

Topics, Communities, and Diversity. Overlaying Cognitive and Social Dimensions in Argumentation Studies

Natalija Todorovic* and Benedetto Lepori**

* natalija.todorovic@usi.ch

Institute of Communication and Public Policy, Università della Svizzera Italiana, Lugano, Switzerland

** blepori@usi.ch

Institute of Communication and Public Policy, Università della Svizzera Italiana, Lugano, Switzerland

Abstract

In this work we analyse the social structure of the broad field of argumentation and its relationship with the topical structure to discover the dynamics of connections that exist across different communities in the field. In previous work, we demonstrated the topical richness and versatility of the community formed around the term ‘argumentation’ that enabled integration of multiple theoretical frameworks and contexts of usage.

Now, we analyse the scientific production of intellectual leaders, seeking to unveil the topical diversity of their work in the field. Furthermore, we perform social network analysis of co-authorship in the field, to be able to compare it with the cognitive structure we obtained in previous work. In addition, we compare the profiles of authors in different types of communities. The analysis is performed on database containing about 10,000 publications indexed by Scopus with the word ‘argumentation’ in title, abstract, and keywords. The combination of scientometric techniques and social network analysis allows overlaying cognitive and social structure of communities and highlights the similarities and differences between the two. When combined with analysis of individual scientific profiles of the most productive and influential individuals, the work provides insights of individual strategies scientists use when choosing their topics of research and research problems.

Our findings show that the structure of the field is rich with respect of diversity of topics and communities of authors. In addition, there are numerous connections among these that enable the diffusion of ideas across topics as well as several specific communities whose engagement contributes to delving deep into single topics. Individuals as drivers of these two mechanisms tend to show more diversity in their works as their production increases. These normal ecologies of science have been studied before; however, we explore it from new viewpoint that is suitable for fields not having standardized bibliographic databases with annotated articles. These open some interesting questions for future exploration.

1. Introduction

Scientific communities are built on cognitive and socio-cultural dimension and to truly understand the scientific production we should develop accounts for capturing both dimensions (Nersessian, 2005). Empirical analyses of the dimensions have been vastly covered by literature albeit rarely together, as scholars tend to both develop tools and examine emergent trends and scholarly communities separately (Yan, Ding, Milojević, & Sugimoto, 2012).

It is widely accepted that the appropriate approach to study cognitive dimension of science is through terminology-based studies to better understand ideas, knowledge, and relationships between them (Milojević, Sugimoto, Yan, & Ding, 2011). Traditionally, the cognitive dimension has been analysed using co-word analysis, e.g., Callon, Courtial, Turner, & Bauin (1983) to identify clusters from documents. In the era of internet and increasing computing power, topic modelling becomes standard technique for identification and tracking of knowledge domains in science (Foster & Evans, 2011). An example of topic modelling

technique is Latent Dirichlet Allocation which determines hidden topics in corpus by using probabilistic distributions of words over documents and documents over topics (Blei, Ng, & Jordan, 2003). Besides topic modelling, the cognitive mapping of the field can be performed using other methods, e.g. the work of Foster, Rzhetsky, & Evans (2015) who explored the knowledge categories and relationships among them by using annotations of chemical compounds to identify knowledge clusters as “subfields” in their dataset.

Literature suggests that social interaction influences emergence of topics, e.g., Zhou, Ji, Zha, & Giles, (2006), Gruhl, Guha, Liben-Nowell, & Tomkins (2004), Backstrom, Huttenlocher, Kleinberg, & Lan (2006). Thus, studying topics and social communities is meaningful and necessary for improved understanding of research fields and science in general. This sort of hybrid approach to studying social and cognitive structure (Yan et al., 2012) can reveal different types of communities and their members. By looking at the differences and similarities between them, we can paint the better picture of the field itself. An example of research that combines structural and semantic features is analysis of academic team formation by Taramasco, Cointet, & Roth (2010). Some other related works are summarized by Cambrosio, Cointet, & Abdo (2020).

According to Yan et al. (2012), research topics and research communities are not disconnected from each other, and it is important to study them together to be able to tell in what topic a community is specialised and how communities are related via topics. Yan et al. (2012) argue that analysis that includes topic detection and community identification can contribute to our understanding of interdisciplinarity and scholarly communication. In addition, such an analysis can contribute to understand the relationships between researchers and topics, how topics interact one with another and how communities trade topic and researchers (Osborne, Scavo, & Motta, 2014).

Social network analysis is a popular strategy for investigating social structures (Otte & Rousseau, 2002). A branch of social network analysis, community detection is an approach that aims at identifying cohesive groups in real-world graphs (Zhao, Li, Zhang, Chiclana, & Viedma, 2019). By performing such an analysis, we gain better understanding of mechanisms of exchange in social network itself. In science, there are several approaches to creation of social networks. Some studies that focus on collaboration, e.g., Newman (2004) or citation patterns out of which groups can emerge (Waltman & van Eck, 2012).

Although social and cognitive dimension are closely related and interconnected, we can expect that there is no complete overlap between them. As Yan, Ding, & Jacob (2012) demonstrate, after analysing paper-author and paper-word matrices of articles published by selection of authors in 16 journals in library and information science, the topic of study only partially drives the social structure of a community.

One can thus expect that there are communities that cover single topics, as well as communities that work on multiple topics. Engaging in multiple topics, i.e., diversity in research, is a phenomenon that is found to be correlated to impact of research and an indicator of interdisciplinarity (Enduri, Reddy, & Jolad, 2015; National Academy, National Academy, & Institute, 2005). At the same time, focusing on single topic at the time might enable deeper understanding the phenomenon under observation and is an important driver of knowledge building (March, 1991; O’Kane, Cunningham, Mangematin, & O’Reilly, 2015). Even if situated within the same research field, different communities and topics show different behaviours and communication patterns (Waltman & van Eck, 2012; Yan et al., 2012). Thus, it is important to consider topic-level analysis of each community to identify those patterns (Yan, 2014). Identification of different types of communities sheds a light on landscape and knowledge

pattern diffusions within and among them, and furthermore provide a solid basis for identification of leading individuals and their circumstances in relationship with topics and communities. Together with our familiarity with community in which we perform analysis, this approach offers a fine-grained perspective on topic-centred research communities, useful for both researchers and policymakers.

In previous work, we analysed the usages and meaning of the word ‘argumentation’ (and few complements) by using topic modelling techniques (Blei et al., 2003), as well as network visualisation techniques applied to words, authors, and sources (Cobo, López-Herrera, Herrera-Viedma, & Herrera, 2011) and citation analysis (Boyack & Klavans, 2010). The results offer a systematic view of cognitive territories in the field. In this work, we used the information on topics to create overlay maps of social and cognitive structure on documents containing word argumentation, whose usage in literature expanded in meanings, communities, and contexts in last few decades. In such way, we aim at expanding and complementing the systematic analysis of intellectual roots and scholarly communities that besides argumentation theory include several domains of studies, such as discourse analysis, informatics, and education (van Eemeren & Verheij, 2018).

Our data contains almost 10,000 publications containing word ‘argumentation’, as well as several related terms in title, abstract, and keywords, which we extracted from Scopus. We create social network of co-authorship in the field and seek to discover communities of collaboration. This social structure is then overlayed with cognitive structure, to unveil the differences and similarities among them. In addition, by exploring the diversity in individual production of intellectual leaders and combining the results with cognitive and social structure delineation, we aim at disclosing the connections between different parts of the field. We hope that the findings will contribute to further understanding of individuals strategies scientists use to choose their research problems and topics. While works with similar aims have been previously executed (e.g., Foster et al. (2015)) our approach allows for studying the fields without annotated entities. Additionally, as we do not use journal subject categorization to delineate topics within a field of study, this approach is suitable to be used without exploring additional databases for categorization.

2. Method

2.1. LDA and Case study

For the purpose of this study, we perform an analysis in the field of argumentation studies. Argumentation studies traditionally has been a small area of inquiry across philosophy and linguistics, where researchers studied how, by the means of human reasoning, a shared conclusion in dialogical process of discussion can be reached (Toulmin, 2003; Van Eemeren et al., 2014). The roots of the study are grounded in ancient Greek rhetoric and during 25 centuries of its history, there were only few redirections of subject, among which significant ones happened in Renaissance and in 20th century (Zarefsky, 2005).

With the beginning of the new century, the concept of argumentation expanded from classical rhetoric, logic, and dialectic to other areas of research as well. Today, argumentation is regarded an interdisciplinary research territory at the crossroads between computer science and philosophy (Reed & Koszowy, 2011). Techniques and results obtained in argumentation theory are used in artificial intelligence, education, law, etc. (Abbas & Sawamura, 2009), thus forming a new area of inquiry with a common reference to “argumentation”. Many suggest that the argumentation today has an important interdisciplinary appeal (Reed & Koszowy, 2011;

Van Eemeren et al., 2014). We intend to analyse the dynamics and patterns of interaction across different areas of inquiry within the field.

For that purpose, we used the dataset obtained from Scopus database containing the documents in English language, published between 2005 and 2019, with word ‘argumentation’ and several related terms that we included after consultation with expert from the field (e.g., ‘argument mining’, ‘argumentative discourse’, etc.). We excluded from results subject categories such as zoology, physics, and biochemistry, as we did not expect to find documents related to argumentation studies there. The querying was performed in September 2021 and contained 11,765 results. We manually excluded from the analysis the false positive results, the ones that are mostly related to usage of the term to describe author’s reasoning (e.g., “this line of argumentation, “my/his/her/our argumentation”, etc.) rather than to describe content. Our final dataset includes 9,550 documents.

We then performed LDA topic modelling (Blei et al., 2003), to identify ‘latent’ topics within our set of documents using Dirichlet distribution and process. The procedure was performed in Stata using ‘ldagibbs’ package. This procedure computes topics based on statistically significant distribution of words over documents assuming that documents are created as a mixture of topics and that topics are created as mixture of words. The procedure requires choosing the number of topics in advance, and solutions with numerous topics tended to generate topics that are less clearly distinguished. The choice of 8 topics showed the good balance between delineation of topics and the level of detail. This solution showed a document specificity (i.e. the mean of the topic assignment) between 0.74 (cluster 6) and 0.55 (cluster 8) with an average contrast of 0.62 – i.e. on the average the second topic assignment is only 62% of the first one. Therefore, the topics are sufficiently distinct and well-delineated, and display remarkable coherence in terms of content, authors, and sources.

Table 1. Topics in argumentation studies field

	<i>Topic name</i>	<i>Description</i>
topic1	Argumentation theory	Philosophical and rhetorical approach, studying argumentation as dialogical practice aiming to achieve common conclusion and to convince interlocutor during human dialogue (Van Eemeren et al., 2014).
topic2	Argumentation mining	Deals with automatic extraction of argumentation and identification of argumentative structures in natural texts using computer programs (Lippi & Torroni, 2016) .
topic3	Discourse and language	Application of formal semantic approaches to the description of argument schemes and dialogue; studying the theoretical relationship between basic categories of linguistic meaning (Van Eemeren et al., 2014).
topic4	Artificial intelligence	Deals with developing tools and methods for argument evaluation and argument invention, “logic continued by other means” (Rahwan & Simari, 2009).
topic5	Science	Studies of argumentation as an integral part of

	Education	instruction and learning (Erduran & Jiménez-Aleixandre, 2008).
topic6	Argumentation frameworks	Formalisms rooted in classical deductive reasoning used to reach conclusions in automatic and unambiguous manner (Bench-Capon & Dunne, 2007).
topic7	Public argumentation	Studying argumentation about public affairs in public setting (Zenker et al., 2019).
topic8	On-line argumentation	Dealing with argumentation shared in online environments and developing tools for the analysis.

We identified two major understandings of the word ‘argumentation’ in our data, with distinct main authors and publication venues, and vocabulary following it. First one is related to the rhetorical and philosophical tradition of argumentation theory (topic 1), which understands argumentation as dialogical practice aiming to achieve common conclusion and to convince interlocutor during human dialogue (Van Eemeren et al., 2014). The second understanding is related to logical tradition of argumentation frameworks (topic 6) in informatics and conceives argumentation as a set of formal rules through which conclusions can be reached unambiguously and automatically, hence being rooted in mathematics and formal logic (Dung, 1995). With growth of usage of the term ‘argumentation’ in literature new topics have been introduced.

In social sciences and humanities, specialized communities have emerged, applying the concepts of ‘argumentation theory’ to specific contexts of usage. The most prominent one has been the study of argumentation in science education (topic 5). In this broad tradition we identified communities that aim at facilitating the dialogue with broader communities in social sciences, such as discourse studies (topic 3) and public communication (topic 7), for example around the notion of communication contexts (Rigotti and Rocci 2006).

Rooted in information science, two new topics have emerged that cut across the divide between social sciences and humanities on the one hand, and informatics on the other hand, i.e. argumentation mining (topic 2) and on-line argumentation (topic 8).

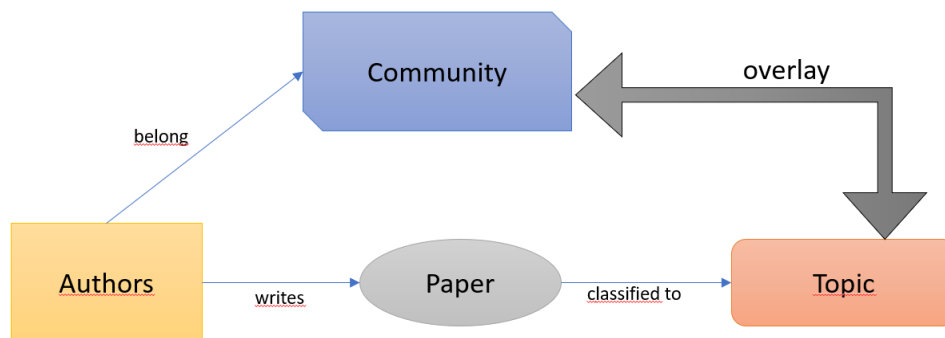
In such way, we found the evidence of multiple topics and approaches existing in the broad area of inquiry formed around the term argumentation. As these are not isolated one from another, there must exist the ways in which the connections appear. This can be achieved for example by integrating multiple topics in one’s work or by collaborating with people working in other topics. In this analysis we seek to identify those connections by identifying individuals whose work spans across multiple topics, i.e., authors with high topical diversity, as well as communities of authors connecting multiple topics.

2.2. Topical diversity in individuals’ production

As aforementioned, using the LDA analysis we classified papers to 8 topics. The authors of those papers are furthermore classified to communities (see next section for method). As each document has been authored by persons attributed to communities, we can overlay communities of authors and topics of documents to see in which way these overlap or not and to further identify communities and topics working on single or multiple topics. These findings can contribute to our understanding of the dynamics in the field and the circulation of the important findings. In addition, this work can shed the light on individuals whose workings

enable development and “drill down” of topics and the ones who contribute to the expansion of the topics across communities.

Figure 1. Overlaying communities and topics.



For that purpose, the first step of our analysis is identifying the profiles of individual authors. Besides the descriptive statistics of the number of publications, for individual authors we calculate the indices of diversity in their production. This measure is calculated in RStudio and is based the results of LDA classification of documents.

We first calculated the cosine (dis)similarity of topics based on the words they have in common (that we obtained in LDA) using the package ‘lsa’. LDA defines topics as probability distribution of terms within documents (Blei et al., 2003) and as such topics can be similar (or not) based on number and proportion of words they have (or don’t have) in common. As expected, topics 2, 4, 6 and 8 related to informatics and computer science are the most similar one to another, whereas topic 5 shows least similarity to other topics. Next, for each author we calculated Rao Stirling index of topic diversity, based on the topic cosine dissimilarity and the attribution of author’s documents to topics, following the procedure for computation of Rao Stirling diversity (Rafols & Meyer, 2010).

2.3. Network analysis

By tracing the links that appear when authors collaborate on a paper together, we can create social networks of co-authorships. Social network analysis studies social ties among actors, by detecting and interpretation social ties among actors (De Nooy, Mrvar, & Batagelj, 2011). In the case of co-authorship social network, the aim is creating a graph to represent the structure of the network, with a set of vertices which represent authors and set of links between vertices which represent co-authorship relationship. By connecting with some over the others, social actors create communities, which are densely connected groups of people with sparser connections between groups (Newman, Mark EJ, 2006). Besides community detection, social network analysis can shed light on position of individuals, connections, and distributions. Yet, as Yan et al. (2012) noted, communities in co-authorship networks do not contain information on cognitive structure and to obtain a better insight on dynamics in scientific community of interest, it is thus necessary to include both layers in the analysis. Although very informative, social network analysis and community detection based on collaboration or citation patterns fail in capturing the topicality of knowledge being exchanged. Thus, to “paint a better picture” these approaches should be enriched to include cognitive dimension as well.

The next step of our analysis is to create a social network of co-authorship from our dataset. Working together on an article reflects mutual intellectual and social influence (Wagner, Whetsell, & Mukherjee, 2019) and as such is an appropriate outlet for investigation of social and cognitive structure within a field.

Out of 9550 documents retrieved from Scopus, 6145 have at least two authors. There is one document with 55 authors that we exclude from analysis, since it influences largely the centrality measures of authors, yet most of the authors do not appear again in our dataset. This exclusion contributes to better visualisation in the co-authorship network, as well. The next document with the highest number of co-authors has 22 authors. We pair the authors of remaining co-authored documents (6144 documents in our data set) to create adjacency matrix and to create a social network of co-authorship in RStudio. We create the co-authorship network using Gephi (Bastian, Heymann, & Jacomy, 2009), software for social network analysis. Our graph has 8826 nodes (authors) and 17525 edges (co-authorships). We use Openord algorithm for creating the layout with default parameters, as well as Noverlap to ensure the readability of the graph. Many of the nodes are not connected to the largest component. If we exclude those authors, we are left with 2573 nodes (29,15%) and 8306 edges (47,4%) in giant component. We proceed with analysis of authors in the giant component.

For visualization, we set the size of the nodes to depend on the number of the document an author has in our dataset (the larger the node the more documents), and the colour of the node represents the social community of the node. We calculated the communities in the giant component using Gephi's modularity class algorithm (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008) with resolution parameter set to 1. We have tried calculation of the modularity with other values for, however the default value of 1 is the most informative one and groups together persons as expected. This approach yielded 42 communities with modularity 0.862. The modularity close to 1 indicates strong community structure (Clauset, Newman, & Moore, 2004).

To proceed, we turned back to the results of LDA classification and for each community to overlay them with each social community. Authors belong to one community yet can have documents in multiple topics. For each community, we thus look how many documents each community has attributed to topics. This approach allows for identification of different types of communities – ones that have documents belonging to single topic or multiple topics.

3. Results

3.1. Individual Profiles of Scientists

There are 10,694 individual authors in our data set with average number of documents 2.09. Out of those, 7894 authors have only one document and 1328 authors have 2 documents. The highest number of documents an author has is 126.

Types of authors

To proceed with the analysis, we divide the authors in 5 groups based on their number of documents as follows: 1) authors with 1 or 2 documents, 2) authors with 3 to 10 documents, 3) authors with 11 to 15 documents 4) authors with 16 to 30 documents, 5) authors with 31 or more documents.

The RS diversity index has low correlation (0.4) with number of documents an author has. After excluding authors with only few documents (less and equal than 3) as these people would automatically have low diversity, the correlation becomes lower (0.2).

Most of the authors (90%) have low diversity, between 0 and 0.09. Again, we divide the authors into five categories according to their RS index: 1) people with index between 0 and 0.05, 2) people with index between 0.06 and 0.1, 3) people with index between 0.1 and 0.15, 4) people with index between 0.15 and 0.2, 5) index above 0.2.

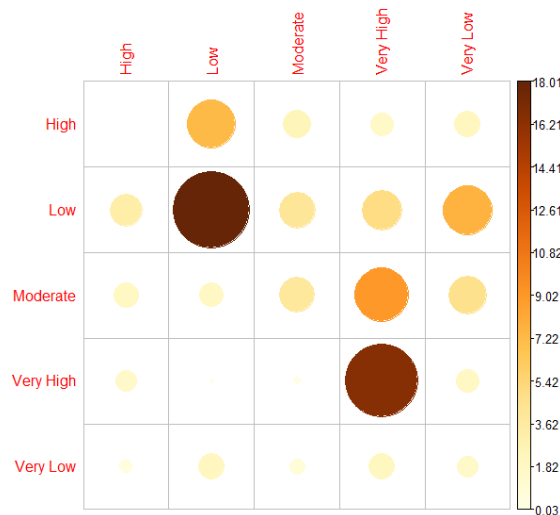
Then, by combining the measures of productivity in terms of number of documents and of topical diversity we classified authors into several categories, excluding the people with just one document as these will have 0 diversity by definition.

Table 2. Classification of authors (with at least 2 documents) based on number of documents and topical diversity.

		Diversity					
			<=0.05	<=0.1	<=0.15	<=0.2	>0.2
			Very Low	Low	Moderate	High	Very High
Documents	<=10	Very Low	1871	52	150	213	65
	<=20	Low	33	25	24	28	17
	<=40	Moderate	14	7	15	15	14
	<=60	High	0	5	5	2	3
	>60	Very High	1	1	0	5	8

The results of chi square testing show that the number of documents and the topical diversity are statistically significantly associated.

Figure 2. Pearson's Chi-squared test contribution (X-squared = 528.31, df = 16, p-value < 2.2e-16)



As expected, the topical diversity increases with the number of documents and there are statistically significant differences among the classes. Very low number of documents contributes highly to very low diversity and very high number of documents contributes highly

to very high diversity. As soon as authors produce a lot of documents their topical diversity increases. However, if we closely look into the profiles of individuals with high numbers of documents, we can see that there are different patterns of diversity, ranging from very low to very high.

There are 15 people with at least 60 documents in our dataset. We present their name, number of documents, topical diversity index, and the distribution of their documents across topics in the following table.

Table 3. Authors with highest number of documents, their topical diversity and attribution of documents to topics

Name	documents	diversity	topic1	topic2	topic3	topic4	topic5	topic6	topic7	topic8
Simari G.R.	126	0.153303	0	11	0	27	0	87	0	1
Walton D.	121	0.186439	72	31	2	11	0	3	0	2
Toni F.	119	0.225312	0	15	0	38	0	52	0	14
Bench-Capon T.	109	0.258915	16	28	0	27	1	28	5	4
Amgoud L.	97	0.168928	1	4	0	33	0	59	0	0
van Eemeren F.	94	0.066769	84	1	3	0	1	0	5	0
Prakken H.	90	0.240379	13	26	0	18	0	30	2	1
Atkinson K.	79	0.258772	13	9	1	28	0	19	1	8
Woltran S.	76	0.008181	0	1	0	0	0	75	0	0
Hunter A.	75	0.206273	1	8	0	26	1	37	0	2
Villata S.	75	0.255089	3	21	3	12	0	26	1	9
Reed C.	74	0.23803	7	30	4	18	0	4	0	11
Parsons S.	66	0.175281	1	6	0	42	0	13	0	4
Macagno F.	62	0.154528	46	4	5	1	6	0	0	0
Modgil S.	61	0.201224	1	3	1	23	0	31	0	2

The results show that the most productive persons have backgrounds in computer science, while only few authors work within argumentation theory, according to the attributions of documents they authored to topics. Looking at the attribution of documents to topics, we can identify people with numerous documents in topics 4 and 6, which we showed that are topics related to computer science. These authors can be grouped into following categories – a) authors with no documents in other topics, b) authors with documents in topics 2 and 8, c) authors with documents in topics 1 and 2.

The first group of authors are “pure” computer scientists. These people have most of their documents in topics artificial intelligence and argumentation frameworks, where they are dealing with defining principles and automated methods for machine deliberations, using formal logic for argumentation of autonomous systems, programming languages for autonomous agents, etc. As these two areas of inquiry are overlapping, these two topics as well

are closely resembling each other in outlets of publication as well as terms that they cover. Consequently, authors in this group have moderate topical diversity, nonetheless their work spans across more than one topic. In this group we have following authors, with their topical diversity in brackets: Simari G.R. (0.15), Amgoud L. (0.16), Parsons S. (0.17), Modgil S. (0.20).

Second category are authors whose work touches upon natural language processing and argument mining. This area of research uses computer programs for automatic extraction and identification of argumentation in natural languages (Lippi & Torroni, 2016). As the inputs in form of human-generated texts for this sort of analysis can be connected from online environments, topic 8 (online argumentation) can be easily connected to this category. Thus, besides formal logic and multi agent systems, the expertise of the authors in this group might include computational linguistics and are oriented towards more applied settings as well. Therefore, the topical diversity of authors in this group should be slightly higher than the previous one. Here we have Toni F. (0.22), Atkinson K. (0.25), Villata S. (0.25), Reed (0.23).

In third category we have authors whose work besides computer science reaches into argumentation theory, without documents in online argumentation topic, however. At the crossroads between argumentation analysis, argumentation mining and artificial intelligence, workings of these people integrate findings from social sciences and humanities as well as engineering. In this group we have authors Bench-Capon T. (0.26) and Prakken H. (0.24) with very high diversity. Considering different cognitive traditions of these topics, higher diversity is what we expect to find.

Among authors with most documents in argumentation theory topic we have two authors with moderate diversity, Walton D. and Macagno F. Macagno has most of the documents in topic 1 (74%). However, the results show that diversity in his work comes from the fact that his remaining documents belong to topics that are dissimilar to each other. Besides topic 2 argumentation mining, Macagno's work is in topic 3 discourse and language, as well as topic 5 science and education.

Walton D. per contra has numerous documents in topic 2 argumentation mining, as well as topic 4 artificial intelligence, thus he is creating the bridge between theory of argumentation and computer science. This author, however, differs from others who have numerous documents in computer science topics, as Walton's primary topic, in terms of number of documents, is topic 1 argumentation theory.

The creator of pragma dialectics approach to argumentation, Frans van Eemeren has almost 90% of documents classified to topic 1 argumentation theory, as well as 74% of his documents in our dataset published in journal "Argumentation library". The low diversity index confirms that he remained focused on single topic during the 15 years our dataset covered.

Woltran S. with 98% of his documents in topic 6, argumentation frameworks, is among the least diverse authors when it comes to distribution of documents across topics, with topical diversity of 0.008.

We can see thus that there are different patterns and drivers of connections between computer science and theory of argumentation. Even though most productive people show substantial diversity in their workings, there are some people who instead of covering multiple topics, show tendency of traditional knowledge exploitation and drill down into single topics.

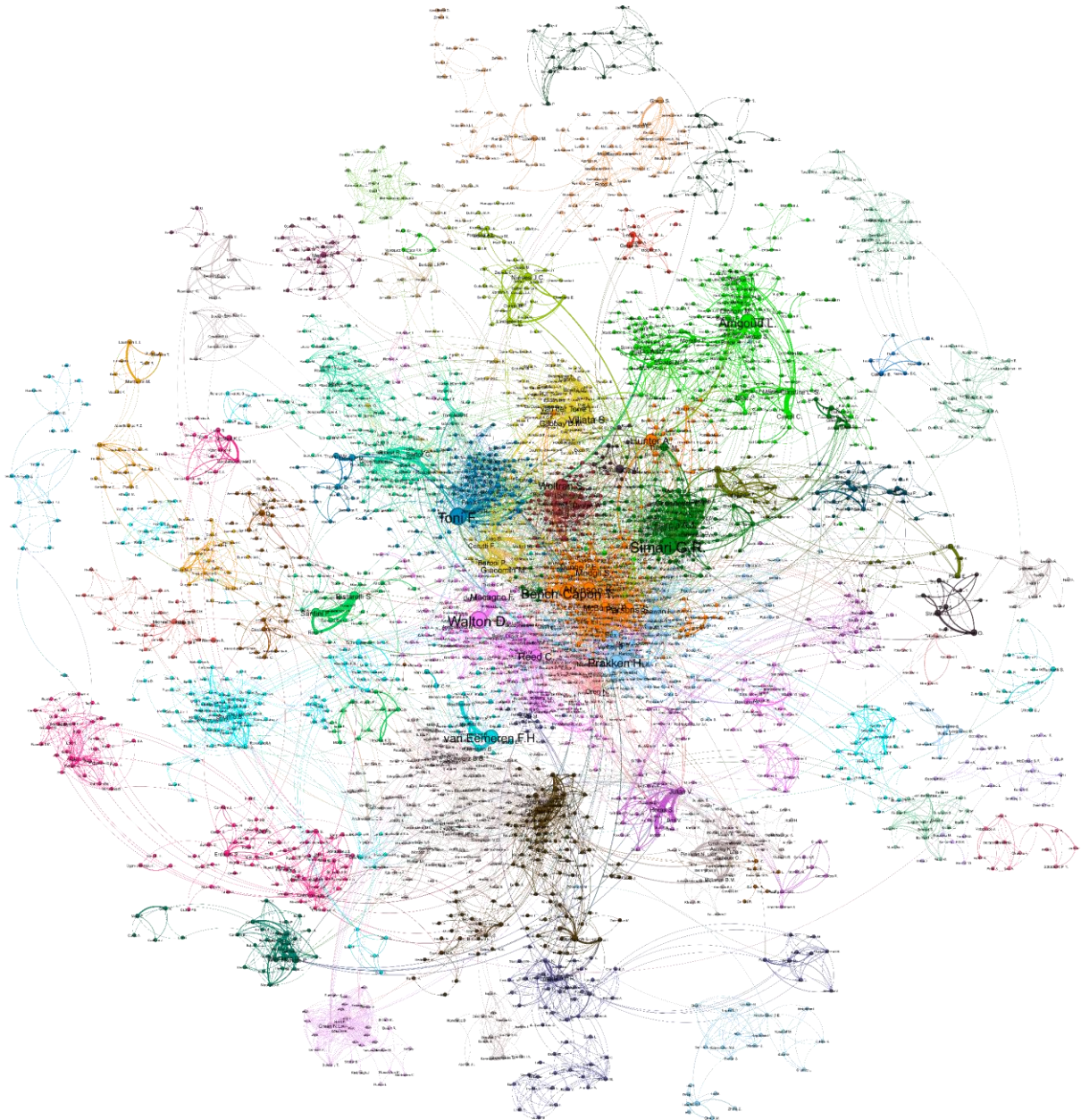
We are planning to extend these results with citation networks of these authors, to see how much difference/overlapping there is with current findings, as well as to extend the analysis of individual profiles to include authors whose numerous numbers of documents have been categorized top topics 3, 5, and 7.

3.2. Characteristics of Communities

The average degree of nodes in entire network is 3.9. There are 1755 weakly connected components. The giant component accounts for 1573 nodes (29.15%) and 8306 edges (47.4%).

Figure 3. Social network of co-authorship in argumentation.

Giant component with different colours representing different modularity classes.



There are 42 modules or communities in giant component in co-authorship social network. The average degree of nodes is 6.456. The author with the highest weighted degree is Simari G.R., the author with the highest closeness centrality is Toni F., the author with the highest betweenness centrality is Fischer F., while the author with the highest eigenvector centrality is Modgil S.

The figure clearly shows that there are many more communities than topics we identified. This indicates that communities overlap with topics in different ways. On periphery of the network, we have several communities where authors have connection to the remainder of the network only through one other author or very few authors. The core however is far more intertwined, with numerous connections. One would thus expect to find communities covering single topic, as well as communities covering few topics and numerous topics.

When we attribute documents of the authors in each community to the topics, we create the overlay of communities and topics. The results show that there are communities that are very “specific”, as most of their documents belong to the one topic. Most of these communities have the authors working in “Science Education” tradition. It is interesting to note that there is one community with tradition in informatics that is as well highly specific (community in deep red positioned centrally in the network) with 382 documents (96%) belonging to “argumentation frameworks” topic. We can identify very “diverse” communities as well, whose documents spread across multiple topics. An example of those is community formed around the author Walton D. (purple on the network).

We classify communities based on attribution of the documents of all authors to different topics as follows: a) monotopic communities – with at least 75% of the documents in single topic, or with second largest topic containing less than 15% of documents; b) two topic communities – with at least 80% of documents in top two topics (by number of documents), where second largest group has at least 15% of documents; c) distributed communities – communities with less than 80% of documents in top two topics and with at least 3 topics containing at least 15% of documents. The following table shows the number of communities attributed to each topic and classified according to the classification above.

Table 4. Categorization of communities and their numbers across topics

	topic1	topic2	topic3	topic4	topic5	topic6	topic7	topic8	Total communities
Mono	1	3	0	3	9	2	0	0	18
Two	1	3	0	5	3	4	1	3	10
Distributed	3	11	1	10	1	8	0	7	14

We identified 18 communities that have documents in single topic. Most of them (9) are in science and education topic, topic 5, that we saw in previous work is the most closed one in terms of sources and authors, as these have publications in specialized journals. We identified one mono topic community with documents in argumentation theory, as well as several communities that work on single topic with computer science background (topics 2, 4, and 6).

The table shows that computer science related topics have major number of two topics and distributed topic communities. Two topics communities are communities which are associating two different topics. We can see that there is 1 group that connects topic 1 with another topic, 3 communities that are connecting topic 2 with other topics, etc. To illustrate this, we can look at community to which author Amgoud L. belongs (bright green in the co-authorship network). This community has 27% of documents in topic 4 and 57% of documents in topic 6. Another example is community with author Fisher F. (in dark brown on co-authorship network) that bridges topics 5, science and education with 68% of documents, and topic 2, argumentation

mining with 23% of documents. Fisher and co-authors published numerous documents in conference proceedings named “Computer-Supported Collaborative Learning”.

There are 14 communities whose documents are distributed over more than 2 topics. Here, we have 41 connections across topics, where 11 of the communities connect three topics and remaining 3 communities spread across 4 topics. An illustration is community in purple with authors Walton D. and Reed C., connecting topic 1 argumentation theory (30% of documents), topic 2 argumentation mining (30% of documents), and topic 8 online argumentation (15% of documents), with documents published in sources such as “Argumentation”, “Informal Logic”, “Frontiers in Artificial Intelligence and Applications”, “Argument and Computation”. Another example is community in mint with authors Sartor G. and Rotolo A. Authors in this community have 34% of documents in topic 4 artificial intelligence, 24% of documents in topic 6 argumentation frameworks, 18% of documents in topic 2 argumentation mining and 17% of documents in topic 8 online argumentation, thus covering only topics with cognitive background in computer science.

We intend to extend this work with calculations of Rao Stirling diversity of each community.

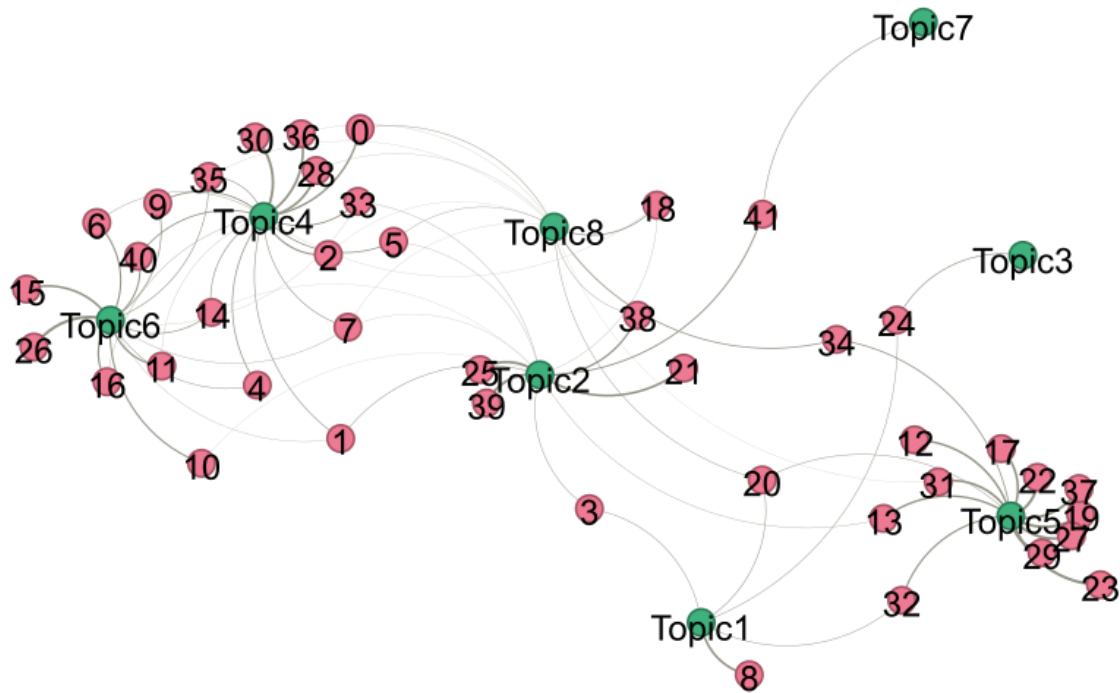
The figure below shows co-authorship map with colours representing the type of community – orange for mono topic, green for two topics, and lilac for distributed topics.

perceived usefulness of ideas (Deichmann et al., 2020), one would expect that that this exposure to new knowledge would be translated into more diversity in one’s work and covering multiple topics. This is because high closeness centrality means high exposure to various disparate social circles in social network, central individuals are prone to facilitated broader thinking and connecting unrelated areas (Perry-Smith, 2006), which might translate to higher topical diversity and working of several topics. The communities with Woltran S. and Van Eemeren F. are not in accordance with that, as they are centrally positioned yet belonging to monotopic communities. These are also very productive authors with low diversity.

3.3. Linking Communities to Topics

The following map represents the bi-partite network of communities and topics. For better representation, we display only edges where community has at least 15% of documents in topic, as well as rescaled thickness of edges 0.1 to 2.0.

Figure 5. Bipartite network of communities and topics.



First, there is a lot of overlapping between communities around topic 4 – Artificial Intelligence and topic 6 – Argumentation frameworks. There exist certain overlaps between these two topics and topic 2 – Argumentation mining, as well as few connections with topic 8 – On-line argumentation. Topics 2 and 8 act as connectors of computer science part with humanities part of argumentation field. Topic 5 – science and education has numerous “specific” communities with documents belonging just to that topic as well as few communities with documents in common with topic 1, topic 2 and topic 8. Topics 3 and 7 remain peripheral, without specific documents. We have seen in our earlier work that these topics are the most distinct in terms of shared with remainder of communities. Interestingly, topic 1 – argumentation theory came to occupy position that is much less central and with fewer connections to other topics.

The communities that are connected to topic 1 show some interesting results. There is one community that is very specific to this topic – community 8. The most prominent member of this community, in terms of number of documents is author van Eemeren, who himself has low

topical diversity. The same goes for his co-authors, as the results show that 80 % of all documents of members of this group are classified in topic 1. Next community connected to topic 1 is community 3 – the least specific one. The results show that members of this group act as a bridge between the argumentation mining topic and argumentation theory topic. The most productive members of this community are among authors with highest topical diversity (Walton D., Reed C., Macagno F.), suggesting that their collaboration is fuelling the exchange between computer science and humanities part of argumentation. Community 24 connects topics 1 and 3, thus enabling the connection between argumentation theory and discourse and language analysis. Community 32 connects argumentation theory and science and education topics; however, it is a small community with number of documents insufficient for analysis. The same goes for community 20.

This shows that topics have both specific communities and communities that connect them with other topics. We might extend this by saying that there are groups of people who create ideas and groups of people who disseminate the ideas. The first category is related to focusing activities important for deepening understanding within topics through specialization in given topic, developing taxonomy and methods. Specialties as such are important in science because of their crucial role for creation and validation of scientific knowledge (Morris & Van der Veer Martens, 2008). In second category, people combine ideas from multiple topics, thus potentially creating bridges or even new directions for research. This combining of ideas from multiple cognitive areas is another important driver of science evolution (National Academy et al., 2005; Wang, Veugelers, & Stephan, 2017). Both mechanisms are part of normal ecology of science and highlight some important features of interdisciplinary fields – brokerage and development of new ideas.

4. Conclusions

The results we obtained in this analysis go in hand with suggestions in literature that to truly understand the dynamics within scholarly communities, we should study both its topical and social structure. Not only there is distinctiveness and interconnections in topics that a community covers, but we can also see that individuals can be carriers of both diversity and specificity. Communities can form around one single topic when authors whose areas of interest are highly similar engage in collaboration (example community 8). On the other hand, when authors whose individual work shows low diversity, come to co-author with others that themselves also have low topical diversity, but different research interests, thus formed communities show low specificity (case of community 8). In addition, works of individuals can be very diverse, and these people tend to belong to communities with low specificity (case of community 3). All topics seem to be related to two types of communities – the bridging ones and the focused ones.

Our findings highlight important mechanisms of circulation of ideas in science. We have seen some communities that take up roles of brokerage, reminding of findings of Burt (2004). Others focus on single topics and contribute to developing and deeper examination and validation of ideas. This is consistent to normal ecology of science as both spanning multiple topics and drill down of single topic contributes to further development of science.

When it comes to analysis at the individual level, we have shown that high productivity does necessarily imply high diversity in research. Although major part of relevant and productive authors has substantial diversity in the topics they engage in, there are some exceptions. Even when taking in consideration central positions of authors, and thus exposure to multiple ideas, some people do not engage in works characterized with high topical diversity.

These findings open series of questions. First is for which reasons people choose to work on multiple topics over single topic (or vice versa). As literature shows, engaging in research that combines multiple knowledge domains has potential of achieving high influence and awards, yet it carries substantial risk of failing (Foster et al., 2015), why do people still decide to engage in bridging multiple topics? On the other hand, if bridging multiple topics allows for more originality, why do people choose to engage in traditional knowledge exploring? The next would be to investigate the implications of such choices for the overall development of science.

Open science practices

We obtained the data for this analysis from the Scopus, which requires subscription to access the data, with Open Access filters that provide open access options. After consideration of options to obtain the data, we opted for Scopus because of its wide coverage of the titles belonging to social sciences and computer science, which is very relevant for this research. We intend to submit the related work in open access journal and thus make the data publicly available. The software (RStudio and Gephi) used for the analysis have open-source licences, and the code can be available upon request.

Competing interests

Authors have no conflict of interest.

References

- Abbas, S., & Sawamura, H. (2009). Developing an argument learning environment using agent-based ITS (ALES). International Working Group on Educational Data Mining,
- Backstrom, L., Huttenlocher, D., Kleinberg, J., & Lan, X. (2006). Group formation in large social networks: Membership, growth, and evolution. Paper presented at the Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 44-54.
- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. Paper presented at the Proceedings of the International AAAI Conference on Web and Social Media, , 3. (1) pp. 361-362.
- Bench-Capon, T. J. M., & Dunne, P. E. (2007). Argumentation in artificial intelligence. *Artificial Intelligence*, 171(10), 619-641. doi:<https://doi.org/10.1016/j.artint.2007.05.001>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan), 993-1022.
- Blondel, V. D., Guillaume, J., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008.
- Boyack, K. W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology*, 61(12), 2389-2404.
- Burt, R. S. (2004). Structural holes and good ideas. *American Journal of Sociology*, 110(2), 349-399.
- Burt, R. S. (2000). The network structure of social capital. *Research in Organizational Behavior*, 22, 345-423. doi:10.1016/S0191-3085(00)22009-1
- Callon, M., Courtial, J., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, 22(2), 191-235.
- Cambrosio, A., Cointet, J., & Abdo, A. H. (2020). Beyond networks: Aligning qualitative and computational science studies. *Quantitative Science Studies*, 1(3), 1017-1024.
- Clauset, A., Newman, M. E., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E*, 70(6), 066111.
- Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). Science mapping software tools: Review, analysis, and cooperative study among tools. *Journal of the American Society for Information Science and Technology*, 62(7), 1382-1402.
- De Nooy, W., Mrvar, A., & Batagelj, V. (2011). *Exploratory social network analysis with pajek*. Cambridge, Mass.: Cambridge University Press.
- Deichmann, D., Moser, C., Birkholz, J. M., Nerghe, A., Groenewegen, P., & Wang, S. (2020). Ideas with impact: How connectivity shapes idea diffusion. *Research Policy*, 49(1), 103881. doi:10.1016/j.respol.2019.103881
- Dung, P. M. (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77(2), 321-357.
- Enduri, M. K., Reddy, I. V., & Jolad, S. (2015). Does diversity of papers affect their citations? evidence from american physical society journals. Paper presented at the 2015 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), pp. 505-511.
- Erduran, S., & Jiménez-Aleixandre, M. P. (2008). *Argumentation in science education. Perspectives from Classroom-Based Research*. Dordre-Cht: Springer,
- Foster, J. G., & Evans, J. A. (2011). Metaknowledge. *Science*, 331(6018), 721-725. doi:10.1126/science.1201765
- Foster, J. G., Rzhetsky, A., & Evans, J. A. (2015). Tradition and innovation in scientists' research strategies. *American Sociological Review*, 80(5), 875-908.
- Gruhl, D., Guha, R., Liben-Nowell, D., & Tomkins, A. (2004). Information diffusion through blogspace. Paper presented at the Proceedings of the 13th International Conference on World Wide Web, pp. 491-501.

- Lippi, M., & Torroni, P. (2016). Argumentation mining: State of the art and emerging trends. *ACM Transactions on Internet Technology (TOIT)*, 16(2), 1-25.
- March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, 2(1), 71-87.
- Milojević, S., Sugimoto, C. R., Yan, E., & Ding, Y. (2011). The cognitive structure of library and information science: Analysis of article title words. *Journal of the American Society for Information Science and Technology*, 62(10), 1933-1953.
- Morris, S. A., & Van der Veer Martens, B. (2008). Mapping research specialties." annual review of information science and technology. edited by blaise cronin. medford NJ: Information today.
- National Academy, o. S., National Academy, o. E., & Institute, o. M. (2005). Facilitating interdisciplinary research. Washington, DC: The National Academies Press. doi:10.17226/11153
- Nersessian, N. J. (2005). Interpreting scientific and engineering practices: Integrating the cognitive, social, and cultural dimensions. *Scientific and Technological Thinking*, , 17-56.
- Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 5200-5205.
- Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23), 8577-8582.
- O'Kane, C., Cunningham, J., Mangematin, V., & O'Reilly, P. (2015). Underpinning strategic behaviours and posture of principal investigators in transition/uncertain environments. *Long Range Planning*, 48(3), 200-214.
- Osborne, F., Scavo, G., & Motta, E. (2014). A hybrid semantic approach to building dynamic maps of research communities. Paper presented at the Knowledge Engineering and Knowledge Management: 19th International Conference, EKAW 2014, Linköping, Sweden, November 24-28, 2014. *Proceedings 19*, pp. 356-372.
- Otte, E., & Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. *Journal of Information Science*, 28(6), 441-453. doi:10.1177/016555150202800601
- Perry-Smith, J. E. (2006). Social yet creative: The role of social relationships in facilitating individual creativity. *Academy of Management Journal*, 49(1), 85-101.
- Rafols, I., & Meyer, M. (2010). Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience. *Scientometrics*, 82(2), 263-287.
- Rahwan, I., & Simari, G. R. (2009). *Argumentation in artificial intelligence* Springer.
- Reed, C., & Koszowy, M. (2011). The development of argument and computation and its roots in the Lvov-Warsaw school. *Studies in Logic, Grammar and Rhetoric, Special Issue of the Argumentation Series on Argument and Computation*, Ed.Koszowy, M, 23(36), 15-37.
- Taramasco, C., Cointet, J., & Roth, C. (2010). Academic team formation as evolving hypergraphs. *Scientometrics*, 85(3), 721-740.
- Toulmin, S. E. (2003). *The uses of argument* Cambridge university press.
- Van Eemeren, F. H., Garssen, B., Krabbe, E. C., Henkemans, A. F. S., Verheij, B., & Wagemans, J. H. (2014). *Handbook of argumentation theory*.
- van Eemeren, F. H., & Verheij, B. (2018). *Argumentation theory in formal and computational perspective* College Publications.
- Wagner, C. S., Whetsell, T. A., & Mukherjee, S. (2019). International research collaboration: Novelty, conventionality, and atypicality in knowledge recombination. *Research Policy*, 48(5), 1260-1270.
- Waltman, L., & van Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology*, 63(12), 2378-2392. doi:10.1002/asi.22748
- Wang, J., Veugelers, R., & Stephan, P. (2017). Bias against novelty in science: A cautionary tale for users of bibliometric indicators. *Research Policy*, 46(8), 1416-1436. doi:https://doi-org.proxy.sbu.usi.ch/10.1016/j.respol.2017.06.006
- Yan, E. (2014). Topic-based PageRank: Toward a topic-level scientific evaluation. *Scientometrics*, 100(2), 407-437.
- Yan, E., Ding, Y., & Jacob, E. K. (2012). Overlaying communities and topics: An analysis on publication networks. *Scientometrics*, 90(2), 499-513.

- Yan, E., Ding, Y., Milojević, S., & Sugimoto, C. R. (2012). Topics in dynamic research communities: An exploratory study for the field of information retrieval. *Journal of Informetrics*, 6(1), 140-153. doi:10.1016/j.joi.2011.10.001
- Zarefsky, D. (2005). *Argumentation: The study of effective reasoning*
- Zenker, F., von Laar, J. A., Abreu, P., Bengtsson, M., Castro, D., Cooke, M., et al. (2019). Goals and functions of public argumentation.
- Zhao, Z., Li, C., Zhang, X., Chiclana, F., & Viedma, E. H. (2019). An incremental method to detect communities in dynamic evolving social networks. *Knowledge-Based Systems*, 163, 404-415. doi:10.1016/j.knosys.2018.09.002
- Zhou, D., Ji, X., Zha, H., & Giles, C. L. (2006). Topic evolution and social interactions: How authors effect research. Paper presented at the Proceedings of the 15th ACM International Conference on Information and Knowledge Management, pp. 248-257.