the socio-cultural "arbitrageurs" that connect them, but the larger decrease in system-level diversity that results from putting more cultural forms in contact and competition. As mentioned in the text, this suggests that the second law of thermodynamics may have a manifestation in culture, such that the distribution of cultural forms in a social system becomes more *even* as social relationships increase the mutual exchange of information.

### *Name disambiguation of authors and institutions*

In order to disambiguate WoS authors we use a hybrid algorithm. For each name (including family name and initials), we construct a network of relevant papers connected on the basis of a similarity measure that considers shared co-authors, references and citations. Disconnected components of this network are assumed to correspond to distinct authors (Wu, Wang, & Evans, 2017). We obtain 10 million scholars who contributed to 22 million papers. Eighty-five per cent of these scholars contributed to three or more papers, and 44 % to four or more. We use the 2017 Open Researcher and Contributor ID (ORCID) dataset to validate the name disambiguation results, and find that precision is 78 % and recall is 86 % among the 118,094 ORCID scholars with three or more papers. We also test the results using a dataset of 31,070 Chinese scholars and 253,786 papers retrieved from the project outcome reports of research funded by National Natural Science Foundation of China with precision 79 %, and recall 65 %.<sup>13</sup>

APS publication data is a freely accessible online (<https://journals.aps.org/datasets>), containing all meta-data for articles published in *Physical Review Letters*, *Physical Review (A, B, C, D, E and X)*, and *Review of Modern Physics*, some of which predate the APS, going back to 1893. In order to analyze collaborations between scholars and institutions, we first needed to disambiguate author and institution names: the same name may refer to multiple people, and distinct affiliation names reference the same institution. To extract unique author names, we matched the APS meta-data with WoS data using the DOIs (Digital Object Identifier) of papers published during 2001–2013. From this disambiguated data, we successfully identified 372,495 authors from the analyzed 359,395 APS papers.

<sup>13</sup> "Precision" here is defined as the percentage of cases that are "correctly identified"(true positive), and "Recall" is the percentage of total relevant cases that are actually achieved by the method.

To resolve ambiguity among institution names, we extracted affiliation names and fed them into the Google Geolocation API to identify whether the same latitudes and longitudes are returned. Google Geolocation provides results within 25 feet if the address translation is correct. Leveraging this approach, we merged 330,560 institution names into 55,317 unique ones.

#### *Modeling complex social and semantic networks using the Poincaré disk model*

The initial algorithm used to represent social networks in hyperbolic space was slow and inefficient (Papadopoulos, Psomas, & Krioukov, 2015). The primary bottleneck involved the Hamiltonian Monte Carlo method used to estimate node positions, which repeatedly calculated the gradient or fit between network and geometric embedding and became burdensome as datasets scaled up. This changed dramatically when computer scientists discovered that artificial neural networks were powerful tools for embedding optimization. Moreover, analysts have demonstrated that low dimensional hyperbolic embedding improve upon higher dimensional Euclidean embedding for many tasks, including predicting the collaboration between scientists in weighted undirected collaboration networks (Chamberlain et al., 2017; Nickel & Kiela, 2017).

In Fig. A2, we graph the institution collaboration network over time, with all 290 institutions (dots) colored by region: blue for North America, green for Europe, orange for Asia, red for Australia, and purple for South America. Edges represent the co-authoring relationship between scholars from each institution. Between 2002 and 2011, the number of edges increased from 3331 to 7,647, and the total weight increased from 29,117 to 94,336 over the same period.

In Fig. A3 we illustrate the movement of institutions in collaboration space over time. We pick Columbia University as the focal institution (the red triangle) and investigate the temporal evolution of its distance to two collaborating institutions, the University of Notre Dame (the orange rectangle) and University of Southern California (the blue rectangle). Over time, the University of Notre Dame moves towards Columbia University, whereas the University of Southern California moves away from it.

We also highlight the top ten institutions closest to in Columbia University in 2002 (purple dots) and their position in the following years to provide visual guidance for distance in the hyperbolic space, which differ substantially from 2D Euclidean distance.

In Fig. A4, we illustrate the movement of institutions in the semantic space of physics over time. We constructed the culture space as shown in Fig. 3D, and display all 5819 PACS codes in the panels above as gray dots. Next, we calculated the representative locations of Columbia (red triangle), Notre Dame (orange rectangle) and University of South California (blue rectangle), as shown in Fig. A3, but within the culture space based on the positions of their PACS codes using the method introduced below and illustrated in Fig. A5. We also fitted Gaussian kernel density curves to show the concentrate and spread of PACS codes for each of these three institutions.

In Fig. A5, we illustrate the empirical distribution of angles and radius for PACS codes from Fig. 3D. Each row shows the 5819 PACS codes from one of ten subfields, which are colored following the same scheme as Fig. 3D. Due to the size limitations of the figure, abbreviated names of the subfields are used. The first column shows the two estimated Poincaré disk parameters, angles  $\theta_i$  (the x-axis) and radius  $1 - r_i$  (the y-axis; for figure readability rather than  $r_i$ ) of the  $i$ th PACS code. The second column shows the empirical distribution of angles  $\theta_i$  and corresponding Gaussian kernel density estimation. The red lines show the peak of the angle distribution we select to represent the summary research direction of subfields or institutions based on their PACS codes. The third column shows the empirical distributions of radius  $1 - r_i$ . We found that due to the asymmetric property of the hyperbolic space, the values of most  $r_i$  approach 1. As such, neither mean nor median can characterize the difference in hierarchy between two universities or subfields. Therefore we take the extreme values by calculating the average of the ten smallest  $r_i$  to represent the university or subfield hierarchy.

After we constructed a university collaboration network, we also measured the density of social collaborative connections. Following convention, we define social density as the ratio of observed network edges to the maximum number possible (Wasserman & Faust, 1994). We denote the set of weighted edges connecting the  $i$ th node into the global institutional collaboration network as  $E_i$ , which can be further separated into  $E_i = E_i^{in} + E_i^{out}$ , where  $E_i^{in}$  denotes self loops and  $E_i^{out}$  edges connecting node  $i$  with other nodes (Palla, Barabási, & Vicsek, 2007). For institution  $i$ , we defined a weighted clustering coefficient:

$$c_i = \sum_{kh} (|E_k^{out} \cap E_h^{out}| |E_i^{out} \cap E_k^{out}| |E_i^{out} \cap E_h^{out}|)^{1/3} / (\max(w) * |E_i^{out}| * (|E_i^{out}| - 1))$$

where  $k$  and  $h$  include all unique pairs of institutions that constitute closed triplets with  $i$  and  $\max(w)$  represents the maximum weight of  $i$ 's ego network. We use  $c_i$  to measure social density as an alternative to measures defined in the hyperbolic social space. We also calculated the entropy of PACS codes as a measure of semantic diversity. In Fig. A6 we find that social density and semantic hierarchy are positively correlated and that social density and semantic diversity are negatively correlated. Therefore, our expectation that the broad association between social interaction and semantic contraction appears to hold for 21<sup>st</sup> Century Physics.

If position in semantic hierarchy is defined as radius  $r_{ci}$  and semantic diversity  $S_{ci}$  as the entropy over their PACS code angles on the Poincaré disk, then the two main figures show the temporal evolution of  $c_i$ ,  $r_{ci}$  and  $S_{ci}$  for two extreme cases. The left panel shows how social density ( $c_i$ , the left y-axis) and semantic hierarchy ( $r_{ci}$ , the right y-axis) move in the same direction. More specifically, both  $c_i$  (dotted red line) and  $r_{ci}$  (solid red line) for the University of Innsbruck decreases over time. We plot these as ranks and not original

values and also adjust the scale of ranks to provide visual guidance for comparing trends. By contrast, both  $c_i$  (dotted blue line) and  $r_{ci}$  (solid blue line) for the Naval Research Lab increase over time. The right panel shows how social density ( $c_i$ , the left y-axis) and semantic diversity ( $S_{ci}$ , the right y-axis) go in the opposite direction. The  $c_i$  for the University of Innsbruck decreases over time (dotted red line), whereas its PACS code entropy increases (solid red line). By contrast, the  $c_i$  of Johns Hopkins University increases over time (dotted blue line), but its PACS code entropy decreases (solid blue line). The two insets confirm that the full sample social density and semantic hierarchy are positively correlated and the full sample social density and semantic diversity are negatively correlated.

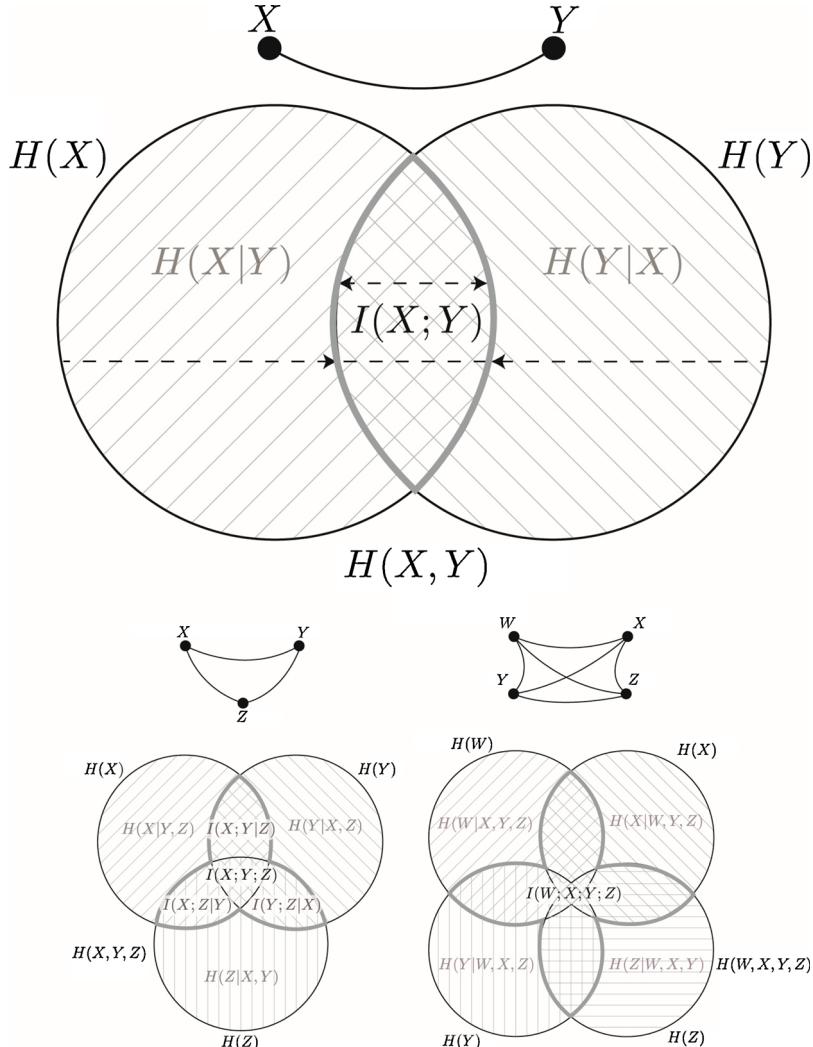
In Fig. A10, we explore the directional prediction of social and semantic distance, and their component parts—radius and angular distance—over time, by lagging each variable by one year and predicting the others. Even though it takes time for collaboration to yield a research article with particular PACS codes, both collaboration and PACS codes are assessed from publications, and so predicting future collaborations from present PACS codes indicates the likelihood that a similar publishing focus at time  $t$  will attract like-minded physicists to work together at time  $t + 1$ . Conversely, predicting future PACS codes from current collaborations indicates the likelihood that collaborations with a partner institution at time  $t$  anticipates a convergence in publishing focus at time  $t + 1$ . When we perform these regressions between each pair of universities, we discover the degree to which social centralization leads to convergence, and its converse. When we perform them for each university and all other universities, we unearth the extent to which social centralization leads to semantic contraction. Fig. A10 summarizes all of these Granger causal regressions. The number decorating each arrow in the graph represents the percentage of significant Granger causal regression estimates that post a positive influence.

From these figures, we see that the apparent association causality between social and semantic distances is bidirectional, suggesting that both socialization and selection are driving the attention of universities' physicists. Nevertheless, present collaborative distance drives future semantic or topical distance slightly more consistently and significantly than its converse. 67 % of all regressions predicting next year's topic distance from each pair of universities' current social distance post a positive effect, compared with only 57 % that do so in the opposite direction. By contrast, 89 % of all regressions predicting next year's topic distance from current global collaborations between a university and all others post a positive effect, compared with 79 % from collaboration to shared focus. These patterns provide moderate evidence that social centralization leads to semantic convergence between pairs of universities, but strong evidence that more global social centralizations lead to semantic contraction.

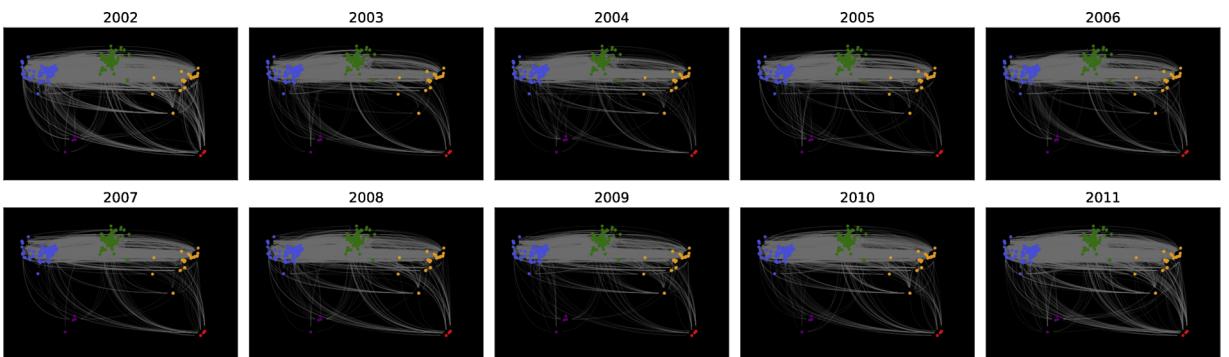
When we decompose the hyperbolic distance between pairs of universities, and between each university and all others, into the difference between their angle ("divergence") and radius ("hierarchy"), we find that hierarchical distance in the social or semantic domain is always a more consistent predictor of hyperbolic distance in the other domain than angular distance. This suggests a more consistent sensitivity of social and semantic distances to differences in the centrality of a university's physicists in the alternative space. Greater centrality provides universities and their physicists with a wider menu of options, and less centrality gives them far fewer choices regarding what they can do or where they can go next. As a result, differences in social centrality more consistently magnify differences in semantic distance, and the converse, than differences in angle, which represents a university's specific topical focus or collaboration cluster. This provides substantial support for the importance of using a hyperbolic representation that highlights hierarchy for the social and semantic systems characterized by it.

Below we provide additional examples of hyperbolic embedding in the academic context. Fig. A11 shows how the aggregated density distribution of PACS codes for four top universities precisely characterizes their topic focus in Physics. Fig. A12 shows that all tags of questions in the online Physics Q&A community on Stack Exchange can be effectively embedded using the Poincaré disk model to reveal their hierarchy and similarity.

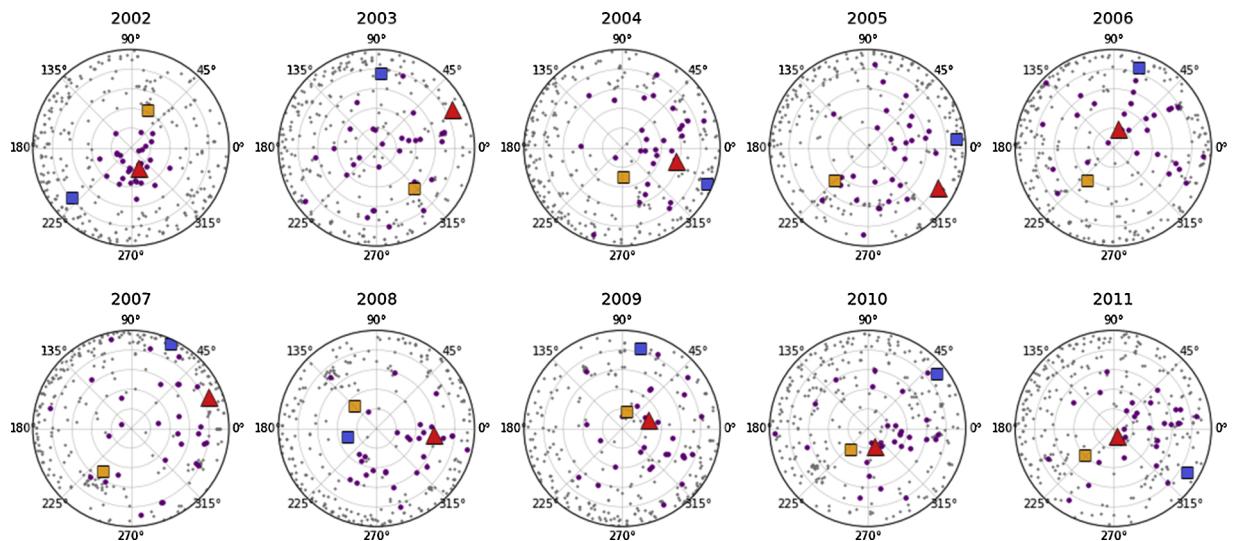
## Appendix



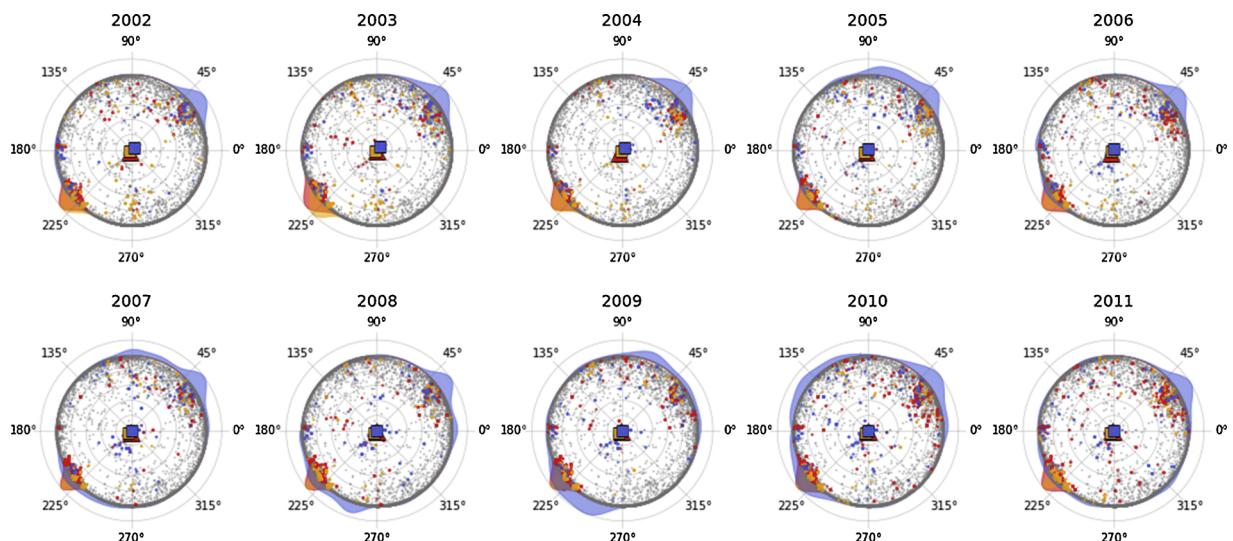
**Fig. A1.** Venn diagrams for information measures calculated for culturally correlated actors  $W, X, Y$  and  $Z$ . For the simple  $\{X, Y\}$  network,  $H(X)$  constitutes the entropy or average information associated with  $X$ 's cultural forms and  $H(Y)$  the entropy of  $Y$ 's cultural forms. As their relationship and similarity increases, their mutual information  $I(X;Y)$  rises and their joint entropy  $H(X,Y)$ —the total information in the system—shrinks.  $H(X|Y)$  is the conditional entropy of  $X$  given  $Y$ , or the average cultural information or surprise another actor  $A$  would experience when communicating with  $X$  after having communicated with  $Y$ . The graphs and Venn diagrams at the bottom of the figure illustrate the information scenario for clustered networks  $\{X, Y, Z\}$  and  $\{W, X, Y, Z\}$  (with a missing  $W, Z$  intersection for the  $\{W, X, Y, Z\}$  network), where the joint entropy shrinks linearly with the union of all multivariate mutual informations (see footnote 3).



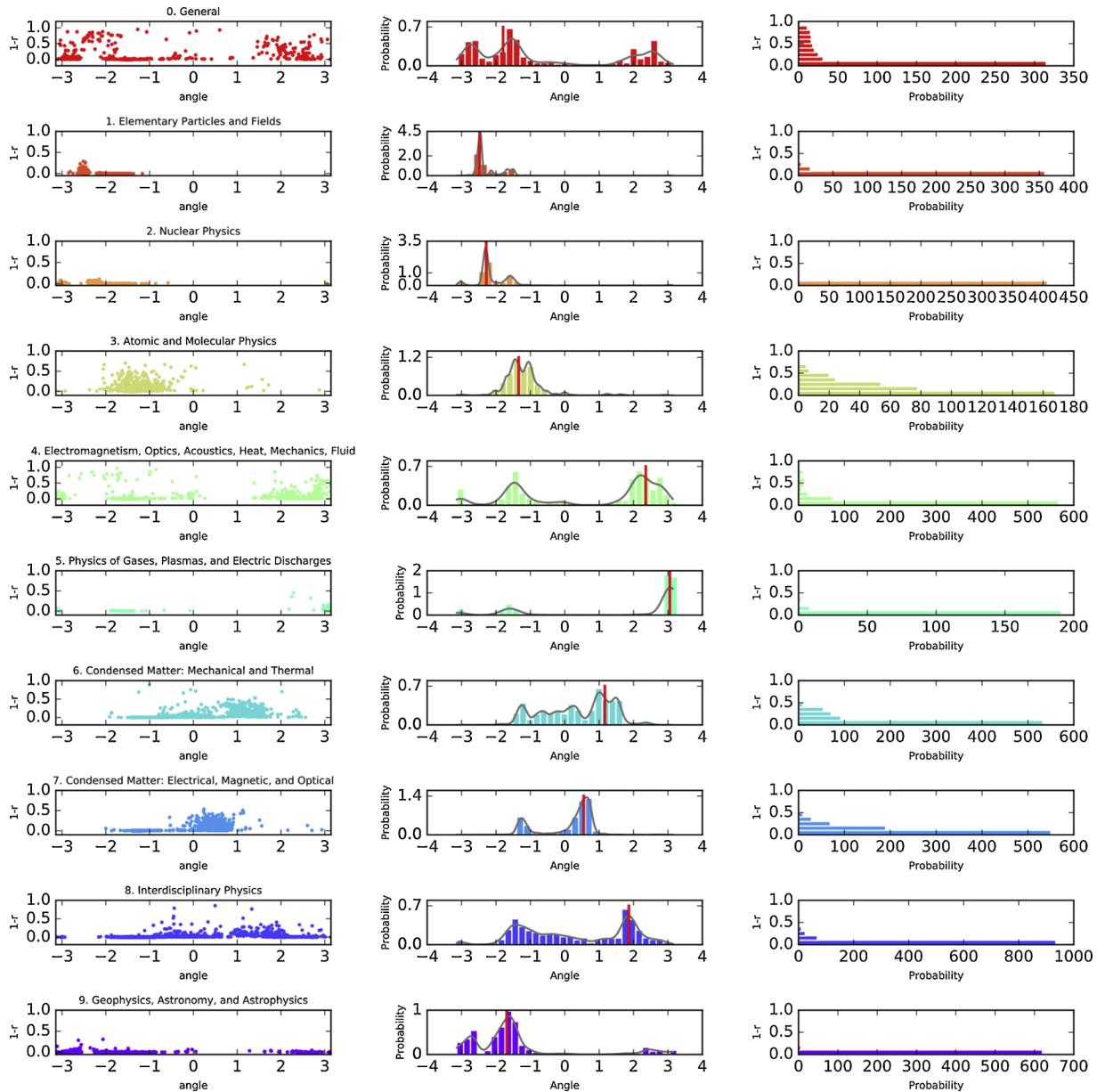
**Fig. A2.** Institution collaboration network, 2002–2011.



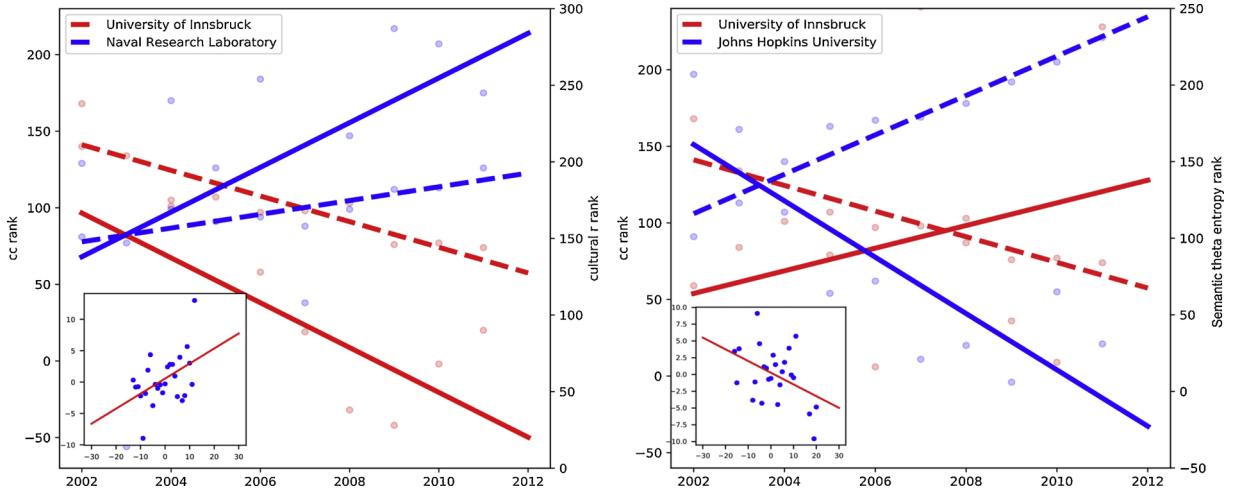
**Fig. A3.** Movement of institutions in collaboration space, 2002–2011.



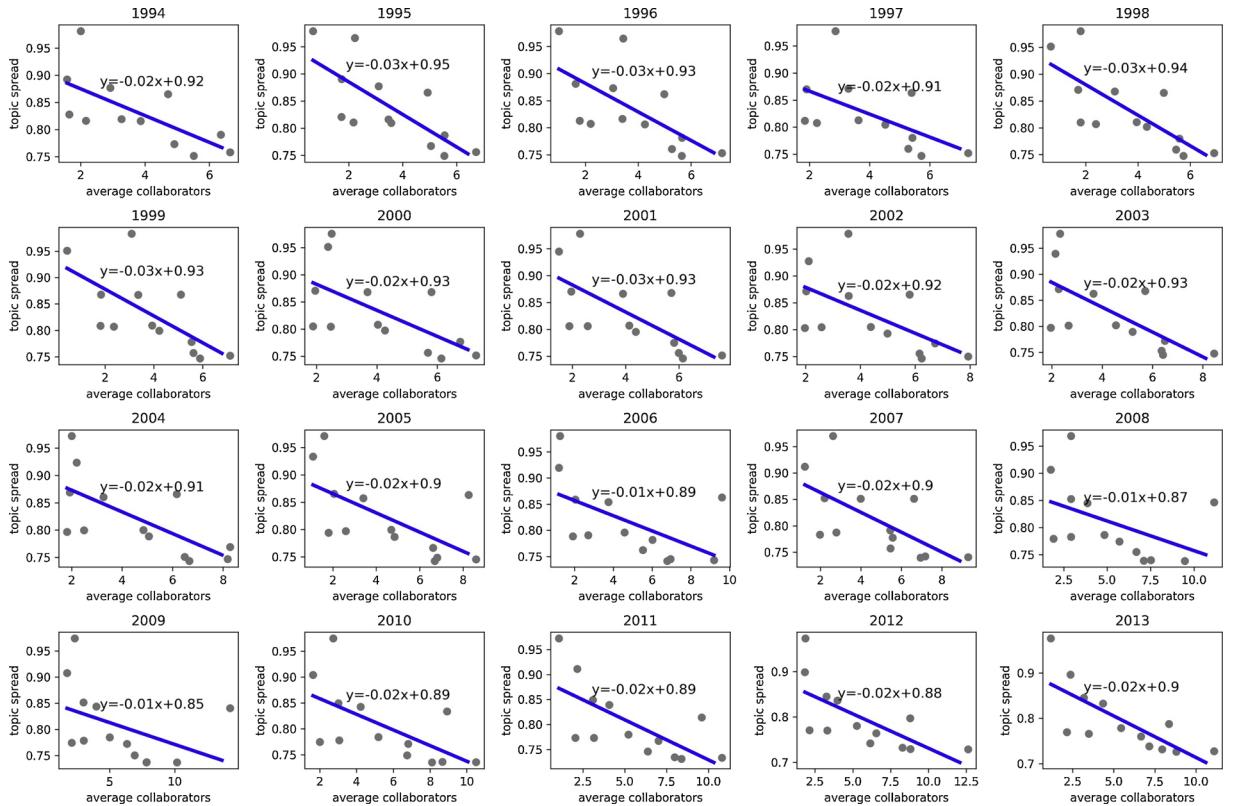
**Fig. A4.** Movement of institutions in the “semantic space” of physics topics, 2002–2011.



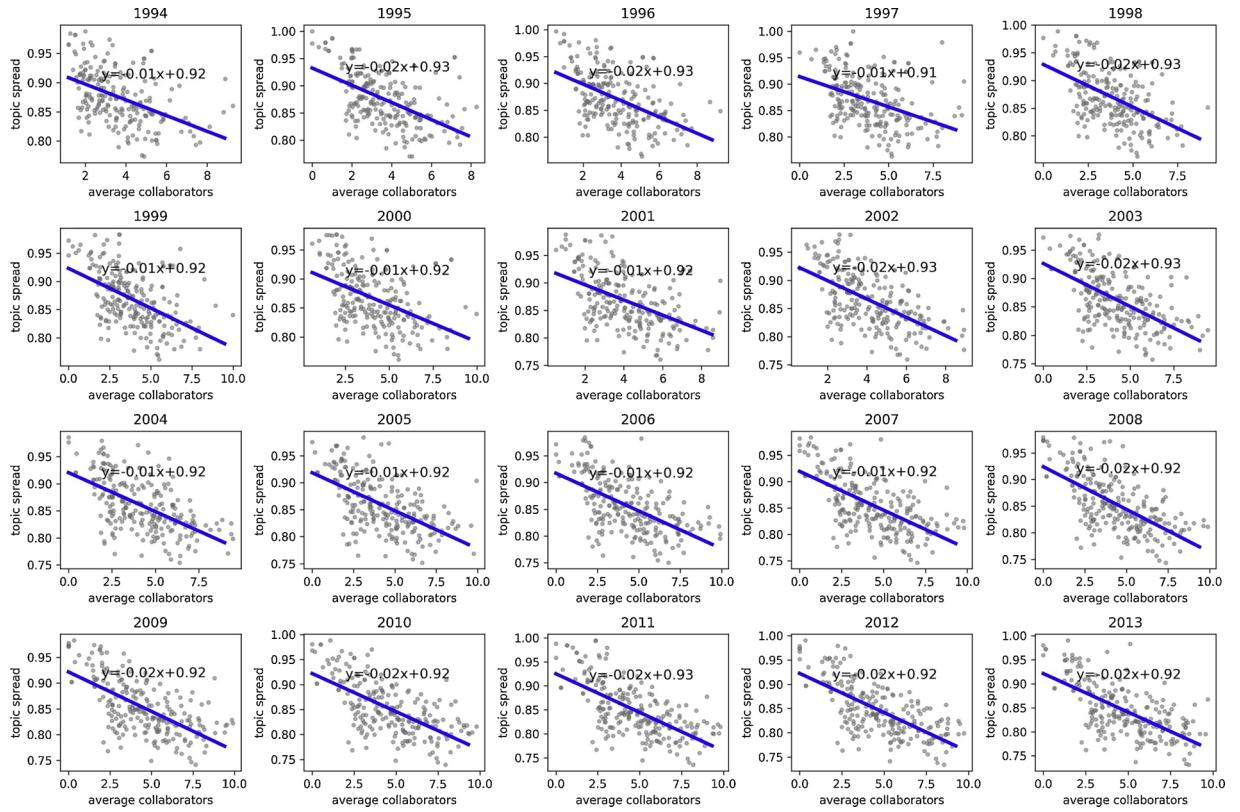
**Fig. A5. Illustration of the empirical distribution of angles and radius for all PACS codes from Fig. 3D.** we displayed the probability density function and not probability mass function. The former does not sum up to one, but these two functions are of the same shape.



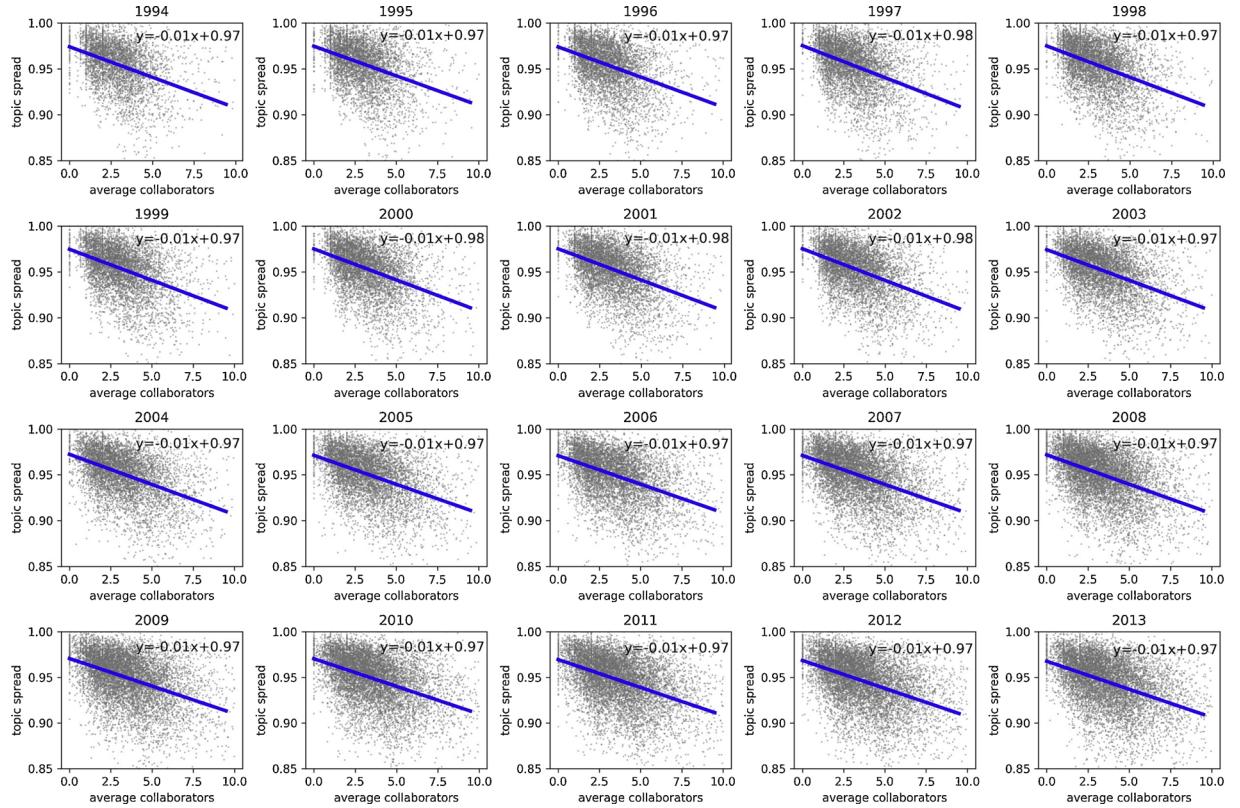
**Fig. A6.** Correlation between social density, semantic hierarchy, and semantic diversity.



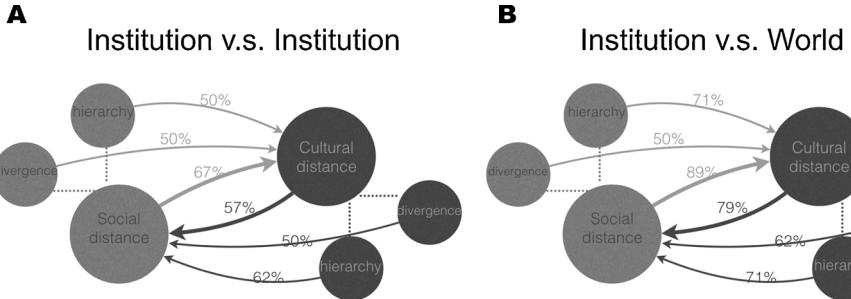
**Fig. A7.** Correlation between social densification and topic convergence across 13 fields, 1994–2013. In each subplot, a data point represents a field in science. The average collaborators (x-axis) and the topic spread (y-axis) are computed in the same way as in Fig. 5. The OLS regressions are used to fit the data points, a majority (85 %) of the regressions are significant ( $p < 0.05$ ) and all the regressions are marginally significant ( $p < 0.1$ ). The regression coefficients (the slopes of blue lines) vary from -0.01 to -0.03.



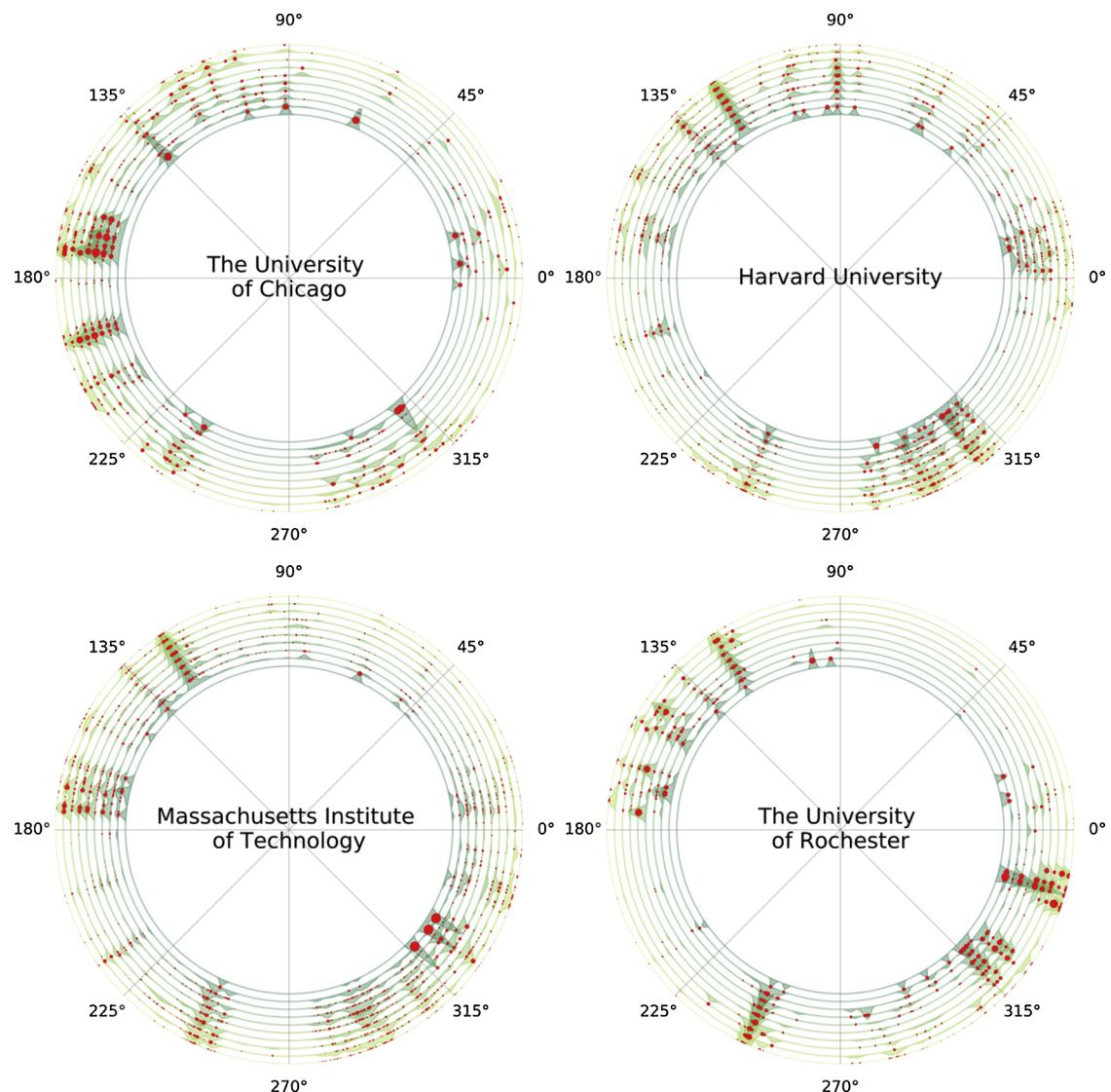
**Fig. A8. Correlation between social densification and topic convergence across more than 200 subfields, 1994–2013.** In each subplot, a data point represents a subfield in science. The subfields are based on WoS meta data of journal classifications. The average collaborators (x-axis) and the topic spread (y-axis) are computed in the same way as in Fig. 5. The OLS regressions are used to fit the data points, all the regressions are significant ( $p < 0.05$ ). The regression coefficients (the slopes of blue lines) vary from -0.01 to -0.02.



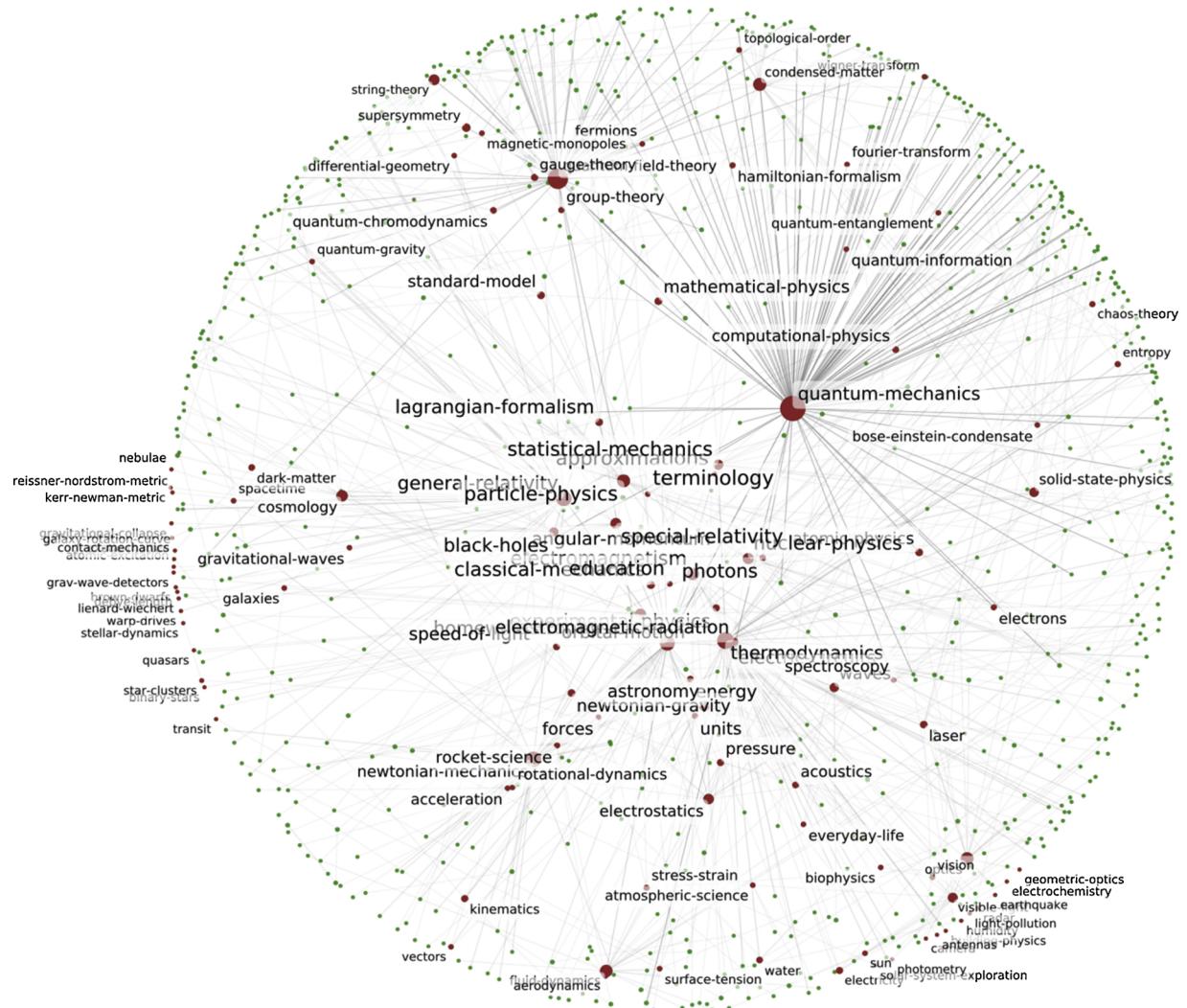
**Fig. A9. Correlation between social densification and topic convergence across over 10,000 journals, 1994–2013.** In each subplot, a data point represents a journal. The average collaborators (x-axis) and the topic spread (y-axis) are computed in the same way as in Fig. 5. The OLS regressions are used to fit the data points, all the regressions are significant ( $p < 0.05$ ). The regression coefficients (the slopes of blue lines) for all subplots are -0.01.



**Fig. A10.** We use Granger causality regression to explore whether the social distance of institutions predicts their semantic distance in future and vice versa. In addition, as the positions of institutions and the distance between them always involves two factors, angle (“divergence”) and radius (“hierarchy”), we can further investigate which factor contribute more in the prediction. We investigate the dynamics between pairwise institutions (“institution vs institution”, panel A), and also aggregate the results to explore how single institutions interact with all other institutions (“institution vs world”, panel B). In both panels, arrows point from the predictors to the predicted variables. The values associated with arrows were the percentages of positive coefficients in the Granger causality regression, which quantifies the predicting power in the positive direction. In general, the social distance tends to predict cultural distances slightly more than the converse, but both are consistent and significant.



**Fig. A11. Hyperbolic embedding of four universities.** The Gaussian kernel density estimation is used to visualize the concentration of the angles of PACS codes in green waves. Red dots visualize the data points of PACS codes, with a size proportional to the number of papers. Each circle represents a year from 2002 (the innermost circle) to 2011 (the outermost circle).



**Fig. A12.** Hyperbolic embedding tags from 47,282 questions in Physics Stack Exchange (<https://physics.stackexchange.com/>).

## References

- Anderson, C. (2006). *The long tail: Why the future of business is selling less of more*. Hachette books.

Anderson, P. W. (1972). More is different. *Science*. <https://doi.org/10.1126/science.177.4047.393>.

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs General and Applied*, 70(9), 1.

Barabasi, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439), 509–512.

Basov, N., & Brennecke, J. (2017). Duality beyond dyads: Multiplex patterning of social ties and cultural meanings. *Structure, Content and Meaning of Organizational Networks*, 87–112 Research in the Sociology of Organizations. Emerald.

Bauer, H. H. (1990). Barriers against interdisciplinarity: Implications for studies of science, technology, and society (STS). *Science, Technology & Human Values*, 15(1), 105–119.

Becker, J., Brackbill, D., & Centola, D. (2017). Network dynamics of social influence in the wisdom of crowds. *Proceedings of the National Academy of Sciences of the United States of America*, 114(26), E5070–76.

Bettencourt, L. M. A., Lobo, J.é, Dirk Helbing, C. K., & West, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences of the United States of America*, 104(17), 7301–7306.

Biron, L. (2015). Our flat universe. *Symmetry: Dimensions of Particle Physics*, 7(April), 2015 [https://www.symmetrymagazine.org/article/april-2015/our-flat-universe?email\\_issue=725](https://www.symmetrymagazine.org/article/april-2015/our-flat-universe?email_issue=725).

Bolukbasi, T., Chang, K.-W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man Is to computer programmer as woman is to homemaker? Debiasing word embeddings. *Advances in Neural Information Processing Systems*, 4349–4357.

Bonola, R. (1955). *Non-euclidean geometry: A critical and historical study of its development*. Courier Corporation.

Boudreau, K. J., & Lakhani, K. R. (2015). 'Open' disclosure of innovations, incentives and follow-on reuse: Theory on processes of cumulative innovation and a field experiment in computational biology. *Research Policy*, 44(1), 4–19.

Bourdieu, P. (1984). *Distinction. Translated by richard nice*. Cambridge, MA: Harvard Univer.

Bourdieu (1989). Social space and symbolic power. *Sociological Theory*, 7(1), 14–25.

Braun, T., & Schubert, A.ás (2003). A quantitative view on the coming of age of interdisciplinarity in the sciences 1980–1999. *Scientometrics*, 58(1), 183–189.

- Breiger, R. L., & Puetz, K. (2015). Culture and networks. *The International Encyclopedia of the Social and Behavioral Sciences*, 557–562.
- Burt, R. S. (2004). Structural holes and good ideas. *The American Journal of Sociology*, 110(2), 349–399.
- Burt, R. S. (2007). Brokerage and closure: An introduction to social capital. *OUP Oxford*.
- Burt, R. S. (2009). *Structural holes: The social structure of competition*. Harvard University Press.
- Buskens, V., & van de Rijt, A. (2008). Dynamics of networks if everyone strives for structural holes. *The American Journal of Sociology*, 114(2), 371–407.
- Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. *Science*, 356(6334), 183–186.
- Carley, K. (1986). An approach for relating social structure to cognitive structure. *The Journal of Mathematical Sociology*, 12(2), 137–189.
- Carley, K. (1991). A theory of group stability. *American Sociological Review*, 56(3), 331–354.
- Carlino, G. A., Chatterjee, S., & Hunt, R. M. (2006). *Urban density and the rate of invention*. August <https://doi.org/10.2139/ssrn.926328>.
- Cetina, K. K. (2009). *Epistemic cultures: How the sciences make knowledge*. Harvard University Press.
- Chamberlain, B. P., Clough, J., & Deisenroth, M. P. (2017). Neural embeddings of graphs in hyperbolic space. *arXiv [stat.ML]*. arXiv. <http://arxiv.org/abs/1705.10359>.
- Danchev, V., Rzhetsky, A., & Evans, J. A. (2019). Centralized communities more likely generate non-replicable results. *eLife*. <http://arxiv.org/abs/1801.05042>.
- Danescu-Niculescu-Mizil, C., Lee, L., Pang, B., & Kleinberg, J. (2012). Echoes of power: Language effects and power differences in social interaction. *Proceedings of the 21st International Conference on World Wide Web* (pp. 699–708).
- Danescu-Niculescu-Mizil, C., West, R., Jurafsky, D., Leskovec, J., & Potts, C. (2013). No Country for Old members: User lifecycle and linguistic change in online communities. *Proceedings of the 22nd International Conference on World Wide Web* 307–18.
- Deegan, R. D., Bakajin, O., Dupont, T. F., Huber, G., Nagel, S. R., & Witten, T. A. (1997). Capillary flow as the cause of ring stains from dried liquid drops. *Nature*, 389(October), 827.
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv [cs.CL]*. arXiv. <http://arxiv.org/abs/1810.04805>.
- DiMaggio, P. (1987). Classification in Art. *American Sociological Review*, 440–455.
- Dunham, D., et al. (2009). The symmetry of ‘Circle limit IV’ and related patterns. *BRIDGES*, 163–168.
- Elberse, A. (2008). Should you invest in the long tail? *Harvard Business Review*, 86(7/8), 88.
- Erickson, B. H. (1988). The relational basis of attitudes. *Social Structures: A Network Approach*, 99(121), 443–475.
- Feld, S. L. (1991). Why your friends have more friends than you do. *The American Journal of Sociology*. <https://doi.org/10.1086/229693>.
- Fellbaum, C. (2005). WordNet and wordnets. In A. Barber (Ed.), *Encyclopedia of language and linguistics* Elsevier 2–665.
- FIRTH, & J. R (1957). A synopsis of linguistic theory, 1930–1955. *Studies in Linguistic Analysis*. <https://ci.nii.ac.jp/naid/10020680394/>.
- Fleming, L., Mingo, S., & Chen, D. (2007). Collaborative brokerage, generative creativity, and creative success. *Administrative Science Quarterly*, 52(3), 443–475.
- Fourcade, M., Ollion, E., & Algan, Y. (2015). The superiority of economists. *Revista de EconomÃa Institucional*, 17(33), 13–43.
- Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 40(1), 35–41.
- Furnas, G. W., Deerwester, S., Dumais, S. T., Landauer, T. K., Harshman, R. A., Streeter, L. A., et al. (1988). Information retrieval using a singular value decomposition model of latent semantic structure. *Proceedings of the 11th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 465–480).
- Garg, N., Schiebinger, L., Jurafsky, D., & James, Z. (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences of the United States of America*, 115(16), E3635–44.
- Garicano, L. (2000). Hierarchies and the organization of knowledge in production. *The Journal of Political Economy*, 108(5), 874–904.
- Goel, S., Broder, A., Gabrilovich, E., & Pang, B. (2010). Anatomy of the Long tail: Ordinary people with extraordinary tastes. *Proceedings of the Third ACM International Conference on Web Search and Data Mining* (pp. 201–210).
- Granovetter, M. (1974). *Finding a job: A study on contacts and careers*. Cambridge, Mass: Harvard University Press.
- Granovetter, M. S. (1973). The strength of weak ties. *The American Journal of Sociology*, 78(6), 1360–1380.
- Guimerà, 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.
- Harari, Y. N. (2014). *Sapiens: A brief history of Humankind*. Random House.
- Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences of the United States of America*, 101(46), 16385–16389.
- Huntington, S. P. (2000). *The clash of civilizations? Culture and politics*. Springer99–118.
- Hutchinson, I. (2011). *Monopolizing knowledge*. Lulu.com.
- Jaeger, H. M., & Nagel, S. R. (1992). Physics of the granular state. *Science*, 255(5051), 1523–1531.
- Johnson, W. B., & Lindenstrauss, J. (1984). Extensions of lipschitz mappings into a hilbert space. *Contemporary Mathematics and Statistics*, 26(189–206), 1.
- Joshi, A., & Roh, H. (2009). The role of context in work team diversity research: A meta-analytic review. *Academy of Management Journal* *Academy of Management*, 52(3), 599–627.
- Kandel, D. B. (1978). Homophily, selection, and socialization in adolescent friendships. *The American Journal of Sociology*, 84(2), 427–436.
- Klemm, K., Eguiluz, Víctor M., Toral, R.úl, & Miguel, M. S. (2003). Nonequilibrium transitions in complex networks: A model of social interaction. *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics*, 67(2 Pt 2), 026120.
- Kozlowski, A. C., Taddy, M., & Evans, J. A. (2018). The geometry of culture: Analyzing meaning through word embeddings. *arXiv [cs.CL]*. arXiv. <http://arxiv.org/abs/1803.09288>.
- Kozlowski, A. C., Taddy, M., & Evans, J. A. (2019). The geometry of culture: Analyzing the meanings of class through word embeddings. *American Sociological Review*, 84(5), 905–949.
- Krackhardt, D., & Kilduff, M. (1990). Friendship patterns and culture: The control of organizational diversity. *American Anthropologist*, 92(1), 142–154.
- Krioukov, D., Kitsak, M., Sinkovits, R. S., Rideout, D., Meyer, D., & Boguñá, M.án (2012). Network cosmology. *Scientific Reports*, 2(November), 793.
- Krioukov, D., Papadopoulos, F., Kitsak, M., Vahdat, A., & Boguñá, M.án (2010). Hyperbolic geometry of complex networks. *Physical Review E, Statistical, Nonlinear, and Soft Matter Physics*, 82(3 Pt 2), 036106.
- Krohn, M. D. (1986). The web of conformity: A network approach to the explanation of delinquent behavior. *Social Problems*, 33(6), s81–93.
- Kuhn, T. S. (1970). *The structure of scientific revolutions*. Harvard University Press.
- Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Harvard University Press.
- Lazarsfeld, P. F., Merton, R. K., et al. (1954). Friendship as a social process: A substantive and methodological analysis. *Freedom and Control in Modern Society*, 18(1), 18–66.
- Lee, M., & Martin, J. L. (2018). Doorway to the dharma of duality. *Poetics*, 68(June), 18–30.
- Lewis, K., & Kaufman, J. (2018). The conversion of cultural tastes into social network ties. *The American Journal of Sociology*, 123(6), 1684–1742.
- Leydesdorff, L., & Ivanova, I. A. (2014). Mutual redundancies in interhuman communication systems: Steps toward a Calculus of processing meaning. *Journal of the Association for Information Science and Technology*, 65(2), 386–399.
- Lizardo, O. (2014). Omnivorousness as the bridging of cultural holes: A measurement strategy. *Theory and Society*, 43(3), 395–419.
- Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences*, 108(22), 9020–9025.
- Malone, E. (2005). *Linguistics and languages; a special report: Endangered languages* 2005. National Science Foundation [https://www.nsf.gov/news/special\\_reports/linguistics/endangered.jsp](https://www.nsf.gov/news/special_reports/linguistics/endangered.jsp).
- Mannix, E., & Neale, M. A. (2005). What differences make a difference? The promise and reality of diverse teams in organizations. *Psychological Science in the Public Interest: A Journal of the American Psychological Society*, 6(2), 31–55.
- March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, 2(1), 71–87.
- Mark, N. (1998). Birds of a feather sing together. *Social Forces; A Scientific Medium of Social Study and Interpretation*, 77(2), 453–485.

- Martin, J. L., & Lee, M. (2018). A formal approach to meaning. *Poetics*, 68(June), 10–17.
- McFarland, D. A., Jurafsky, D., & Rawlings, C. (2013). Making the connection: Social bonding in courtship situations. *The American Journal of Sociology*, 118(6), 1596–1649.
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1), 415–444.
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). “Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, & K. Q. Weinberger (Eds.). *Advances in neural information processing systems 26* (pp. 3111–3119). Curran Associates, Inc.
- Mohr, J. W. (2000). Introduction: Structures, institutions, and cultural analysis. *Poetics*, 27(2), 57–68.
- Montgomery, G. M., & Largent, M. A. (2015). *A companion to the history of american science*. John Wiley & Sons.
- Nickel, M., Kiela, D., et al. (2017). Poincaré embeddings for learning hierarchical representations. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, & S. Vishwanathan (Eds.). *Advances in neural information processing systems 30* (pp. 6338–6347). Curran Associates, Inc.
- Nielsen, M. (2012). *Reinventing discovery: The new era of networked science*. Princeton University Press.
- Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1964). *The measurement of meaning*. University of Illinois Press.
- Pachucki, M. A., & Breiger, R. L. (2010). Cultural holes: Beyond relationality in social networks and culture. *Annual Review of Sociology*, 36(1), 205–224.
- Page, S. E. (2008). *The difference: How the power of diversity creates better groups, firms, schools, and societies - new edition*. Princeton University Press.
- Palla, G., Barabási, A.-L., & Vicsek, T. (2007). Quantifying social group evolution. *Nature*, 446(April), 664.
- Papadopoulos, F., Maksim Kitsak, M., Ángeles Serrano, M.Án. Bñá, & Krioukov, D. (2012). Popularity versus similarity in growing networks. *Nature*, 489(7417), 537–540.
- Papadopoulos, F., Psomas, C., & Krioukov, D. (2015). Network mapping by replaying hyperbolic growth. *IEEE/ACM Transactions on Networking*, 23(1), 198–211.
- Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543.
- Perozzi, B., Al-Rfou, R., & Skiena, S. (2014). DeepWalk: Online learning of social representations. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Poon, W. C. K. (2002). The physics of a model colloid–polymer mixture. *Journal of Physics. Condensed Matter: An Institute of Physics Journal*. <http://iopscience.iop.org/article/10.1088/0953-8984/14/33/201/meta>.
- Prigogine, I. (1947). Le rendement thermodynamique de La thermodiffusion. *Physica*. [https://doi.org/10.1016/0031-8914\(47\)90003-7](https://doi.org/10.1016/0031-8914(47)90003-7).
- Prigogine, I., & Defay, R. (1944). *Thermodynamique chimique conformément aux méthodes de gibbs et De donder*. Vol. 2 Desoer.
- Regier, T. (2008). Whorf was half right. *PsychEXTRA Dataset*. <https://doi.org/10.1037/e527312012-204>.
- Rogers, E. M. (2010). *Diffusion of innovations* (4th edition). Simon and Schuster.
- Rogers, E. M., & Floyd Shoemaker, F. (1971). *Communication of innovations; a cross-cultural approach*. <https://eric.ed.gov/?id=ED065999>.
- Roth, C., & Cointet, J.-P. (2010). Social and semantic coevolution in knowledge networks. *Social Networks*, 32(1), 16–29.
- Saint-Charles, J., & Mongeau, P. (2009). Different relationships for coping with ambiguity and uncertainty in organizations. *Social Networks*, 31(1), 33–39.
- Salganik, M. J., & Watts, D. J. (2008). Leading the herd astray: An experimental study of self-fulfilling prophecies in an artificial cultural market. *Social Psychology Quarterly*. <https://doi.org/10.1177/019027250807100404>.
- Shannon, C., & Weaver, W. (1963). *The mathematical theory of communication*. University of Illinois Press.
- Shi, F., Teplitskiy, M., Dueðe, E., & Evans, J. (2017). The wisdom of polarized crowds. *arXiv [cs.SI]*. arXiv. <http://arxiv.org/abs/1712.06414>.
- Shuicheng, Y., Xu, D., Zhang, B., & Zhang, H.-J. (2005). Graph embedding: A General framework for dimensionality reduction. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2:830–37 Vol. 2.
- Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, 42(3/4), 425–440.
- Strang, D., & Macy, M. W. (2001). In search of excellence: Fads, success stories, and adaptive emulation. *The American Journal of Sociology*. <https://doi.org/10.1086/323039>.
- Suwowiecki, J. (2004). The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business. *Economies, Societies and Nations*, 296.
- Uzzi, B., & Spiro, J. (2005). Collaboration and creativity: The small world problem. *The American Journal of Sociology*, 111(2), 447–504.
- Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science*, 342(6157), 468–472.
- Vaan, M., Vedres, B., & Stark, D. (2015). Game changer: The topology of Creativity1. *AJS; American Journal of Sociology*, 120(4), 1–51.
- Valente, T. W. (1996). Network models of the diffusion of innovations. *Computational and Mathematical Organization Theory*, 2(2), 163–164.
- Vedres, B.Ázs, & Stark, D. (2010). Structural folds: Generative disruption in overlapping groups. *The American Journal of Sociology*, 115(4), 1150–1190.
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge University Press.
- Wasserman, S., & Pattison, P. (1996). Logit models and logistic regressions for social networks: I. An introduction to Markov Graphs Andp. *Psychometrika*, 61(3), 401–425.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘Small-World’ networks. *Nature*, 393(6684), 440–442.
- Whorf, B. L., & Chase, S. (1956). In J. B. Carroll (Ed.). *Language, thought and reality, selected writings of Benjamin Lee Whorf* Foreword by Stuart Chase. Mass.
- Wittgenstein, L., Anscombe, G. E. M., & Wittgenstein, L. (1953). *Philosophical investigations... translated by GEM anscombe. (Philosophische Untersuchungen.)*. Oxford: Eng. & Ger.
- Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, 330(6004), 686–688.
- Wu, L., Wang, D., & Evans, J. A. (2017). Large teams have developed science and technology; small teams have disrupted it. September[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3034125](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3034125).
- Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, 316(5827), 1036–1039.
- Young, H. D., & Freedman, R. A. (2013). *University physics with modern physics technology update: International edition*. Pearson Educacion.
- Zuckerman, E. W., & Jost, J. T. (2001). What Makes You Think You're so Popular? Self-Evaluation Maintenance and the Subjective Side of the ‘Friendship Paradox’. *Social Psychology Quarterly*. <https://doi.org/10.2307/3090112>.

**Linzhuo Li** is a PhD candidate in the Department of Sociology at the University of Chicago. Before that he earned his BA and MA degree from Tsinghua University. His research examines economic and cultural change, and seeks to understand how order emerges, maintains or breaks down in cultural and economic systems. In addition to this project, he has explored rhetorical and ideological transformation in China's newspaper *People's Daily* from 1946 to 2003.

**Lingfei Wu** is an assistant professor in the School of Computing and Information at the University of Pittsburgh. Before that he earned a PhD in Communication from the City University of Hong Kong. His research interests focus on the future of research, education, and employment in an AI economy driven by knowledge production. His research has been published in peer-reviewed journals, including Nature, PNAS, Scientific Reports, etc., and was covered by the New York Times, Forbes, the Atlantic, and other top media outlets. He is also a core, founding member of the Swarma Club, one of the largest and most active open-science communities in China.

**James Evans** is Professor of Sociology, Director of Knowledge Lab and the Computational Social Science program, and member of the Committee on Conceptual and Historical Studies of Science at the University of Chicago, and External Professor at the Santa Fe Institute. His research focuses on the collective system of thinking and knowing, ranging from the distribution of attention and intuition, the origin of ideas and shared habits of reasoning to processes of agreement, certainty, and understanding. He is especially interested in innovation—how new ideas and technologies emerge—and the role that social and technical institutions like the Internet, markets, and collaborations play in collective cognition and discovery.