The importance of Philosophy in the network of academic concepts

(A filozófia jelentősége az akadémiai fogalmak hálózatában)

Author: Hanga Dormán *

Consultant: dr. László Gulyás

12 April 2021

*Eötvös Loránd University, Faculty of Psychology and Education, Computational and Cognitive Neuroscience MSc, Semester 4.

Abstract

This paper explores the network of Wikipedia articles connected by hyperlinks, with the specific purpose to prove the external relevance of *Philosophy* for other academic disciplines. First, field-specific network properties of articles are examined with the help of linear regressions fitted to field membership dummy variables. Next, measures of cross-field contribution to academic knowledge are calculated for individual fields based on a previous study of citation tendencies of published articles in theoretical and applied sciences. For a more detailed picture of one-to-one connections between fields, adjacency matrices of the entire network, as well as its subsets based on broad fields are analysed. To extract the essence of information represented by the cross-field network, statistically significant edges are identified via the backbone extraction method, revealing a loosely connected structure. Finally, the same backbones are extracted for the projections of the indirect hyperlink networks, in which edges signify paths of a given length instead of direct edges between two nodes. With the assumption that the thorough understanding of a concept requires familiarity with the terminology and accumulated knowledge it builds on directly and indirectly, the current analysis demonstrates that Wikipedia articles in *Philosophy* are of central importance in the conceptual structure of academic knowledge.

Absztrakt

Jelen dolgozat a Wikipédia cikkek hiperhivatkozások által összekötött hálózatát tárja fel azzal a céllal, hogy alátámassza a filozófia kiemelt jelentőségét az akadémiai fogalmak szempontjából. A terület-specifikus hálózatos tulajdonságokat lináris regressziók segítségével elemzem, melyeket a területhez tartozás dummy változóira illesztek. Ezek után az akadémiai területek közötti kontribúciót értékelem egy korábbi kutatás során definiált metrikák alapján, mely az elméleti és alkalmazott tudományok területén publikált cikkek hivatkozási tendenciáit vizsgálja. A részletesebb, területek közötti kapcsolatok feltárásának céljából a teljes hálózat, illetve a széles területek szerint felbontott hálózatok adjacencia mátrixát elemzem. Hogy a területek közötti hálózatból kinyerhessem a lényegi információt , a *backbone extraction* módszer segítségével beazonosítom a statisztikai értelemben szignifikáns éleket, ami fényt derít a hálózat lazán összekapcsolódó struktúrájára. Végül ugyanezzel a módszerrel kiválasztom az indirekt, adott hosszúságú ösvények alapján felépített hiperlink hálózat területek közti vetületeinek szignifikáns éleit. Azzal a feltételezéssel élve, hogy egy fogalom alapos megértéséhez szükség van arra a terminológiára és tudásra, amelyre a fogalom direkt vagy indirekt módon épít, jelen eredmények alátámasztják a filozófia kiemelt fontosságát az akadémiai tudás fogalmi struktúrájában.

Contents

List of Figures

1 Introduction

The functional definition of meaning claims that the content of a mental representation derives from its causal, informational or inferential relations to other mental representations (Pitt, 2018). This is consistent with the view of contemporary neuroscience and linguistics that concepts are stored by the brain in a semantic network, where the different types of connections to other concepts define their meaning (Gazzaniga et al., 2009). From this follows that the understanding of a concept presumes familiarity with the surrounding conceptual network, which is one of the key assumptions of the current paper.

Wikipedia is an online encyclopedia edited democratically by internet users - unlike traditional print encyclopedias that are created by an expert elite - which is both its strength due to the efficient accumulation of knowledge that is otherwise dispersed in the community, and its weakness considering the large variance in the contributing individuals' writing skills and expertise (Emigh and Herring, 2005). Wikipedia articles constitute a widely used resource of information, even though it is ill-advised to reference them in academic papers due to the lack of rigorous editorial and citation practices associated with their production. Nevertheless, in the current study I adopt the assumption that the structure of Wikipedia reflects the fundamental architecture of academic concepts, for the reason that it is one of the largest and most readily available open-source repositories of knowledge. For the present analysis, I utilize the hierarchical structure of Wikipedia articles as described by Aspert et al. (2019) to retrieve all the relevant data from theoretical academic disciplines and characterize the network of academic concepts, with the purpose to uncover the importance of *Philosophy* based on the definition of understanding I proposed in the previous paragraph. More specifically, I hypothesise that due to its highly abstract nature and historical prominance, more academic concepts rely on *Philosophy* either directly or indirectly - meaning that in order to gain a thorough understanding of these concepts, one needs to be familiar with some philosophical terms that constitute their conceptual foundation - than on any other field. In what follows, I attempt to confirm this proposal with the help of network scientific methods.

In the next section, I elaborate on the methods of data acquisition and the building of the hyperlink and cross-field networks. In Section 3., I characterize fields by their network attributes. Section 4. covers the cross-field contribution of the different fields to academic knowledge by applying measures defined by Rinia et al. (2002) to the current dataset, as well as highlighting one-to-one connections between individual fields with the help of the adjacency matrix. Furthermore, via the backbone extraction method put forward by Serrano et al. (2009), the section offers a more compact characterization of cross-field referencing tendencies. Finally in Section 5., I dig deeper into the structure of the network and reveal indirect paths of knowledge transmission between fields. The last section concludes.

Humanities and social science		Natural sciences	Formal sciences
Anthropology	Economics	Biology	Computer science
History	Geography	Chemistry	Logic
Linguistics	Interdisciplinary studies	Earth science	Mathematics
Philosophy	Political science	Physics	Applied mathematics
Religion	Psychology	Space science	Statistics
Arts	Sociology	Astronomy	Systems science

Table 1: Classification of academic fields. (Source: self-made, based on Wikipedia (ndb))

2 The Wikipedia hyperlink network

2.1 Data acquisition

The data for the current study has been acquired via Wikipedia's API service (Wikipedia, nda), with the help of the Wikipedia-API Python package (Majlis, 2020), according to the following procedure. Utilizing the hierarchical structure of article categorization in Wikipedia as described by Aspert et al. (2019), I identified broad categories of interest based on Wikipedia's *List of academic fields* (Wikipedia, ndb), which classifies fields on three different levels. I choose three of the four broad fields - *Humanities and social science*, *Natural sciences* and *Formal sciences* - and excluded *Professions and applied sciences* to keep the narrow theoretical focus of the analysis. I paraphrased category labels or omitted them altogether, if I could not find any articles associated with them via Wikipedia's api service (e.g. in case of *Sexual studies*). Next, to dig deeper in the network, I retrieved all first-level subcategories of the selected fields. In the final dataset I only included the broad field- and field-level classification as shown in Table 1., and merged all subfields and subcategories (e.g. *Culinary arts*, *Literature*, *Performing arts* and *Visual arts* are all labelled as *Arts*).

2.2 Building the network

Finally, I built a directed, unweighted network, the nodes of which are all the articles that belong to the selected category labels, and its edges are the hyperlinks connecting them. For technical considerations, I restricted the analysis to the largest strongly connected component of this network. The so obtained network contains 24 638 nodes and 655 873 edges, with an average degree of 26.62. As shown in Figure 1., both the in and out degree distribution of the network follow what could be best described as a "truncated power-law", which may partly be due to the data acquisition method, namely the restriction of categories to academic fields and the exclusion of any sub-categories beyond the first level.

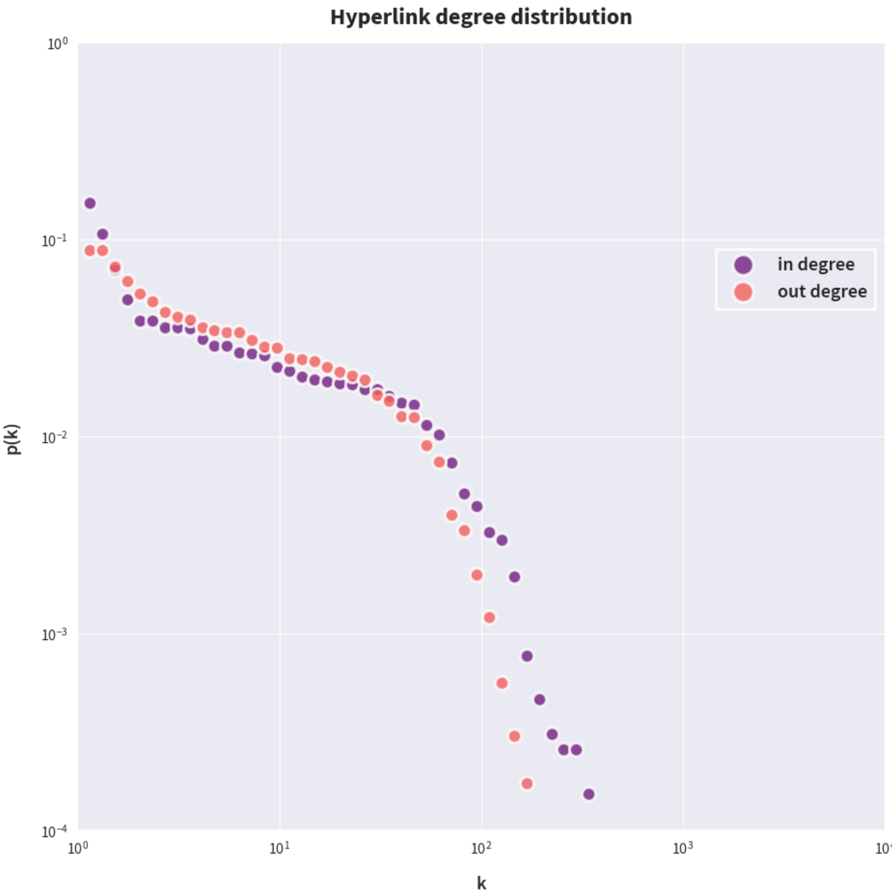

Figure 1: Distribution of in and out degrees the Wikipedia hyperlink network. Each point represents the probability of degree k, according to the logarithmic binning method introduced by Barabási (2016). (Source: self-made)

2.3 Cross-field projection

Articles are distributed between the various fields as shown in Figure 2. As category labels may overlap, I divided the number of articles that belong to multiple fields equally between these fields (an article that belongs to n different fields counts as $\frac{1}{n}$ in each field). As I was specifically interested in the transfer of knowledge and concepts between academic fields, I distinguished edges based on the field labels of their target and source nodes. First, I had to take into account the overlapping structure of category labels: there's no way to tell, which fields a particular hyperlink is associated with, if the two nodes it connects belong to multiple fields. As an example, let's suppose there's a link from Albert Einstein, who belongs to both *Physics* and *Philosophy*, to Kurt Gödel with field labels *Mathematics* and *Philosophy*. What can we say about this connection? Is this a within- or a cross-field edge? Is is between two philosophers, or a physicist and a mathematician, or a physicist and a philosopher? Resolving this problem requires a detailed exploration of context, which is unfeasible due to the large amount of data and it is beyond the scope of the current analysis. Therefore, I decided to give equal weights to all of these possibilities, by splitting the edges between all possible combinations. In the example above, this means that a link from Einstein to Gödel is counted as four distinct edges (from *Philosophy* to *Philosophy*, from *Philosophy* to *Mathematics*, from *Physics* to *Philosophy* and from *Physics* to *Mathematics*), each with a weight of $\frac{1}{4}$, as illustrated in Figure 3.. Based on this weighted network, I calculated the *within-* and *cross-field in* and *out degrees* for each node-field pair, which express how often an article refers to, or is referred to by another article from the same field, as opposed to articles from different fields. (If we consider the example above and suppose that the node *Albert Einstein* has no other outgoing edges, its *cross-field out degree* is $\frac{1}{4}$ in the field of *Philosophy*, and $\frac{1}{2}$ in the field of *Physics*, while its *within-field out degrees* in the two fields are $\frac{1}{4}$ and 0, respectively.)

Finally, I built the weighted cross-field projection of the hyperlink network, in which fields constitute the nodes, and the weight of the edge from field i to field j equals the share of hyperlinks pointing to articles in field j of all outgoing hyperlinks from articles in field i (thus, the out degree of each field equals 1). I derived the edge weights of this projection from the weighted version of the original network, to account for the potential ambiguities described in the previous paragraph.

3 Field-specific network phenomena

As the first step of the analysis, I investigated how the network properties of individual nodes differ across fields. For each field, I created a dummy variable: an $\{0;1\}^N$ array (where N is the total number of nodes), which indicates, whether a particular node belongs to the field or not. Next, I normalized these dummies, so that they sum up to 1 for each node. This was necessary in order to avoid underestimation of coefficients in case of overlapping fields. I fitted linear (OLS) regressions to the normalized field dummies to predict each of the following network attributes sepa-

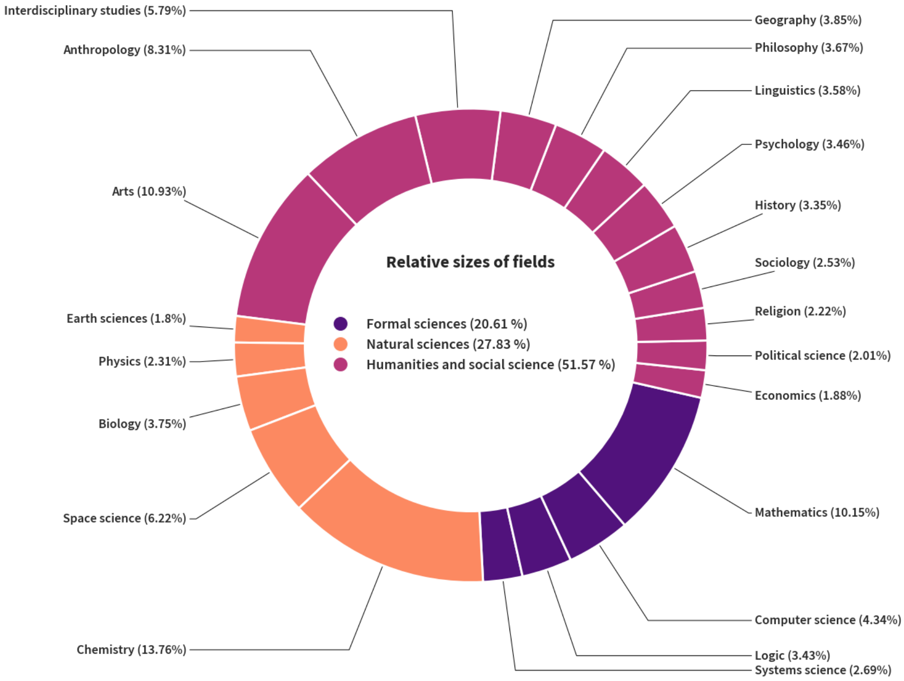

Figure 2: Relative sizes of fields. The weight of each article is reversely proportional to the number of fields it belongs to. (Source: self-made)

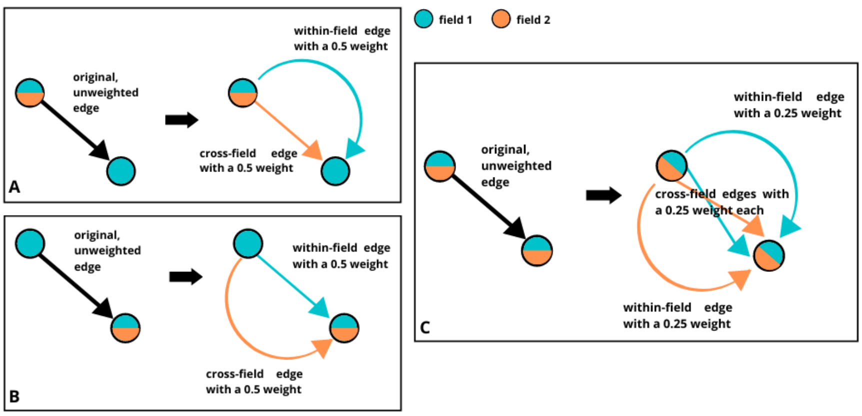

Figure 3: Illustration of edge-splitting between field combinations. (A) The outgoing edge of a multiple-field node pointing to a single-field node is divided into a cross- and a within-field outgoing edge, each with a weight of 0.5. (B) The incoming edge of a multiple-field node pointing from a single-field node is divided into a cross- and a within-field incoming edge, each with a weight of 0.5. (C) An edge between two multiple-field nodes is divided into two cross- and two within-fields edges, each with a weight of 0.25. (Source: self-made)

rately: the logarithm of in and out degree (*log in* and *log out degree*), incoming and outgoing neighbor connectivity[1] (*in* and *out neighbor conn*), clustering coefficient (*clustering*), betweenness centrality (*betweenness*), incoming and outgoing closeness centrality[2] (*closeness in* and *out*), and the proportions $\frac{cross-field\ in\ degree}{in\ degree}$ (*p cross field out*) and $\frac{cross-field\ out\ degree}{out\ degree}$ (*p cross field in*). All target variables were standardized for better comparison. Regression results are presented in Figure 4. (See Appendix A for regression summaries.) For clarification, a coefficient β_{kj} of the normalized field dummy j in regression k means that if a node belongs to field j, its attribute k increases on average by $\frac{\beta_{kj}}{n}$ standard deviations (where n is the total number of fields the node belongs to), ceteris paribus.

What I would like to highlight from these results is that *Philosophy* has the highest average degree, neighbor connectivity and closeness centrality for both the incoming and outgoing directions. Note that these attributes are not independent from each other (for more details, see the correlation matrix of target variables included in Appendix A). High degree and neighbor connectivity both suggest that an article reaches many other articles in the network in a few steps, which is also reflected in high closeness centrality. This implies that *Philosophy* is deeply embedded in the hyperlink network, even though it is a relatively small field based on the number of articles. Furthermore, its *p cross field in* coefficient is also in the top of the range, meaning that in case of *Philosophy*, the share of cross-field links of all incoming references is larger than the average. However, this latter statement applies to several other fields, such as *Sociology*, *Interdisciplinary studies* or *Physics.* On the other end of the range is *Chemistry* that - based on its low coefficients for both *p cross field in* and *out* - appears to be an extremely introverted field.

There are some other notable tendencies, of which I only mention a few. The high *p cross field out* coefficient of *Interdisciplinary studies* implies that articles in this field are the most likely to cite articles from other fields, which follows intuitively from the interdisciplinary character of the field. The outstandingly high clustering coefficient of *Geography*, as well as its low closeness measures indicate that geography articles are densely connected to each other, but not to the rest of the network, for which I could not find any straightforward explanation. In the regression for betweenness centrality, all field coefficients except for *Philosophy* and *History* are insignificant, meaning that the average betweenness centrality of these two fields is slightly higher than - while in the case of other fields it does not differ significantly from - the average.

The major takeaway from the results regarding the role of *Philosophy* within the network of academic concepts is the following: if we accept the premise that an outgoing edge signifies the import of knowledge from the target node, we can claim that a considerable proportion of academic terminology is conceptually rooted in philosophy, either directly or indirectly. In the following sections, I further examine this proposition from multiple angles.

[1] These metrics were calculated as the average in degree of predecessors, and the average out degree of successors, respectively.

[2] The incoming closeness centrality of node i is the reciprocal of the average shortest directed paths from all other nodes in the network to i. Analogously, I calculated the outgoing closeness centrality of i as the reciprocal of the average shortest directed paths from i to all other nodes in the network.

Figure 4: OLS regression coefficients by field and variable. Each columns represents a distinct model, where the target variable is the network attribute indicated by the column label, and rows encode the coefficients of exogenous variables. Insignificant parameters are masked. Rows are sorted lexicographically. (Source: self-made)

4 Cross-field contribution to academic knowledge

4.1 General ranking of academic fields

Rinia et al. (2002) studied knowledge transfer between fields of science via bibliometric methods. Their database contained all papers from the year 1999, published in journals included in the Science Citation Index, and the references between them. Disciplines were categorized based on the ISI classification of journals. Relying on previous studies, the authors developed three different metrics to measure the relative contribution of individual fields to collective scientific progress. According to their assumption, this contribution can be captured in the distribution of cross-field references, which reveals how much researchers in one field utilize results obtained in other fields. Although they acknowledge the weaknesses of this methodology, they conclude that it can provide useful indications about the value of research and potentially inform funding decisions.

They defined the *relative external use* of publications within field j as $\frac{(1-\alpha_j)}{\alpha_j} \cdot \frac{\sum_{i \neq j} R_{i,j}}{\sum_{i \neq j} R_i}$, where α_j is the relative size of j (calculated from the number of publications in each field), and the second fraction is the share of references to field j of all references by other fields (based on Rinia et al., 2002, pp. 351, 365). This measure thus favors small fields, as well as fields that received a large share of all references. As an alternative that takes into account the absolute number of articles and references, the *external citation average* is defined as $\frac{\sum_{i \neq j} R_{i,j}}{P_j}$, where the counter is the number of all references received by j, and the denominator is the number of articles in j (based on Rinia et al., 2002, pp. 351, 367). A third option is the *import/export ratio*: $\frac{\sum_{i \neq j} R_{i,j}}{\sum_{i \neq j} R_{j,i}}$, which expresses the number of references to j from other fields relative to the number of references from j to other fields (based on Rinia et al., 2002, pp. 351, 358). The authors found that based on the first two measures, *Multidisciplinary Sciences* rank the highest, followed by *Basic Life Sciences*, *Pharmacology* and *Environmental Sciences*. Four fields - *Multidisciplinary Sciences*, *Basic Life Sciences*, *Physics* and *Geo Sciences* - were associated with an import/export ratio higher than one, meaning that publications in external fields rely more heavily on them than they rely on external knowledge. (Rinia et al., 2002)

My approach differs from the one described above in several aspects: instead of publications, it uses Wikipedia articles as a measure of cross-disciplinary transfer of knowledge, and the domain of analysis is theoretical academic fields instead of theoretical and applied sciences. Relying on Wikipedia articles that store accumulated knowledge may overlook the short-term, gradual development of fields, which is captured in the separate publication of new discoveries. However, it provides a more general description of how the different terminologies are interrelated, irrespective of their temporal evolution. Figure 5. displays the rankings of fields I obtained based on the total number of external citations and the three measures introduced by Rinia et al., using the weighted cross-field network introduced in Section 2.3. As is apparent in the figure, in three of the four metrics *Philosophy* ranks highest, and it has the second largest import/export ratio, preceded only by *Chemistry*. This observation confirms my proposition about the importance of *Philosophy*. However, in order to draw further conclusions, I need to explore to what extent it

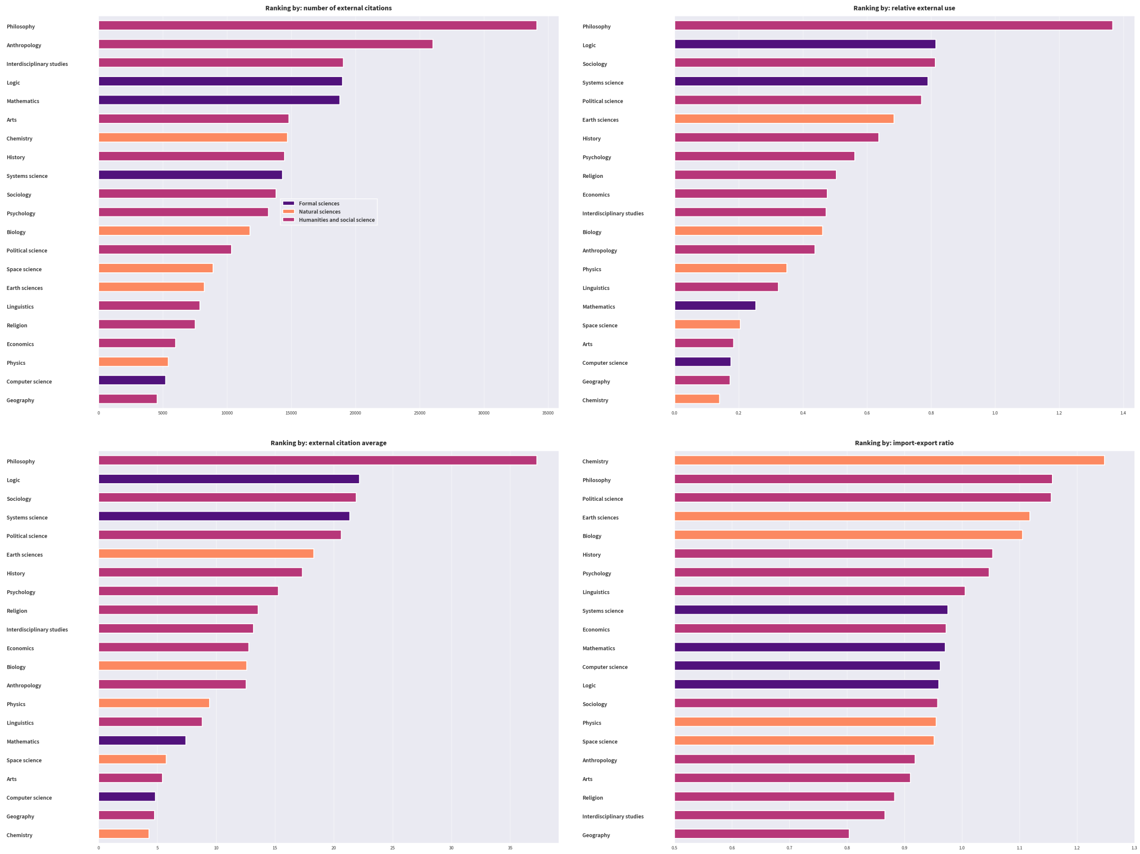

Figure 5: Ranking of academic fields based on Rinia et al.'s measures. Upper left: number of external citations. Upper right: relative external use. Lower left: external citation average. Lower right: import/export ratio. (Source: self-made)

contributes to the individual fields.

4.2 One-to-one contributions

To uncover one-to-one contributions, I derived the adjacency matrix of the cross-field network, which is shown in Figure 6. There are high values in the diagonal, as within-field references are the most frequent. By inspecting the rows we find that *Philosophy* stands out not only for having a large share of all external references, but also for receiving a considerable share of references from many different fields, including *Logic*, *Systems science*, *Arts*, *History*, *Psychology*, *Sociology*, *Political science*, *Economics*, *Linguistics* and *Religion*. In this comparison, *Philosophy* is followed by *Anthropology*, and the third place is shared by *Interdisciplinary studies*, *Mathematics* and *Logic*, all of which receive more than a 5% share of references from numerous fields. The matrix also reveals why *Chemistry* precedes *Philosophy* in the ranking by *import/export* ratio. *Chemistry* has an extremely high proportion of within-field links (there's 87% in the diagonal, while all other cells of its column are below the 5% threshold), which results in a low denominator

$\sum_{i \neq j} R_{j,i}$. On the other hand, only 43% of citations from *Philosophy* point within the field, while it receives many external references. Thus, in this case not a low denominator, but a high counter $\sum_{i \neq j} R_{i,j}$ is responsible for the high *import/export ratio*. In other words, while the advantage of *Chemistry* in this comparison derives from its low reliance on external knowledge - which is also in line with the regression results presented in Section 3. - , *Philosophy* stands out due to its high contribution to other fields.

However, it is evident from the matrix that *Philosophy* owes most of its incoming references to *Humanities and social sciences*. Thus, it is possible that its prominent role in the network of academic concepts is only due to the predominance of this broad field in the data, and the other two broad fields have their own theoretical foundation as well, which does not show in this context. Therefore, I split the original network into two parts: *Humanities and social science* and *Formal and natural sciences*, and built the weighted cross-field projections of these networks as well using the procedure described in Section 2.3. By deriving the adjacency matrices of these restricted cross-field networks, I found that, while none of the fields within *Formal and natural sciences* are as prominent as *Philosophy* within *Humanities and social science*, we can still say that *Systems science*, *Chemistry* and *Mathematics* are of central importance. (See Appendix B for further details.)

As it is difficult to extract the essence of all the information represented by the adjacency matrix, in the next section I employ an information reduction method that helps highlight the core structural properties of the cross-field network.

4.3 Backbone extraction

Serrano et al. (2009) propose a useful method to extract the fundamental structure of weighted networks that takes into account multi-scale variations and preserves local specificities. The authors describe the following procedure to be applied to each node of an undirected network in order to extract its backbone. First, the weights of the edges of node i are normalized, so that they sum up to 1. Were the edge weights distributed randomly between all k neighbors of the node, the normalized weights would be equivalent to randomly splitting the interval $(0, 1)$ into k sections, and taking the length of these sections. In this case, the probability density of length x is $\rho(x) = (k-1)(1-x)^{k-2}dx$ (Serrano et al., 2009, p. 6484). Consequently, the probability that the normalized weight of an edge is greater than or equal to some value w_{ij} - given that it is sampled from the distribution described above - is $\alpha_{ij} = 1-(k-1)\int_0^{w_{ij}} (1-x)^{k-2}dx$ (based on Serrano et al., 2009, p. 6484). An edge counts as statistically significant, if its α_{ij} is smaller than a previously determined threshold α. Note that $\alpha_{ij} \neq \alpha_{ji}$, as both values depend on the degrees of nodes i and j, respectively. Thus, an edge is added to the backbone if it passes the significance test at either of its endpoints.

This method is applicable to a directed network as well, in which case the same procedure is carried out separately for the successors and the predecessors of each node. Note, that in case of the cross-field network, the outgoing

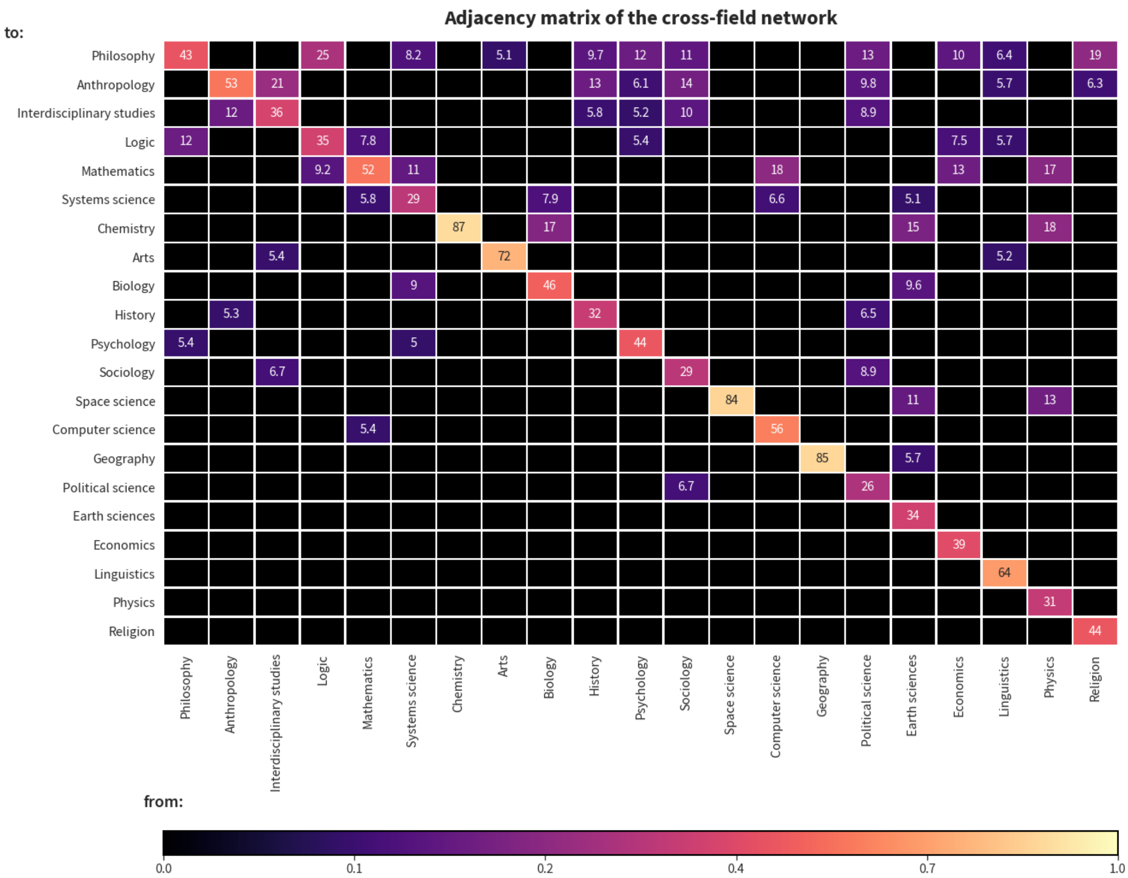

Figure 6: Adjacency matrix of the cross-field network. A value in row j and column i expresses the share (in %) of references received by field j of all references given by field i (values in a column sum up to 100 %). Only proportions higher than 5% are annotated, and rows are sorted based on the number of cells they contain that reach this threshold. (Source: self-made)

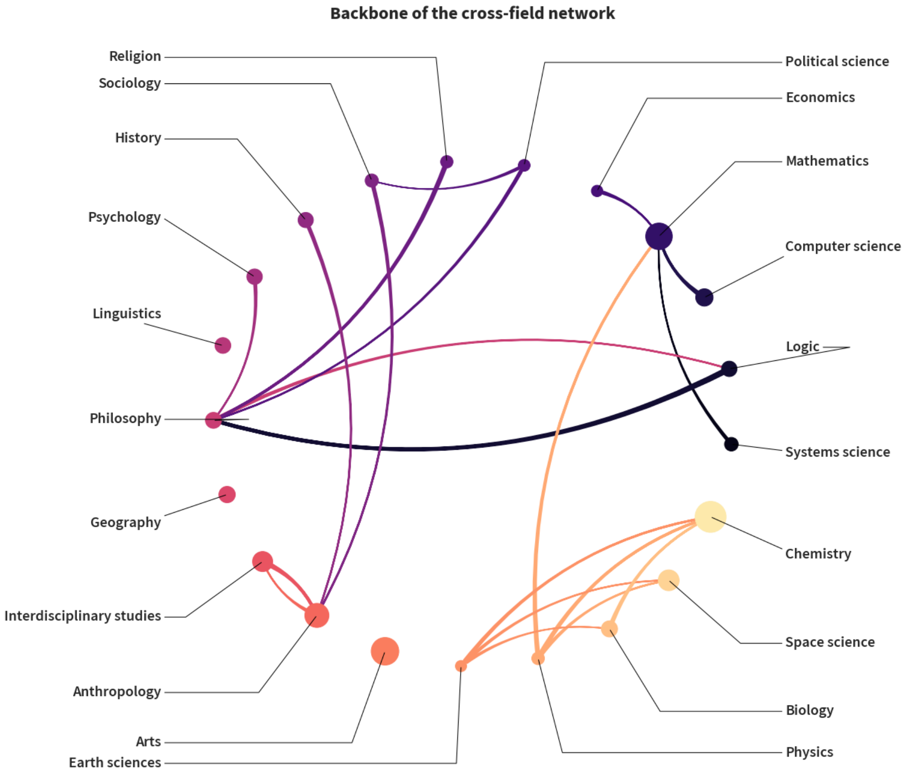

Figure 7: Backbone of the cross-field network. Only those edges are included, for which $\alpha_{ij} < 0.1$ for either the source or the target node, and self-loops (within-field edges) are not shown. Edge color corresponds to the color of the source, and width indicates edge weights. Node sizes represent the relative sizes of fields, as shown in Figure 2. (Source: self-made)

edge weights of each node are already normalized, and so the normalization in case of incoming edges is performed on the ratios instead of the absolute numbers of hyperlinks from other fields. This way, even small source fields with fewer references may pass the significance test at the target field, if they cite that field relatively often compared to other fields. By setting α to 0.1, I obtained the backbone of the cross-field network shown in Figure 7. Based on this figure, three fields in each broad field appear to be of central importance: *Philosophy* in *Humanities and social science*, *Mathematics* in *Formal sciences* and *Chemistry* in *Natural sciences*. It is also apparent that the three broad fields are relatively separated from each other. However, they are loosely connected via a strong reciprocal link between *Philosophy* and *Logic*, and a one-way link from *Physics* to *Mathematics*, which implies a deeper structural importance of these links. In the next section, I am taking the analysis further to uncover the indirect transfer of knowledge between fields.

5 Digging deeper: secondary and tertiary cross-field networks

In the previous section I have shown the detailed field-to-field connections within the network of academic concepts. However, so far I could only derive conclusions regarding direct knowledge transfer through individual hyperlinks. If we think of the academic knowledge represented by Wikipedia as an extended network of interrelated concepts, the in-depth understanding of one concept does not only require familiarity with other Wikipedia pages cited by the article, in which the concept is explained - with network terminology, these are the direct successors of the node. One also needs to rely to some extent on the knowledge accumulated in articles it references indirectly - the successors of the node's successors, and their successors, etc. Based on the backbone of the cross-field network presented in Figure 7., I hypothesised that while the direct importance of *Philosophy* is only apparent in case of *Humanities and social science*, other broad fields also import a considerable amount of knowledge from *Philosophy* indirectly. The most potential for this indirect knowledge transmission is through the links between *Philosophy* and *Logic*, and *Mathematics* and *Physics* that connect the three broad fields.

To measure the strength of indirect cross-field connections, I first built the secondary and tertiary projections of the original hyperlink network. In these networks, there's a weighted edge from node i to node j, if there's at least one path of length 2 from i to j, and the weight of the edge corresponds to the number of such paths. Then, I built the cross-field projections of both the secondary and tertiary hyperlink networks, and extracted their backbones according to the procedure described in Section 4.3. The results are presented in Figure 8. As I expected, secondary and tertiary edges pointing to *Philosophy* reveal its indirect importance for *Formal sciences*, and eventually even for *Natural sciences* through a tertiary link from *Biology.* It is noteworthy that there's a significant tertiary edge to *Philosophy* from *all fields* except for *Chemistry*, *Space sciences*, *Physics*, *Earth sciences* and *Geography.* On the other hand, the strength and nature of the connection between *Philosophy* and *Natural sciences* is rather counter-intuitive. Due to its strong reliance on *Mathematics*, I expected *Physics*, and not *Biology* to be the first to form a significant indirect connection to *Philosophy.* I also thought that *Natural sciences* will get more integrated into the backbone through indirect connections, but the figure indicates that it is still relatively separated from the rest of the network. Nevertheless, the overall message conveyed by figures 7. and 8. can be summarized as follows: the number and strength of significant cross-field edges pointing to *Philosophy* increases and extends to the other two broad fields as we dig deeper into the indirect hyperlink structure of Wikipedia articles. This is a unique tendency not shown by any other field with a similar consistency. What is demonstrated here, in fact, is how the remarkably high closeness centrality of philosophy articles pointed out in Section 3. translates to their indirect cross-field contribution to academic knowledge.

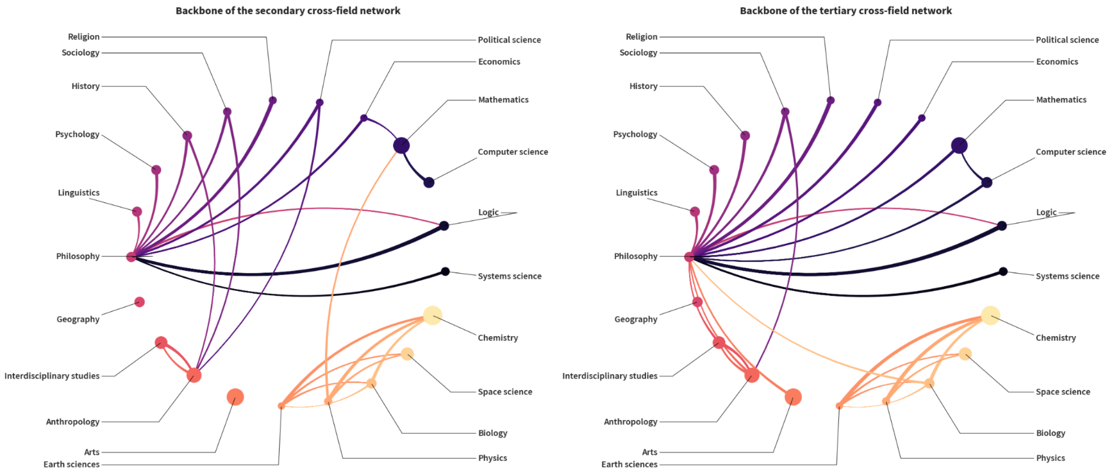

Figure 8: Backbones of secondary and tertiary cross-field networks. The interpretation is the same as in case of Figure 7. (Source: self-made)

6 Conclusion

In this paper, I explored the network of academic concepts, with a specific focus on the cross-field contribution of *Philosophy*. First, I fitted linear regressions to field membership dummy variables, which revealed the outstandingly high average degree, neighbor connectivity and closeness centrality of philosophy articles, suggesting that they are deeply embedded in Wikipedia's hyperlink structure. Next, based on the cross-field projection of the hyperlink network I calculated measures of cross-field contribution to knowledge defined by Rinia et al. (2002). I found that *Philosophy* by far outperforms all other fields in terms of relative external use and external citation average, meaning that the external academic relevance of the average article in *Philosophy* is higher than that of articles in any other field. While *Chemistry* precedes *Philosophy* in the ranking by *import/export* ratio, the adjacency matrix of the cross-field network that represents one-to-one connections between fields revealed that this is due to the introverted nature of *Chemistry*, while the primacy of *Philosophy* in all rankings is due to the high number of fields that refer to articles in *Philosophy* relatively often. A further inspection of the separate networks of *Humanities and social sciences* and *Formal and natural sciences* offered some reasons for caution regarding the conclusions drawn about the prominent role of *Philosophy*: when isolated properly from the large body of *Humanities and social sciences*, some fields in *Formal and natural sciences* also show their central importance within their own broad field. To extract the fundamental structure of the cross-field network, I employed the backbone extraction method introduced by Serrano et al. (2009), and found that the three broad fields are only loosely connected by statistically significant edges within the entire network of academic concepts. However, through the analysis of indirect connections via length 2 and 3 hyperlink paths, a deeper structure unfolded, indicating that all broad fields utilize the concepts and knowledge accumulated in the field of *Philosophy* - *Natural sciences* to lesser, while the other two broad fields to a great extent.

Finally, I would like to emphasize the relevance of this line of research and point out some open questions that may form the basis of further analysis. Studying the timeless network of academic concepts can provide a useful tool to sophisticate educational strategies. By taking into account the complex interrelated structure of academic knowledge, one can identify fields and subfields that are of key importance through the early phases of learning as well as in higher education. For instance, based on the results presented in this paper, teaching *Philosophy* to pupils before they encounter other disciplines might help them putting newly learnt concepts into a deeper context, and consequently enhance the success of public education. Of course, this is a far-fetched conclusion that requires thorough investigation. Therefore, the next step is to simulate different kinds of learning strategies that can be though of as sampling algorithms applied to the network of Wikipedia articles. If we accept the resulting samples as analogies of the knowledge network of an individual, we can define measures that reflect the quality of knowledge and evaluate the outcome of different sampling procedures based on these metrics. To follow up on the previous example, we could pose questions such as: does previously acquired knowledge in the field of *Philosophy* make learning the concepts of an arbitrary new discipline easier (e.g. more time-efficient), or does it encourage interdisciplinary approaches by embedding concepts in a more general context? A clever use of available data and the tools of network science may provide answers to these questions that help us improve on current educational practices and enrich the way we look at academic disciplines.

References

Aspert, N., Miz, V., Ricaud, B., and Vandergheynst, P. (2019). A graph-structured dataset for wikipedia research. *Companion Proceedings of The 2019 World Wide Web Conference*, pages 1188–1193. Downloaded: `https://dl.acm.org/doi/pdf/10.1145/3308560.3316757` (25 May, 2020).

Barabási, A.-L. (2016). Advanced Topic 3.B. Plotting Power-laws. In *Network Science*. [online edition]. Downloaded: `http://networksciencebook.com/chapter/4#advanced-b` (10 April, 2020).

Emigh, W. and Herring, S. (2005). Collaborative Authoring on the Web: A Genre Analysis of Online Encyclopedias. Proceedings of the 38th Annual Hawaii International Conference on System Sciences. Downloaded: `https://www.researchgate.net/publication/261398868_Collaborative_Authoring_on_the_Web_A_Genre_Analysis_of_Online_Encyclopedias` 11 April, 2021.

Gazzaniga, M. S., Ivry, R. B., and Mangun, G. R. (2009). *Cognitive Neuroscience. The biology of the mind*. W. W. Norton & Company, New York, London, third edition edition.

Majlis, M. (2020). Wikipedia-API 0.5.4. *MIT*. `https://pypi.org/project/Wikipedia-API/`.

Pitt, D. (2018). Mental representation. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2018 edition. Downloaded: `https://plato.stanford.edu/archives/win2018/entries/mental-representation/` (10 April, 2021).

Rinia, E., van Leeuwen, T., Bruins, E., van Vuren, H., and van Raan, A. (2002). Measuring knowledge transfer between fields of science. *Scientometrics*, 54(3):347–362. Downloaded: `https://link.springer.com/content/pdf/10.1023/A:1016078331752.pdf` (23 May, 2020).

Serrano, M. Á., Boguñá, M., and Vespignani, A. (2009). Extracting the multiscale backbone of complex weighted networks. *Proceedings of the National Academy of Sciences*, 106(16):6483–6488. Downloaded: `https://www.pnas.org/content/106/16/6483.full.pdf` (7 April 2021).

Wikipedia (n.d.a). API:Main page. `https://www.mediawiki.org/wiki/API:Main_page`.

Wikipedia (n.d.b). List of academic fields. Downloaded: `https://en.wikipedia.org/wiki/List_of_academic_fields` (25 May, 2020).

Appendix A Regression summaries

```
                            OLS Regression Results
==============================================================================
Dep. Variable:       in_neighbor_conn   R-squared:                       0.083
Model:                            OLS   Adj. R-squared:                  0.083
Method:                 Least Squares   F-statistic:                     106.3
Date:                Sun, 11 Apr 2021   Prob (F-statistic):               0.00
Time:                        20:08:17   Log-Likelihood:                -32197.
No. Observations:               23406   AIC:                         6.444e+04
Df Residuals:                   23385   BIC:                         6.460e+04
Df Model:                          20
Covariance Type:            nonrobust
=============================================================================================
                              coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------------
Anthropology               -0.0458      0.023     -2.034      0.042      -0.090      -0.002
Arts                       -0.2353      0.019    -12.287      0.000      -0.273      -0.198
Biology                    -0.0444      0.033     -1.363      0.173      -0.108       0.019
Chemistry                   0.0351      0.017      2.082      0.037       0.002       0.068
Computer science           -0.4551      0.031    -14.768      0.000      -0.516      -0.395
Earth sciences             -0.0734      0.048     -1.515      0.130      -0.168       0.022
Economics                   0.0504      0.046      1.095      0.273      -0.040       0.140
Geography                  -0.2306      0.032     -7.252      0.000      -0.293      -0.168
History                    -0.1057      0.035     -3.045      0.002      -0.174      -0.038
Interdisciplinary studies  -0.0289      0.028     -1.030      0.303      -0.084       0.026
Linguistics                -0.0473      0.034     -1.382      0.167      -0.114       0.020
Logic                       0.3374      0.035      9.658      0.000       0.269       0.406
Mathematics                -0.2920      0.020    -14.310      0.000      -0.332      -0.252
Philosophy                  1.1277      0.035     32.373      0.000       1.059       1.196
Physics                    -0.0746      0.042     -1.766      0.077      -0.157       0.008
Political science           0.1396      0.045      3.097      0.002       0.051       0.228
Psychology                  0.2469      0.034      7.180      0.000       0.179       0.314
Religion                    0.0348      0.042      0.825      0.410      -0.048       0.118
Sociology                   0.5549      0.041     13.633      0.000       0.475       0.635
Space science               0.1523      0.026      5.954      0.000       0.102       0.202
Systems science             0.1399      0.040      3.481      0.001       0.061       0.219
==============================================================================
Omnibus:                    16803.567   Durbin-Watson:                   2.004
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           569066.760
Skew:                           3.060   Prob(JB):                         0.00
Kurtosis:                      26.368   Cond. No.                         2.87
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure A.1: Regression summary for incoming neighbor connectivity

```
                            OLS Regression Results
==============================================================================
Dep. Variable:      out_neighbor_conn   R-squared:                       0.172
Model:                            OLS   Adj. R-squared:                  0.171
Method:                 Least Squares   F-statistic:                     242.7
Date:                Sun, 11 Apr 2021   Prob (F-statistic):               0.00
Time:                        20:08:17   Log-Likelihood:                -31032.
No. Observations:               23406   AIC:                         6.211e+04
Df Residuals:                   23385   BIC:                         6.228e+04
Df Model:                          20
Covariance Type:            nonrobust
============================================================================================
                                coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------------
Anthropology                 -0.0216      0.021     -1.007      0.314      -0.064       0.020
Arts                         -0.3182      0.018    -17.488      0.000      -0.354      -0.283
Biology                      -0.1904      0.031     -6.111      0.000      -0.251      -0.129
Chemistry                     0.0808      0.016      5.030      0.000       0.049       0.112
Computer science             -0.5722      0.029    -19.517      0.000      -0.630      -0.515
Earth sciences               -0.1769      0.046     -3.888      0.000      -0.266      -0.088
Economics                    -0.1234      0.044     -2.832      0.005      -0.209      -0.038
Geography                    -0.3475      0.030    -11.466      0.000      -0.407      -0.288
History                      -0.0799      0.033     -2.407      0.016      -0.145      -0.015
Interdisciplinary studies     0.1771      0.027      6.602      0.000       0.125       0.230
Linguistics                  -0.0303      0.033     -0.929      0.353      -0.094       0.034
Logic                         0.5855      0.033     17.647      0.000       0.520       0.650
Mathematics                  -0.4677      0.019    -24.108      0.000      -0.506      -0.430
Philosophy                    1.6249      0.033     49.145      0.000       1.560       1.690
Physics                      -0.1259      0.040     -3.167      0.002      -0.204      -0.048
Political science             0.0379      0.042      0.894      0.371      -0.045       0.121
Psychology                    0.1373      0.033      4.188      0.000       0.073       0.202
Religion                      0.2857      0.040      7.088      0.000       0.207       0.365
Sociology                     0.7769      0.039     19.799      0.000       0.700       0.854
Space science                 0.1283      0.024      5.287      0.000       0.081       0.176
Systems science               0.0639      0.038      1.671      0.095      -0.011       0.139
==============================================================================
Omnibus:                     4749.054   Durbin-Watson:                   2.002
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            13593.364
Skew:                           1.071   Prob(JB):                         0.00
Kurtosis:                       6.058   Cond. No.                         2.83
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure A.2: Regression summary for outgoing neighbor connectivity

```
                            OLS Regression Results
==============================================================================
Dep. Variable:             clustering   R-squared:                       0.042
Model:                            OLS   Adj. R-squared:                  0.041
Method:                 Least Squares   F-statistic:                     51.57
Date:                Sun, 11 Apr 2021   Prob (F-statistic):          3.48e-201
Time:                        20:08:18   Log-Likelihood:                -32732.
No. Observations:               23406   AIC:                         6.551e+04
Df Residuals:                   23385   BIC:                         6.567e+04
Df Model:                          20
Covariance Type:            nonrobust
=============================================================================================
                                coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------------
Anthropology                  0.1551      0.023      6.773      0.000       0.110       0.200
Arts                          0.0981      0.020      5.025      0.000       0.060       0.136
Biology                      -0.0650      0.034     -1.936      0.053      -0.131       0.001
Chemistry                    -0.1361      0.017     -7.864      0.000      -0.170      -0.102
Computer science             -0.0510      0.031     -1.631      0.103      -0.112       0.010
Earth sciences               -0.2230      0.050     -4.490      0.000      -0.320      -0.126
Economics                     0.1760      0.047      3.729      0.000       0.083       0.268
Geography                     0.7053      0.033     21.619      0.000       0.641       0.769
History                      -0.2593      0.036     -7.230      0.000      -0.330      -0.189
Interdisciplinary studies    -0.1213      0.029     -4.222      0.000      -0.178      -0.065
Linguistics                   0.2217      0.035      6.273      0.000       0.152       0.291
Logic                        -0.0095      0.036     -0.267      0.789      -0.079       0.060
Mathematics                  -0.1784      0.021     -8.572      0.000      -0.219      -0.138
Philosophy                    0.0338      0.036      0.953      0.341      -0.036       0.103
Physics                      -0.2311      0.043     -5.378      0.000      -0.315      -0.147
Political science            -0.1526      0.046     -3.318      0.001      -0.243      -0.062
Psychology                   -0.1222      0.035     -3.469      0.001      -0.191      -0.053
Religion                     -0.0959      0.043     -2.212      0.027      -0.181      -0.011
Sociology                    -0.1388      0.042     -3.322      0.001      -0.221      -0.057
Space science                 0.3039      0.026     11.602      0.000       0.253       0.355
Systems science              -0.0799      0.041     -1.931      0.054      -0.161       0.001
==============================================================================
Omnibus:                     1343.868   Durbin-Watson:                   2.032
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1436.093
Skew:                           0.576   Prob(JB):                         0.00
Kurtosis:                       2.618   Cond. No.                         2.87
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure A.3: Regression summary for clustering coefficient

OLS Regression Results

Dep. Variable:	betweenness	R-squared:	0.002
Model:	OLS	Adj. R-squared:	0.001
Method:	Least Squares	F-statistic:	2.678
Date:	Sun, 11 Apr 2021	Prob (F-statistic):	6.81e-05
Time:	20:08:18	Log-Likelihood:	-33542.
No. Observations:	23406	AIC:	6.713e+04
Df Residuals:	23385	BIC:	6.730e+04
Df Model:	20		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Anthropology	0.0232	0.024	0.975	0.330	-0.023	0.070
Arts	-0.0352	0.020	-1.739	0.082	-0.075	0.004
Biology	0.0281	0.035	0.814	0.415	-0.040	0.096
Chemistry	-0.0125	0.018	-0.700	0.484	-0.047	0.023
Computer science	-0.0353	0.032	-1.086	0.277	-0.099	0.028
Earth sciences	0.0916	0.051	1.812	0.070	-0.007	0.191
Economics	-0.0402	0.049	-0.823	0.411	-0.136	0.056
Geography	-0.0232	0.034	-0.682	0.495	-0.090	0.043
History	0.0968	0.037	2.611	0.009	0.024	0.169
Interdisciplinary studies	-0.0572	0.030	-1.931	0.053	-0.115	0.001
Linguistics	-0.0708	0.036	-1.950	0.051	-0.142	0.000
Logic	0.0067	0.037	0.180	0.857	-0.066	0.079
Mathematics	0.0068	0.022	0.315	0.753	-0.036	0.049
Philosophy	0.1841	0.037	5.003	0.000	0.112	0.256
Physics	-0.0153	0.044	-0.345	0.730	-0.102	0.071
Political science	-0.0430	0.047	-0.908	0.364	-0.136	0.050
Psychology	-0.0031	0.037	-0.084	0.933	-0.075	0.069
Religion	-0.0205	0.045	-0.455	0.649	-0.109	0.068
Sociology	-0.0093	0.044	-0.212	0.832	-0.095	0.076
Space science	-0.0066	0.027	-0.243	0.808	-0.060	0.047
Systems science	0.0700	0.042	1.652	0.099	-0.013	0.153

Omnibus:	54934.463	Durbin-Watson:	2.014
Prob(Omnibus):	0.000	Jarque-Bera (JB):	663185803.382
Skew:	23.449	Prob(JB):	0.00
Kurtosis:	826.296	Cond. No.	2.83

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Figure A.4: Regression summary for betweenness centrality

OLS Regression Results

Dep. Variable:	closeness_in	R-squared:	0.086
Model:	OLS	Adj. R-squared:	0.085
Method:	Least Squares	F-statistic:	110.4
Date:	Sun, 11 Apr 2021	Prob (F-statistic):	0.00
Time:	20:08:18	Log-Likelihood:	-32158.
No. Observations:	23406	AIC:	6.436e+04
Df Residuals:	23385	BIC:	6.453e+04
Df Model:	20		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Anthropology	-0.0235	0.022	-1.049	0.294	-0.067	0.020
Arts	-0.2453	0.019	-12.905	0.000	-0.283	-0.208
Biology	0.1036	0.033	3.151	0.002	0.039	0.168
Chemistry	-0.0230	0.017	-1.367	0.172	-0.056	0.010
Computer science	-0.4958	0.031	-16.150	0.000	-0.556	-0.436
Earth sciences	0.2481	0.048	5.179	0.000	0.154	0.342
Economics	0.0586	0.046	1.277	0.202	-0.031	0.149
Geography	-0.5758	0.032	-18.108	0.000	-0.638	-0.513
History	0.1250	0.035	3.592	0.000	0.057	0.193
Interdisciplinary studies	-0.0926	0.028	-3.286	0.001	-0.148	-0.037
Linguistics	-0.1733	0.034	-5.035	0.000	-0.241	-0.106
Logic	0.4350	0.035	12.453	0.000	0.367	0.503
Mathematics	-0.1034	0.020	-5.085	0.000	-0.143	-0.064
Philosophy	1.0109	0.035	29.226	0.000	0.943	1.079
Physics	-0.1988	0.042	-4.740	0.000	-0.281	-0.117
Political science	0.2576	0.045	5.742	0.000	0.170	0.346
Psychology	0.0928	0.034	2.711	0.007	0.026	0.160
Religion	0.1105	0.042	2.606	0.009	0.027	0.194
Sociology	0.4653	0.041	11.404	0.000	0.385	0.545
Space science	0.0510	0.026	1.994	0.046	0.001	0.101
Systems science	0.3968	0.041	9.774	0.000	0.317	0.476

Omnibus:	97.547	Durbin-Watson:	2.003
Prob(Omnibus):	0.000	Jarque-Bera (JB):	98.680
Skew:	0.157	Prob(JB):	3.73e-22
Kurtosis:	3.046	Cond. No.	2.84

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Figure A.5: Regression summary for incoming closeness centrality

OLS Regression Results

Dep. Variable:	closeness_out	R-squared:	0.118
Model:	OLS	Adj. R-squared:	0.117
Method:	Least Squares	F-statistic:	156.1
Date:	Sun, 11 Apr 2021	Prob (F-statistic):	0.00
Time:	20:08:18	Log-Likelihood:	-31749.
No. Observations:	23406	AIC:	6.354e+04
Df Residuals:	23385	BIC:	6.371e+04
Df Model:	20		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Anthropology	0.1061	0.022	4.798	0.000	0.063	0.149
Arts	-0.2592	0.019	-13.879	0.000	-0.296	-0.223
Biology	-0.0359	0.032	-1.117	0.264	-0.099	0.027
Chemistry	-0.2807	0.017	-16.976	0.000	-0.313	-0.248
Computer science	-0.5450	0.030	-18.117	0.000	-0.604	-0.486
Earth sciences	0.1834	0.047	3.933	0.000	0.092	0.275
Economics	0.0981	0.045	2.181	0.029	0.010	0.186
Geography	-0.5772	0.031	-18.366	0.000	-0.639	-0.516
History	0.2176	0.034	6.376	0.000	0.151	0.284
Interdisciplinary studies	0.1060	0.028	3.825	0.000	0.052	0.160
Linguistics	-0.1285	0.034	-3.789	0.000	-0.195	-0.062
Logic	0.4462	0.034	12.967	0.000	0.379	0.514
Mathematics	-0.1456	0.020	-7.305	0.000	-0.185	-0.107
Philosophy	1.1432	0.034	33.356	0.000	1.076	1.210
Physics	0.0747	0.041	1.820	0.069	-0.006	0.155
Political science	0.1693	0.044	3.843	0.000	0.083	0.256
Psychology	0.2881	0.034	8.563	0.000	0.222	0.354
Religion	0.1717	0.042	4.124	0.000	0.090	0.253
Sociology	0.6067	0.040	15.136	0.000	0.528	0.685
Space science	0.0591	0.025	2.340	0.019	0.010	0.109
Systems science	0.4612	0.040	11.495	0.000	0.383	0.540

Omnibus:	215.877	Durbin-Watson:	1.971
Prob(Omnibus):	0.000	Jarque-Bera (JB):	353.643
Skew:	-0.031	Prob(JB):	1.61e-77
Kurtosis:	3.599	Cond. No.	2.82

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Figure A.6: Regression summary for outgoing closeness centrality

```
                            OLS Regression Results
==============================================================================
Dep. Variable:       p_cross_field_in   R-squared:                       0.144
Model:                            OLS   Adj. R-squared:                  0.143
Method:                 Least Squares   F-statistic:                     196.3
Date:                Sun, 11 Apr 2021   Prob (F-statistic):               0.00
Time:                        20:08:18   Log-Likelihood:                -31458.
No. Observations:               23406   AIC:                         6.296e+04
Df Residuals:                   23385   BIC:                         6.313e+04
Df Model:                          20
Covariance Type:            nonrobust
=============================================================================================
                               coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------------
Anthropology                -0.0459      0.022     -2.105      0.035      -0.089      -0.003
Arts                        -0.3222      0.019    -17.348      0.000      -0.359      -0.286
Biology                      0.1033      0.032      3.267      0.001       0.041       0.165
Chemistry                   -0.5285      0.016    -32.176      0.000      -0.561      -0.496
Computer science            -0.0368      0.030     -1.236      0.216      -0.095       0.022
Earth sciences               0.4161      0.046      8.949      0.000       0.325       0.507
Economics                    0.2756      0.045      6.179      0.000       0.188       0.363
Geography                   -0.3539      0.031    -11.400      0.000      -0.415      -0.293
History                      0.2625      0.034      7.761      0.000       0.196       0.329
Interdisciplinary studies    0.6723      0.027     24.628      0.000       0.619       0.726
Linguistics                  0.0828      0.033      2.487      0.013       0.018       0.148
Logic                        0.3300      0.034      9.784      0.000       0.264       0.396
Mathematics                 -0.1122      0.020     -5.705      0.000      -0.151      -0.074
Philosophy                   0.5985      0.033     17.913      0.000       0.533       0.664
Physics                      0.6310      0.041     15.503      0.000       0.551       0.711
Political science            0.5913      0.043     13.693      0.000       0.507       0.676
Psychology                   0.1287      0.033      3.868      0.000       0.063       0.194
Religion                     0.0921      0.041      2.247      0.025       0.012       0.172
Sociology                    0.6800      0.039     17.267      0.000       0.603       0.757
Space science               -0.3929      0.025    -15.848      0.000      -0.442      -0.344
Systems science              0.5170      0.039     13.253      0.000       0.441       0.593
==============================================================================
Omnibus:                    13930.503   Durbin-Watson:                   2.003
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           488036.054
Skew:                           2.292   Prob(JB):                         0.00
Kurtosis:                      24.895   Cond. No.                         2.83
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure A.7: Regression summary for incoming cross field degree ratio

```
                            OLS Regression Results
==============================================================================
Dep. Variable:      p_cross_field_out   R-squared:                       0.168
Model:                            OLS   Adj. R-squared:                  0.167
Method:                 Least Squares   F-statistic:                     235.4
Date:                Sun, 11 Apr 2021   Prob (F-statistic):               0.00
Time:                        20:08:19   Log-Likelihood:                -30982.
No. Observations:               23406   AIC:                         6.201e+04
Df Residuals:                   23385   BIC:                         6.218e+04
Df Model:                          20
Covariance Type:            nonrobust
==============================================================================================
                                coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------------
Anthropology                  0.0380      0.021      1.771      0.077      -0.004       0.080
Arts                         -0.2649      0.018    -14.602      0.000      -0.300      -0.229
Biology                       0.0497      0.031      1.593      0.111      -0.011       0.111
Chemistry                    -0.6506      0.016    -40.568      0.000      -0.682      -0.619
Computer science             -0.1432      0.029     -4.925      0.000      -0.200      -0.086
Earth sciences                0.4751      0.045     10.499      0.000       0.386       0.564
Economics                     0.1130      0.044      2.597      0.009       0.028       0.198
Geography                    -0.3437      0.030    -11.354      0.000      -0.403      -0.284
History                       0.3809      0.033     11.495      0.000       0.316       0.446
Interdisciplinary studies     0.9189      0.027     34.360      0.000       0.867       0.971
Linguistics                   0.0528      0.033      1.619      0.105      -0.011       0.117
Logic                         0.3203      0.033      9.653      0.000       0.255       0.385
Mathematics               -3.511e-05      0.019     -0.002      0.999      -0.038       0.038
Philosophy                    0.4474      0.033     13.674      0.000       0.383       0.512
Physics                       0.3004      0.040      7.439      0.000       0.221       0.380
Political science             0.4995      0.043     11.637      0.000       0.415       0.584
Psychology                    0.1243      0.032      3.854      0.000       0.061       0.188
Religion                      0.1268      0.040      3.145      0.002       0.048       0.206
Sociology                     0.5378      0.039     13.859      0.000       0.462       0.614
Space science                -0.3944      0.024    -16.216      0.000      -0.442      -0.347
Systems science               0.6089      0.038     15.977      0.000       0.534       0.684
==============================================================================
Omnibus:                    19025.197   Durbin-Watson:                   2.027
Prob(Omnibus):                  0.000   Jarque-Bera (JB):          1721764.118
Skew:                           3.343   Prob(JB):                         0.00
Kurtosis:                      44.482   Cond. No.                         2.82
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Figure A.8: Regression summary for outgoing cross field degree ratio

OLS Regression Results

Dep. Variable:	log_in_degree	R-squared:	0.039
Model:	OLS	Adj. R-squared:	0.038
Method:	Least Squares	F-statistic:	47.25
Date:	Sun, 11 Apr 2021	Prob (F-statistic):	1.64e-183
Time:	20:08:19	Log-Likelihood:	-32755.
No. Observations:	23406	AIC:	6.555e+04
Df Residuals:	23385	BIC:	6.572e+04
Df Model:	20		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Anthropology	0.2246	0.023	9.777	0.000	0.180	0.270
Arts	-0.0319	0.019	-1.639	0.101	-0.070	0.006
Biology	-0.0446	0.034	-1.327	0.185	-0.111	0.021
Chemistry	-0.0250	0.017	-1.445	0.148	-0.059	0.009
Computer science	-0.3793	0.031	-12.072	0.000	-0.441	-0.318
Earth sciences	0.0541	0.049	1.101	0.271	-0.042	0.150
Economics	-0.1526	0.047	-3.243	0.001	-0.245	-0.060
Geography	-0.0539	0.033	-1.648	0.099	-0.118	0.010
History	-0.0114	0.036	-0.319	0.750	-0.082	0.059
Interdisciplinary studies	-0.1427	0.029	-4.929	0.000	-0.199	-0.086
Linguistics	-0.0198	0.035	-0.560	0.576	-0.089	0.049
Logic	0.1954	0.036	5.474	0.000	0.125	0.265
Mathematics	-0.2176	0.021	-10.417	0.000	-0.259	-0.177
Philosophy	0.5675	0.036	15.892	0.000	0.498	0.638
Physics	-0.3673	0.043	-8.482	0.000	-0.452	-0.282
Political science	0.0439	0.046	0.957	0.339	-0.046	0.134
Psychology	-0.1702	0.035	-4.846	0.000	-0.239	-0.101
Religion	0.0312	0.043	0.719	0.472	-0.054	0.116
Sociology	0.0954	0.042	2.270	0.023	0.013	0.178
Space science	0.3185	0.026	12.173	0.000	0.267	0.370
Systems science	0.1475	0.041	3.593	0.000	0.067	0.228

Omnibus:	1968.253	Durbin-Watson:	2.013
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1270.585
Skew:	0.450	Prob(JB):	1.25e-276
Kurtosis:	2.298	Cond. No.	2.84

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Figure A.9: Regression summary for the logarithm of in degree

OLS Regression Results

Dep. Variable:	log_out_degree	R-squared:	0.058
Model:	OLS	Adj. R-squared:	0.057
Method:	Least Squares	F-statistic:	71.48
Date:	Sun, 11 Apr 2021	Prob (F-statistic):	3.80e-282
Time:	20:08:19	Log-Likelihood:	-32512.
No. Observations:	23406	AIC:	6.507e+04
Df Residuals:	23385	BIC:	6.524e+04
Df Model:	20		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Anthropology	0.1692	0.023	7.408	0.000	0.124	0.214
Arts	-0.1334	0.019	-6.890	0.000	-0.171	-0.095
Biology	-0.0887	0.033	-2.681	0.007	-0.154	-0.024
Chemistry	0.0180	0.017	1.054	0.292	-0.015	0.051
Computer science	-0.4988	0.031	-16.076	0.000	-0.560	-0.438
Earth sciences	0.0076	0.048	0.157	0.875	-0.087	0.103
Economics	-0.1954	0.046	-4.220	0.000	-0.286	-0.105
Geography	0.4158	0.033	12.751	0.000	0.352	0.480
History	-0.0631	0.035	-1.785	0.074	-0.132	0.006
Interdisciplinary studies	-0.1499	0.029	-5.239	0.000	-0.206	-0.094
Linguistics	-0.0490	0.035	-1.408	0.159	-0.117	0.019
Logic	0.2082	0.035	5.868	0.000	0.139	0.278
Mathematics	-0.2741	0.021	-13.311	0.000	-0.314	-0.234
Philosophy	0.6605	0.035	18.733	0.000	0.591	0.730
Physics	-0.3143	0.043	-7.342	0.000	-0.398	-0.230
Political science	-0.0325	0.045	-0.721	0.471	-0.121	0.056
Psychology	-0.0813	0.035	-2.327	0.020	-0.150	-0.013
Religion	-0.0275	0.043	-0.638	0.524	-0.112	0.057
Sociology	0.0991	0.041	2.394	0.017	0.018	0.180
Space science	0.3694	0.026	14.150	0.000	0.318	0.421
Systems science	0.1376	0.041	3.381	0.001	0.058	0.217

Omnibus:	1110.055	Durbin-Watson:	2.006
Prob(Omnibus):	0.000	Jarque-Bera (JB):	529.567
Skew:	0.167	Prob(JB):	1.01e-115
Kurtosis:	2.343	Cond. No.	2.84

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Figure A.10: Regression summary for the logarithm of out degree

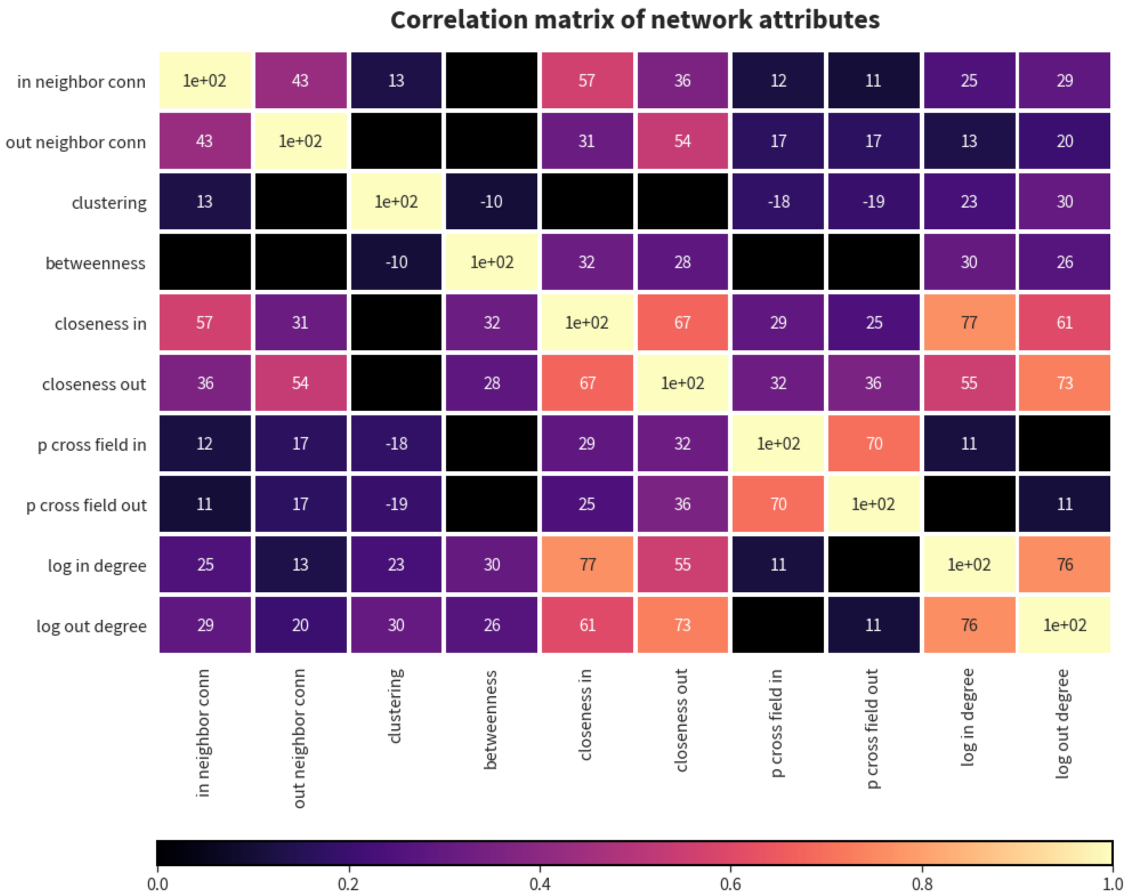

Figure A.11: Correlation matrix of target variables. (Source: self-made)

Appendix B Broad field networks

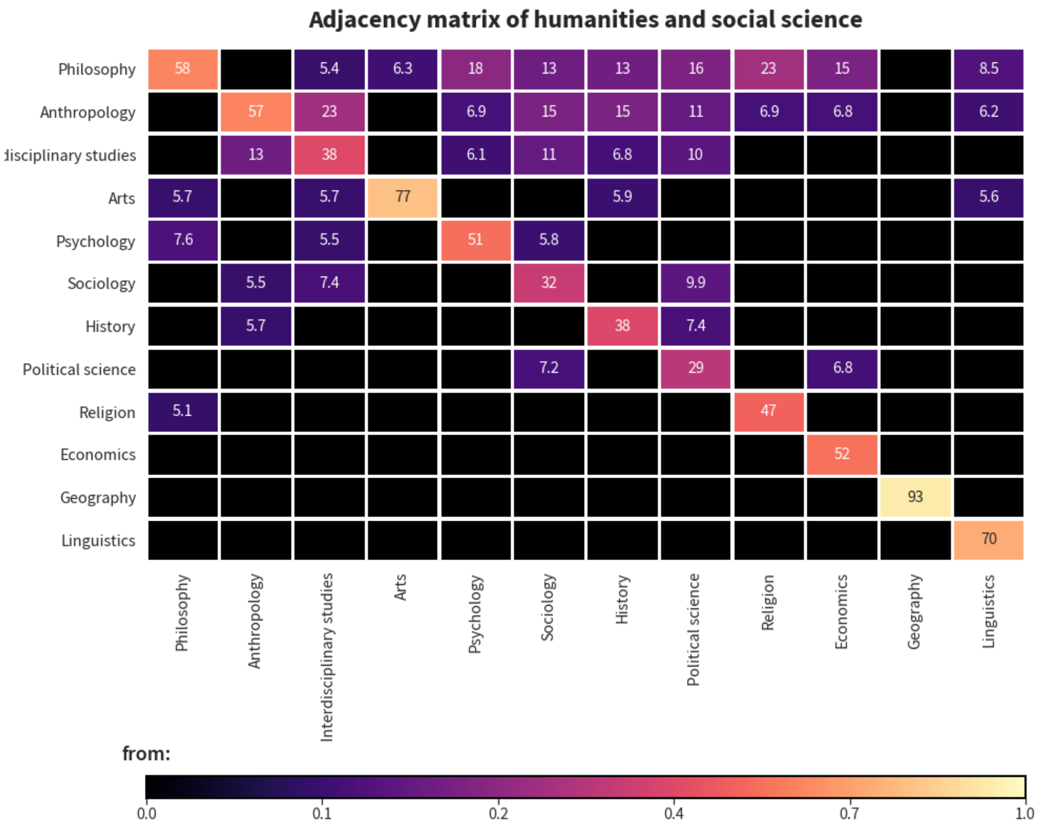

Figure B.1: Adjacency matrix of humanities and social science. (Source: self-made)

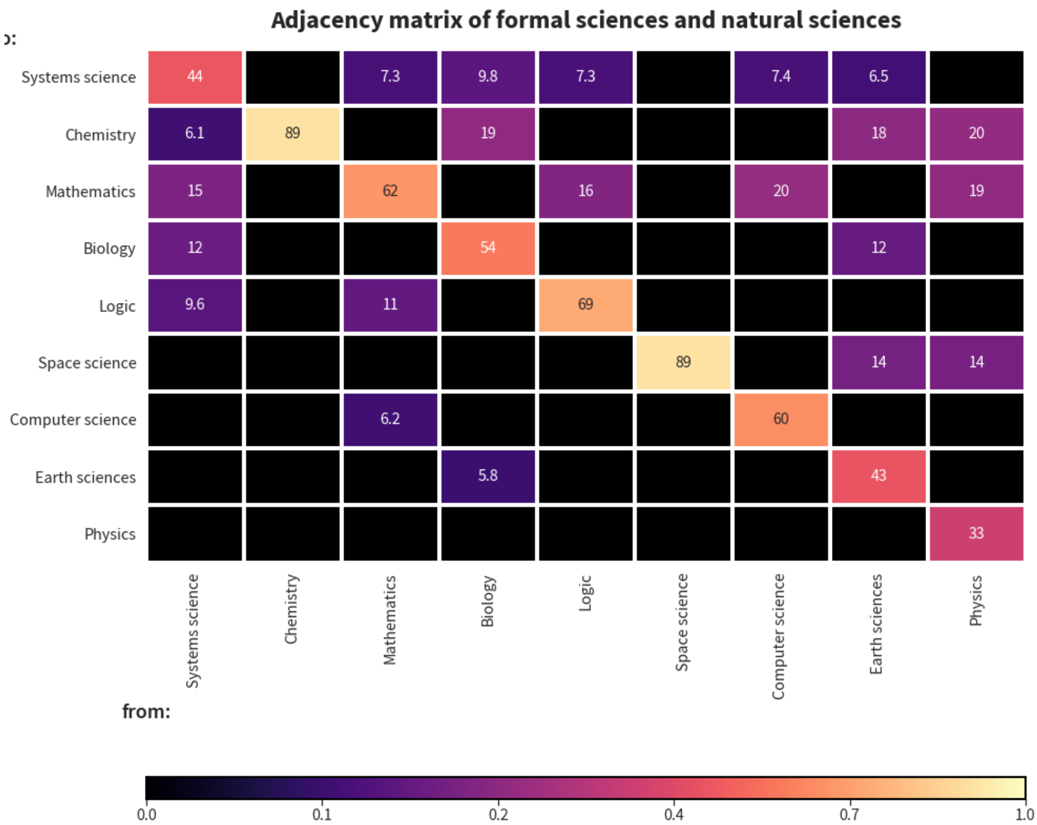

Figure B.2: Adjacency matrix of formal and natural sciences. (Source: self-made)