The Structure and Dynamics of Disciplines

A Comparative Analysis

Vladimir Borel

A thesis presented for the degree of
Doctor of Philosophy

Sociology
University of California, Riverside
United States
April 13, 2023

Abstract

This paper examines the structure and dynamics of science across 16 fields by analyzing both bibliographic and semantic networks. The study finds that there are few structural differences between the fields, indicating that the underlying structure of scientific knowledge is relatively uniform. The analysis reveals that there are common patterns of co-citation and term co-occurrence, highlighting the interconnected nature of scientific research.

1 Introduction

The debates on the structure and dynamics of science have often been based on abstract ideas rather than empirical evidence. However, the recent trend towards digitalization of peer-reviewed scientific articles has made it possible to test these theories using empirical data. This provides a valuable opportunity to advance our understanding of the structure and dynamics of science.

These theories ask a range of questions about the nature of science and its relationship to other fields of knowledge. Some of the key questions they address include:

- Are there regular patterns in the way that the structure of science changes?
- Is science a distinct field of knowledge or is it embedded in larger domains, is the knowledge generated by scientific methods fundamentally different from other types of knowledge?
- Are there clear distinctions between science and non-science or between different types of science (e.g. "hard" and "soft" sciences)?

These theories also consider the potential impact of factors such as technological innovations, social, political, and economic events on the restructuring of science.

This paper aims to make three types of contributions. The first set of contributions is theoretical and involves synthesizing and unifying different conceptions of scientific development in graph-theoretic terms. The second set of contributions is empirical and involves deriving and testing hypotheses from the literature on scientific development. The third set of contributions is practical and focuses on developing a standard analysis pipeline applicable to a wide variety of fields. This is necessary due to the increasing difficulty of coordinating research efforts and bridging "academic silos" in the face of the exponentially increasing output of technical knowledge and the fact that researchers only have local knowledge about their immediate neighbors.

When discussing the structure of a body of knowledge, I am referring to two main aspects: the citation structure and the semantic structure. The citation structure refers to the way in which ideas and information are connected through references and citations, while the semantic structure refers to the meaning and relationships between the concepts and ideas within the body of knowledge. Both of these aspects of structure are important for understanding the organization and development of a particular field of knowledge.

2 Networks

2.1 Citation

A citation network represented as a directed unweighted graph, with nodes representing individual documents and edges indicating a citation from one document to another.

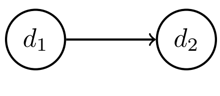

Figure 1: Citation

Figure 1 depicts a citation diad in which document d_1 cites document d_2.

2.2 Co-Citation

A co-citation network represented as an undirected weighted graph, with nodes corresponding to individual documents and edge weights representing the frequency with which two documents are co-cited.

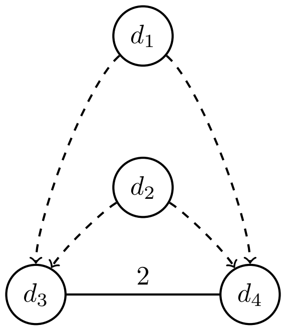

Figure 2: Co-Citation

As shown in Figure 2, the edge weight of 2 between documents d_3 and d_4 reflects the fact that they are co-cited by two documents, d_1 and d_2.

2.3 Co-Occurrence

A co-occurrence network represented as an undirected weighted graph, with nodes representing individual terms and edge weights indicating the frequency with which two terms co-occur within a document.

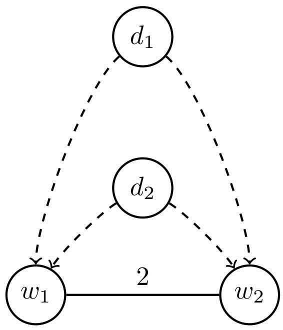

Figure 3: Co-Occurrence

As shown in Figure 3, the edge weight of 2 between terms w_3 and w_4 reflects the fact that they co-occur in both documents, d_1 and d_2.

3 Data

Having established the possibility of linking citation networks and concept networks as clusters at various levels of analysis, I will now delve into a discussion of the data used to identify the structure and dynamics of scientific communities. The analysis of this data will enable me to evaluate the accuracy of the various theories of scientific development discussed earlier and to provide a realistic account of the fields analyzed, given that the data is reliable and sufficient.

3.1 Sample

There are two main ways of bounding citation networks. The first involves sampling based on journals (Beam et al., 2014; Moody and Light, 2006), while the second involves sampling based on a topical search (Borrett et al., 2014; Callon et al., 2005; Chavalarias and Cointet, 2013). In this study, I used the former method, selecting articles that were published in one of the 5 peer-reviewed journals with the highest impact factor.

The main advantage of sampling based on journals is that it will allow us to focus on articles published in high-quality, reputable journals. This can help ensure that the data used in the study is of a high quality and is representative of the broader field of research. Additionally, journal-based sampling can make it easier to compare the results of different studies, as it provides a common frame of reference.

Bradford's law is a principle that describes the relationship between the number of scientific journals in a particular field and the number of citations that articles in those journals receive. It states that a small number of highly-cited journals tend to publish a disproportionate number of highly-cited articles, while a large number of low-cited journals tend to publish a disproportionate number of low-cited articles.

This law can be used to support the choice of only selecting a small number of journals in each scientific field under study. Focusing on a select group of highly-cited journals, can increase the chances of finding and reading the most influential and highly-cited articles in that field. It's worth noting that Bradford's law is not a hard and fast rule, and there may be cases where it is necessary or beneficial to review a larger number of journals in a particular field. However, as a general principle, Bradford's law can be a useful guide when selecting journals for review.

The data used in this study consists of peer-reviewed scientific articles collected from Web of Science. These data include each document's citation information, as well as its title, abstract, key words, and publication date.

3.2 Fields

The fields being analyzed include Artificial Intelligence, Astronomy & Astrophysics, Economics, Ethnic & Cultural Studies, Gender Studies, Genetics & Genomics, Geometry, Geophysics, Human Resources & Organizations, Immunology, International Business, Language & Linguistics, Law, Material Engineering, Neurology, Political Science, Probability & Statistics, Sociology, Biochemistry. A complete list of the journals for each of the fields can be found in the Appendix.

These fields were selected for analysis because they represent a diverse range of disciplines, including both hard sciences, as well as soft sciences. This diversity allows for a comprehensive examination of how scientific communities function across different areas of study. Additionally, these fields are dynamic and rapidly advancing, providing a rich source of data for analysis and making them well suited for studying the structure and dynamics of scientific communities.

Furthermore, these fields were selected because they enable not only the analysis of individual fields but also the use of a comparative framework. Comparing and contrasting the structures and dynamics of scientific communities across different fields allows for a deeper understanding of the similarities and differences among them, and the opportunity to identify common as well as idiosyncratic patterns and trends that may be relevant to theories of scientific development.

3.3 Tools

Python was used for the analysis (van Rossum and Drake, 2010), and several libraries were particularly noteworthy, including Networkx (Hagberg et al., 2008) and PyMC (Salvatier et al., 2016).

4 ERGMs

Exponential Random Graph Models are used to model and test different hypotheses about the structural features of social networks. This approach enables us to investigate the different local and global properties of a network and establish their importance in explaining its formation. By testing for various network properties, such as triangles or the average shortest path between nodes, we can determine whether these properties played a significant role in the network's formation.

According to the Hammersly and Clifford Theorem (1971), any network model can be expressed in the exponential family with counts of graph statistics. The probability of observing a particular graph y on n nodes out of the set of all possible graphs on n nodes, Y, denoted as P, can be calculated using a set of network statistics S and corresponding parameters θ.

$$P_{y,\theta}(Y = y|\theta) = \frac{exp\{\theta^T S(y)\}}{\sum\limits_{y' \in Y} exp\{\theta^T S(y')\}}$$

The denominator is a normalizing constant that guarantees the distribution adds up to one. This constant requires summing over space of possible networks on n nodes. However, the number of possible configurations (size of Y) grows exponentially with the number of nodes, specifically to $2^{(n(n-1)/2)}$ for undirected graphs and $2^{(n(n-1))}$ for directed graphs. Thus, an exact computation of this sum is not feasible.

Because of this, it is custommary to use Marcov Chain Monte Carlo method to generate samples. It estimate the values of the parameters that maximize the likelihood of the observed network. These estimates represent the strength and direction of the effects of various network statistics on the likelihood of observing the network.

The above formula can be rewritten in terms of the covariate vector θ:

$$logit(Y_{ij}|y_{ij}^c) = \theta' \delta(y_{ij})$$

Where

- y_{ij}^c is the complement of y_{ij}, i.e. all dyads in the network other than y_{ij}
- y_{ij}^+ as the same network as y except that $y_{ij} = 1$
- y_{ij}^- as the same network as y except that $y_{ij} = 0$
- $\delta(y_{ij})$ is given by $g(y_{ij}^+) - g(y_{ij}^-)$ which measures how the sufficient statistic $g(y)$ changes if the (i, j)th edge is "toggled" on or off.

In sum, for each of the following sufficient statistics a $n \times n$ matrix s is constructed. The entry s_{ij} indicates how the presence of the edge between i and j changes the network statistic, holding the rest of the network constant.

5 Measurements

5.1 Density effect (Edges)

Density refers to the proportion of connections in a social network relative to the total possible connections.

This is interpreted as the intercept of the model. The entry of the ‘density‘ matrix M_{ij} indicates how the presence of that edge changes the number of edges in the graph, holding the rest of the network constant. It a matrix of ones - with the upper right triangle masked in the case of undirected networks.

5.2 Small-World effect (Triangles)

Social networks tend to exhibit a much higher number of triangles and transitive triads than what would be predicted by random graphs with comparable density.

The entry of the triangle matrix M_{ij} indicates how the presence of that edge changes the number of triangles in the graph, holding the rest of the network constant.

Overall tendency of networks to be composed of local densities (closure). Are the ties randomly distributed of do they form local structures. A graph with a low triangle effect would look homogeneous Distribution of substructures is what we would expect from a random graph Lumps appear (local structures)

5.3 Clique effect (Cliques)

Tendecy for there to be parts of the graph where multiple nodes all co-occur more than by random change If I discuss A all other concepts in the clique it belongs to. highly related to each other distinct topical or functional units within the network

This build on triangles (which are a clique) but captures more information It asks whether triangles overlap to form larger structurees rather than just isolated (random) triangles Densly connected always connected

5.4 Silo effect (Components)

silos refer to isolated areas of research or knowledge that are not well connected to other areas or fields limit the flow of information and influence within the network Hyper specialization of field of knowledge Talk to one another but never talk to nodes outside their primary group

5.5 Popularity effect (k-star)

The popularity effects in most classes were not obvious, which meant the degree (the total number of actors selecting an individual) of individuals in the class networks had little difference.

What we mean by a strong k star effect Control for 9 stars in order to determine if 10 star is significant 10 star is a bunch of 9 stars

5.6 Mediating effect (betweenness centrality)

Nodes with high betweenness centrality in this type of network may represent important bridge terms or pivot terms that connect different topics or themes within the corpus of documents. The betweenness centrality c_b of node n is given by:

$$c_b(i) = \sum_{j \neq k} \frac{\sigma_{jk}(i)}{\sigma_{jk}}$$

Where σ_{ij} is the total number of shortest paths from node i to node j and $\sigma_{ij}(n)$ is the number of those paths that pass through node n. In other words, it is the proportion of all shortest paths between nodes i and j that pass through node n. The average betweenness centrality over all nodes of the network is taken as another network statistic.

Indicates good integration but maybe structural fragility (if central nodes are removed, speperate components)

5.7 Community effect (Louvain)

The Louvain community detection method consists of two main steps. Initially, each node is assigned to a separate community. Then, for each node, the algorithm attempts to optimize the modularity of the network by evaluating the potential gain in modularity achieved by moving the node to each of its neighboring communities. If no gain is achieved, the node remains in its original community.

In the second step, a new network is constructed where each node represents a community from the previous step. The edges between the new nodes are weighted by the sum of the weights of the edges between the nodes in the corresponding communities in the original network. The Louvain method is then applied to this new network, and the process is repeated until no further improvement in modularity can be achieved.

Modularity is a measure of the degree of segregation of a network into communities. It is calculated as the difference between the fraction of edges within a given group and the expected fraction of edges if they were randomly distributed in the network. The change in modularity ΔQ achieved by moving the node to each of its neighboring communities is measured as

$$Q = \frac{1}{2m} \sum_{i,j} [A_{ij} - \frac{k_i k_j}{2m}] \delta(c_i, c_j)$$

where:

- Q is the modularity index
- m is the total number of edges in the network
- A_{ij} is the weight of the edge between nodes i and j
- k_i and k_j are the degrees of nodes i and j, respectively
- c_i and c_j are the community assignments of nodes i and j
- $\delta(c_i, c_j)$ is the Kronecker delta, which is equal to 1 if nodes i and j are in the same community and 0 otherwise.

5.8 Importance effect (closeness)

Nodes with high closeness centrality in these networks may represent important key concepts or core ideas that are central to the understanding of the corpus of documents or, conversly, those that constitute the periphery within the network and are relatively more isolated. On a graph of n nodes, closeness centrality is defined as

$$c_c(i) = \frac{n-1}{\sum\limits_{i \neq j} d(i,j)}$$

Where $d(i,x)$ is the distance (length of the shortest path) between nodes i and y. The average closeness centrality over all nodes of the network is taken as another network statistic.

Integration of the field Concepts play integrating Separated by few or many logical steps

5.9 Community centrality (Eigenvector)

Eigenvector centrality is a network measure that assigns a score to each node in a network based on the node's connections to other high-scoring nodes. This measure takes into account not only the number of connections a node has, but also the quality or importance of those connections. The average closeness centrality over all nodes of the network is taken as another network statistic.

5.10 Concentration effect (Centralization)

Measure of the degree of concentration of edges. Centralization is obtained by computing the ratio of the differences between the highest scoring node and all other nodes for both the observed and a star network of the same size.

$$D(G) = \sum_i (max(c) - c_i)$$

where c_i is the centrality c of node i and $max(c)$ is the largest centrality score in graph G. by the maximum theoretical score for a graph with the same number of vertices:

$$C(G) = \frac{D(G)}{D(G_{star})}$$

5.11 Closure effect (Transitivity)

A triad is a group of three nodes in a graph. A triad can either be open or closed. An open triad is a group of three nodes that are connected by two edges, while a closed triad is a group of three nodes that are connected by three edges. Transitivity is defined as the ratio of the number of closed triads in the graph to the number of open triads in the graph.

$$T = \frac{3t_c}{t_o}$$

Where t_c is the number of closed triads and t_o is the number of open triads.

5.12 Inequality effect (Gini coefficient)

The Gini coefficient in a network can be defined as the ratio between the area between the Lorenz curve and the line of perfect equality (A) to the area under the line of perfect equality ($A + B$). The line of perfect equality is a straight line that represents a situation in which all nodes has the same number of edges. The Lorenz curve plots the cumulative proportion of edges against the cumulative proportion of nodes.

$$G = \frac{A}{A + B}$$

As A grows relative to B, the fraction tends towards 1, indicating perfect inequality. Conversely, As B grows raltive to A, the faction tends towards 0, indicating perfect equality.

5.13 Geodesic distance

To calculate the average geodesic distance, the shortest path length between every pair of nodes is first computed, and then the average of all these distances is taken. The result is a single value that represents the typical distance between any two nodes in the graph.

6 Results

Table 1: ERGM Results

	Artificial Intelligence		Economics		Ethnic & Cultural Studies		Gender Studies		Genetics & Genomics	
	G_c	G_o	G_c	G_o	G_c	G_o	G_c	G_o	G_c	G_o
Density	-3.712*** (0.379)	-5.224*** (0.349)	-4.884*** (0.339)	-4.97*** (0.323)	-3.93*** (0.385)	-4.203*** (0.246)	-2.193*** (0.302)	-4.793*** (0.298)	-3.509*** (0.492)	-5.313*** (0.34)
Triangles	3.353*** (0.633)	0.972*** (0.238)	2.489*** (0.522)	0.664** (0.241)	2.684*** (0.521)	2.685*** (0.379)	0.546* (0.297)	0.168 (0.513)	1.274** (0.502)	-0.498 (0.505)
Stars	0.061*** (0.009)	0.007*** (0.001)	0.01*** (0.003)	0.011*** (0.001)	0.018*** (0.003)	0.013*** (0.003)	0.013 (0.01)	0.006*** (0.0)	-0.027 (0.018)	0.003*** (0.0)
Betweenness	0.009 (1.007)	0.015 (0.989)	-0.009 (0.99)	-0.021 (1.02)	0.001 (0.982)	0.003 (0.996)	-0.005 (0.995)	0.002 (0.98)	-0.001 (0.965)	0.005 (0.993)
Closeness	-0.016 (1.009)	0.006 (1.006)	-0.067 (0.998)	-0.074 (0.986)	-0.029 (0.994)	-0.047 (1.002)	-0.023 (1.015)	0.021 (1.011)	-0.007 (1.001)	0.037 (0.998)
Eigenvector	0.018 (0.989)	-0.009 (1.002)	-0.01 (1.012)	-0.009 (1.024)	0.029 (1.024)	-0.007 (0.99)	0.005 (1.0)	-0.007 (0.985)	0.006 (0.998)	-0.012 (1.009)
Centralization	-0.008 (1.015)	-0.019 (1.013)	0.032 (1.017)	-0.432 (1.017)	0.001 (1.015)	0.057 (0.999)	-0.026 (1.004)	-0.091 (1.003)	-0.018 (0.998)	0.014 (1.03)
Clustering	0.032 (0.997)	-0.289 (0.978)	0.228 (0.973)	-0.046 (0.996)	0.253 (0.985)	-0.035 (0.996)	0.228 (0.99)	0.003 (1.002)	0.014 (0.998)	-0.295 (1.027)
Transitivity	0.032 (1.021)	0.012 (0.996)	0.098 (0.996)	-0.008 (0.993)	0.046 (0.996)	0.023 (1.002)	0.119 (1.018)	-0.022 (1.01)	0.013 (1.006)	-0.013 (0.989)
Cliques	-1.04*** (0.318)	0.016 (0.036)	-0.272 (0.211)	0.038 (0.044)	-0.648* (0.354)	-0.867*** (0.183)	0.05 (0.059)	0.226 (0.232)	0.339 (0.266)	0.496** (0.185)
Components	2.164*** (0.501)	-2.356** (0.859)	-0.423 (0.421)	-2.008*** (0.424)	1.58*** (0.417)	-1.714*** (0.441)	3.979*** (0.451)	-3.257*** (0.736)	1.443** (0.562)	-4.145*** (0.655)
Gini	-0.141 (0.998)	0.034 (1.02)	-0.328 (1.002)	-0.021 (1.007)	-0.263 (0.986)	0.013 (0.994)	-0.194 (1.021)	0.058 (1.006)	-0.036 (0.995)	0.01 (1.007)
Louvain	-5.128*** (0.391)	0.098 (0.339)	-0.723*** (0.171)	-0.446* (0.204)	-2.66*** (0.387)	-3.614*** (0.458)	-2.689*** (0.392)	0.362 (0.236)	-3.257*** (0.501)	0.123 (0.258)

	Geometry		Geophysics		Human Resources & Organizations		Immunology		International Business		Language & Linguistics	
	G_c	G_o	G_c	G_o	G_c	G_o	G_c	G_o	G_c	G_o	G_c	G_o
Density	-3.747*** (0.365)	-5.434*** (0.308)	-4.262*** (0.536)	-5.543*** (0.285)	-4.337*** (0.398)	-5.081*** (0.245)	-5.346*** (0.663)	-6.778*** (0.501)	-5.016*** (0.286)	-4.924*** (0.252)	-3.811*** (0.294)	-5.013*** (0.277)
Triangles	1.732*** (0.325)	2.435*** (0.481)	1.978** (0.762)	2.385*** (0.287)	1.996*** (0.614)	0.977** (0.394)	0.614 (0.787)	0.632 (0.822)	2.565*** (0.399)	1.334*** (0.409)	2.985*** (0.459)	1.402*** (0.366)
Stars	-0.013 (0.008)	0.013*** (0.001)	0.193*** (0.033)	0.012*** (0.002)	0.043*** (0.005)	0.004*** (0.0)	0.003** (0.001)	0.002*** (0.0)	0.013*** (0.002)	0.005*** (0.001)	0.021*** (0.006)	0.006*** (0.001)
Betweenness	-0.013 (1.005)	-0.435 (0.998)	0.0 (1.005)	-0.057 (1.026)	-0.006 (0.984)	-0.003 (1.026)	-0.018 (1.009)	0.0 (0.987)	-0.017 (1.005)	-0.01 (0.998)	-0.041 (1.002)	-0.005 (1.01)
Closeness	-0.033 (0.999)	-0.822 (1.013)	0.002 (0.997)	-0.062 (1.027)	-0.033 (0.997)	0.015 (0.98)	-0.021 (1.022)	-0.027 (1.004)	-0.054 (1.001)	-0.094 (1.009)	-0.01 (1.001)	0.022 (1.001)
Eigenvector	-0.022 (0.997)	-0.011 (0.982)	-0.019 (0.995)	-0.014 (0.998)	-0.016 (0.977)	-0.012 (1.004)	-0.01 (1.006)	-0.018 (1.015)	0.01 (1.006)	0.014 (1.017)	0.022 (0.994)	-0.004 (1.013)
Centralization	0.088 (0.986)	-0.351 (0.991)	-0.057 (0.994)	0.03 (1.005)	-0.081 (0.991)	-0.025 (0.986)	0.007 (1.019)	0.007 (1.004)	-0.037 (1.029)	0.008 (1.007)	0.03 (0.998)	-0.04 (0.993)
Clustering	0.148 (0.985)	-0.109 (1.0)	0.0 (0.992)	0.0 (0.99)	-0.049 (0.999)	-0.289 (1.0)	-0.189 (0.987)	-0.269 (0.994)	0.239 (1.003)	-0.111 (0.986)	0.156 (0.978)	-0.203 (0.982)
Transitivity	0.12 (1.003)	0.004 (0.991)	0.311 (1.0)	0.014 (1.002)	0.036 (0.996)	0.005 (0.996)	0.001 (0.998)	0.004 (0.991)	0.025 (0.991)	-0.016 (0.995)	0.035 (0.99)	-0.003 (0.984)
Cliques	-0.107* (0.065)	-0.489* (0.215)	0.177 (0.414)	-0.264*** (0.058)	-0.22 (0.215)	0.076 (0.156)	0.422 (0.533)	0.373 (0.414)	-0.463** (0.154)	0.0 (0.164)	-0.856*** (0.211)	-0.174 (0.134)
Components	0.495 (0.415)	-2.057*** (0.368)	3.684*** (0.594)	-2.922*** (0.334)	3.004*** (0.582)	-4.864*** (0.551)	1.95*** (0.55)	0.111 (0.82)	-1.859*** (0.359)	-2.162*** (0.403)	1.582*** (0.458)	-2.409*** (0.74)
Gini	-0.19 (0.985)	-0.208 (0.989)	-0.471 (0.988)	-0.133 (1.007)	-0.265 (0.988)	0.061 (0.998)	-0.211 (1.001)	-0.017 (1.018)	-0.088 (0.995)	0.011 (0.981)	-0.128 (0.999)	0.048 (1.011)
Louvain	-0.53** (0.187)	-0.675*** (0.16)	-4.313*** (0.454)	-0.58*** (0.144)	-5.475*** (0.423)	-0.515** (0.21)	-6.406*** (0.386)	-3.959*** (0.67)	-0.718*** (0.169)	-0.892*** (0.259)	-4.336*** (0.411)	-3.634*** (0.659)

	Material Engineering		Neurology		Political Science		Probability & Statistics		Sociology	
	G_c	G_o	G_c	G_o	G_c	G_o	G_c	G_o	G_c	G_o
Density	-2.486*** (0.361)	-4.592*** (0.288)	-3.794*** (0.438)	-5.132*** (0.336)	-2.581*** (0.644)	-4.825*** (0.292)	-4.482*** (0.346)	-5.157*** (0.34)	-2.984*** (0.312)	-5.485*** (0.318)
Triangles	2.494*** (0.577)	1.26*** (0.284)	1.137* (0.506)	1.447*** (0.242)	2.483*** (0.703)	2.067*** (0.503)	1.023*** (0.194)	0.583* (0.332)	2.227*** (0.315)	2.455*** (0.596)
Stars	0.047*** (0.01)	0.008*** (0.001)	-0.028* (0.017)	0.008*** (0.001)	0.046*** (0.014)	0.008*** (0.001)	0.013** (0.005)	0.005*** (0.0)	-0.026* (0.014)	0.003*** (0.0)
Betweenness	0.006 (0.994)	-0.021 (1.014)	0.025 (1.021)	-0.045 (0.998)	0.005 (0.991)	-0.005 (0.997)	-0.036 (0.989)	-0.007 (1.012)	0.001 (0.99)	-0.025 (1.002)
Closeness	-0.001 (0.974)	-0.057 (1.013)	-0.007 (0.97)	-0.087 (0.991)	-0.014 (0.982)	-0.027 (1.0)	-0.029 (1.003)	-0.036 (0.986)	-0.04 (0.988)	-0.03 (0.991)
Eigenvector	-0.006 (0.99)	-0.016 (0.982)	-0.026 (0.99)	0.01 (1.003)	-0.012 (0.985)	0.011 (0.981)	0.006 (0.983)	-0.021 (0.98)	-0.005 (1.009)	0.015 (0.999)
Centralization	0.016 (1.003)	0.007 (0.989)	-0.013 (0.993)	-0.013 (1.004)	0.007 (0.998)	-0.063 (1.043)	0.027 (0.995)	-0.02 (1.016)	0.088 (1.01)	0.007 (0.984)
Clustering	0.101 (1.009)	0.204 (0.996)	-0.025 (1.002)	0.14 (1.019)	0.163 (0.999)	0.133 (0.974)	0.201 (0.992)	-0.13 (0.986)	0.365 (0.998)	-0.051 (1.002)
Transitivity	0.06 (1.002)	-0.007 (1.022)	0.035 (1.003)	0.002 (0.994)	0.221 (1.029)	-0.035 (0.99)	0.033 (1.019)	0.003 (0.999)	0.074 (1.004)	-0.002 (0.998)
Cliques	-0.764*** (0.209)	-0.062 (0.075)	0.438 (0.32)	-0.08* (0.041)	-0.806 (0.542)	-0.559** (0.232)	-0.016 (0.018)	0.121* (0.061)	-0.227** (0.074)	-0.398 (0.294)
Components	3.358*** (0.529)	-1.502*** (0.413)	1.414*** (0.45)	-1.681*** (0.429)	2.241*** (0.427)	-2.158*** (0.445)	1.789*** (0.403)	-1.517** (0.515)	1.958*** (0.438)	-1.03* (0.471)
Gini	-0.101 (0.982)	-0.034 (1.006)	0.003 (0.975)	-0.018 (0.99)	-0.563 (1.002)	0.01 (0.99)	-0.097 (0.988)	0.002 (0.996)	-0.059 (1.032)	-0.015 (1.003)
Louvain	-6.43*** (0.384)	-0.06 (0.19)	-6.155*** (0.506)	-0.203 (0.152)	-1.352*** (0.16)	-0.665*** (0.198)	-3.05*** (0.257)	0.78** (0.288)	-6.059*** (0.512)	-5.339*** (0.543)

6.1 Density Effect (Edges)

Both co-citation and co-occurrence graphs are found to be sparse for every discipline, which is indicated by the negative and significant coefficients. A sparse network is characterized by a relatively low proportion of connections between individuals in comparison to the total number of potential connections that can exist within the network.

6.2 Small-World effect (Triangles)

As suggested by prior research, the social networks under investigation demonstrate a higher prevalence of triangles than what would be expected by random chance. Triangles, which represent a set of three documents or concepts who are mutually connected, are a common feature of social networks and are thought to play an important role in social cohesion and network structure (Kossinets and Watts, 2006; Newman, 2018). The observed excess of triangles in the current study is consistent with previous findings on the presence of triadic closure in social networks, where nodes tend to form connections with the neighbors of their neighbors (Burt, 2000).

All coefficients that are statistically significant for triangles exhibit a positive sign, which suggests that triangles are present in the social network under investigation at a rate higher than that which would be expected by chance alone.

References

Beam, Elizabeth, L. Gregory Appelbaum, Jordynn Jack, James Moody, and Scott A. Huettel. 2014. "Mapping the Semantic Structure of Cognitive Neuroscience." *Journal of Cognitive Neuroscience* 26:1949–1965.

Borrett, Stuart R., James Moody, and Achim Edelmann. 2014. "The Rise of Network Ecology: Maps of the Topic Diversity and Scientific Collaboration." *Ecological Modelling* 293:111–127.

Burt, Ronald S. 2000. "The Network Structure Of Social Capital." *Research in Organizational Behavior* 22:345–423.

Callon, M., J. Courtial, and F. Laville. 2005. "Co-Word Analysis as a Tool for Describing the Network of Interactions between Basic and Technological Research: The Case of Polymer Chemsitry." *Scientometrics* 22:155–205.

Chavalarias, David and Jean-Philippe Cointet. 2013. "Phylomemetic Patterns in Science Evolution—The Rise and Fall of Scientific Fields." *PLOS ONE* 8:e54847.

Hagberg, Aric A, Daniel A Schult, and Pieter J Swart. 2008. "Exploring Network Structure, Dynamics, and Function Using NetworkX." .

Kossinets, Gueorgi and Duncan J. Watts. 2006. "Empirical Analysis of an Evolving Social Network." *Science* 311:88–90.

Moody, James and Ryan Light. 2006. "A View from above: The Evolving Sociological Landscape." *The American Sociologist* 37:67–86.

Newman, Mark. 2018. *Networks*. Oxford, New York: Oxford University Press, second edition, new to this edition:, second edition, new to this edition: edition.

Salvatier, John, Thomas V. Wiecki, and Christopher Fonnesbeck. 2016. "Probabilistic Programming in Python Using PyMC3." *PeerJ Computer Science* 2:e55.

van Rossum, Guido and Fred L. Drake. 2010. *The Python Language Reference*. Number Pt. 2 in Python Documentation Manual / Guido van Rossum; Fred L. Drake [Ed.]. Hampton, NH: Python Software Foundation, release 3.0.1 [repr.] edition.