Exploring Ideation: Knowledge Development in Science Through the Lens of Semantic and Social Networks

Christine Moser
Dept. Of Organization Sciences, VU University Amsterdam
c.moser@vu.nl

Julie M. Birkholz
Dept. Of Organization Sciences, VU University Amsterdam
j.m.birkholz@vu.nl

Dirk Deichmann
Dept. Of Organization Sciences, VU University Amsterdam
d.deichmann@vu.nl

Iina Hellsten
Dept. Of Organization Sciences, VU University Amsterdam
i.r.hellsten@vu.nl

Shenghui Wang
OCLC Research, Leiden
shenghui.wang@oclc.org

Abstract

In this paper, we explore changes in both structural and semantic characteristics of a scientific social network. We trace the emergence of knowledge, what we refer to as ideation, through publication data from two conferences in a sub-field of Computer Science. Social network analysis is used to determine structural characteristics of the co-authorship networks, and we perform semantic network analysis on title words of articles to trace content of topics over time. We find that the emergence of new topics is accompanied by a tendency toward less dense and transitive, but more evenly distributed, social networks. We show that ideation is fostered in a relatively loose environment, as the field does not cling to one particular topic. On the contrary, there is enough ‘slack’ for new knowledge topics to emerge and consolidate over time. Findings contribute to research on social network antecedents, innovation and science studies.

1. Introduction

‘Hot’ research topics promise several rewards for researchers, the affiliated institutions, and society at large. Novel and useful knowledge have the potential to not only generate new knowledge or address and solve particular societal problems, but also generate new sources of income or secure established sources of income, put researchers and universities in the spotlight, and thereby enhance a researcher’s career and an institution’s position in the rankings. Research has shown that the generation of new knowledge is a co-evolution of both social interaction and ideas [1]. In order for ideas to develop, stability has to be found in the ‘epistemic community’, with specific structures that better facilitate knowledge development than others.

Theories on science suggest that new knowledge emerges from the interactions among scientists themselves [2], are guided by the disciplined practices of previous knowledge [3] and constrained by the formal organizational bodies that attempt to steer the scientists producing knowledge [4]. The production of new knowledge, which is often seen as the ultimate aim of science, is communicated through the dissemination of scientific publications. However, the intricate ways in which scientists aim to develop new knowledge are understudied. As such, we contribute to this debate by conducting an exploratory study of the emergence of ideation through the study of semantic and social networks. We focus on the evolution of structural characteristics of both social and semantic networks to explain the process of ideation. In doing so, we contribute to the literature on social networks, innovation and science studies which often have taken a more static view on this issue in the past.

In the next section we layout the theoretical framework of ideation. The methods section describes the use of social and semantic network analysis and the case we explore to understand ideation in science. Finally, we present our findings and discuss contributions, limitations and future research directions.

2. Theory

2.1 Ideation

The central aim of this paper is to show how both social and semantic networks play a role in ideation, defined as the process of the generation of knowledge topics (ideas) [5]. Earlier research addressed facets of this question [6-7], from an organizational [8] or creativity perspective [9], focusing on the dialectical process of knowledge creation [10] or individual attributes [11]. These studies suggest the process of ideation is a practice of reaching consensus. New groundbreaking knowledge topics emerge through the blending of existing conceptions and structures [12].
For example, in science, the context we explore in this study, this consensus is the cycle of communication practices which leads to the acceptance or rejection of a knowledge topic into a community. Through establishing connections between known concepts and solutions, innovative themes emerge that question the status quo [13] and that render existing views and structures obsolete, similar to a process of ‘creative destruction’ [14]. We specifically aim to address the role of social and semantic networks in facilitating ideation.

### 2.2 Social networks and ideation

Through social ties, people communicate knowledge, absorb different views and feedback, discuss improvements to their ideas [15] and subsequently reach consensus. More specifically, relationships are used to leverage each other’s experiences, get emotional support and bundle resources for novel efforts [16-17]. Working on breakthrough issues requires a knowledge-intensive environment in which people must know and trust each other [18-19], as the success or ‘hotness’ of an idea is often unclear at the very beginning.

A number of studies have investigated which network structures might be beneficial for ideation and/or knowledge production and innovation. They have shown that sparse networks lead to an increase in access to diverse information. Sparse networks facilitate performance for non-complex, less knowledge-intensive tasks [20-21]. In contrast, dense social network structures provide a beneficial role in the handling and transfer of complex information that is difficult to verify [18, 20, 22], which applies to new knowledge. Strong ties, characterized by frequent interaction, long duration, and emotional closeness, are most beneficial for ideation - that is, the generation of new knowledge topics [5, 23]. These structures make exchange processes more efficient and less risky as a result of shared understandings, habits, and experiences [16, 24].

However, a dense network structure can also harm future chances of ideation, as it potentially restrains actors from engaging in other, potentially more knowledgeable relationships [16]. This locks people into familiar, trusted, and immediately available relationships, ultimately leading to increasing similarity [25], where the pool of information homogenizes too much [26]. On the other hand, if there are too many actors in the field it becomes difficult to reach consensus and integrate all of the different views of the various people [27]. Without consensus, the groups remain in a stage of brainstorming and exchange, endlessly generating more and more ideas, but failing to realize them due to the lack of collective support where concrete themes can be formulated. With network structures too far from optimal levels, people are more likely to disclose and advocate ideas that conform to commonly held expectations [28]. This has a negative effect on the success of novel research topics.

Specifically in studies on science networks, the topic of interest of this paper, co-author networks display small world properties which have a high number of nodes that are not directly connected to one another, but are connected by short distances via neighboring nodes [29-31]. These networks have properties of clustering hierarchies [32], which are fed by transitivity [33].

As a result, in regards to social network structures density and transitivity both have consequences for the development of new knowledge topics.

### 2.3 Semantic networks and ideation

An idea does not solely garner fame or acceptance through the interaction of the right people, in dense networks of communities. The content of the idea also plays a role in how ideas come to be accepted.

Semantic network analysis provides a tool to identify the content of a set of texts. Semantic networks have been defined as networks in which terms are the nodes and co-occurrences of those terms form the ties. Co-word analysis is a common example of semantic network application. Semantic co-word analysis takes into account both the co-occurrences of words in text documents and the positioning of the words in a set of documents [34-35]. The analysis of these networks has been defined by three measures: density (frequency of co-occurrences), conductivity (directional ties in and out of the node) and consensus (sharing of ties and links) [36].

The consideration of semantic network influence on social network structures has only recently been explored. Most studies have concentrated on establishing the relationship and co-evolution between social and semantic networks [1, 37-38]. Recent studies have established that the position of a node in either a semantic or a social network influences the position in the other network [39, 40]. Consequently, central topics play a role in the positions that actors occupy in a social network.

There is a clear need for more studies that specifically address the shift in the characteristics of the network, as these help to explain the duality
between structures and individuals [41]. Consequently, we question the potential relationship between semantic and social network structures in stimulating ideation. Instead of connecting single topics to individual actors we investigate the overall connections of structural network characteristics between both semantic and social networks to determine the dual effect of social and semantic networks on acceptance of new knowledge.

3. Methods

To investigate our research question, how ideation occurs considering the effects of social and semantic network structures, we conducted network analyses on publication data. This exploratory approach enables us to provide a relational network with meaning, and theorize on how characteristics of both networks may be intertwined in generating new knowledge.

3.1 Setting

In this paper, we investigate ideation in the field of Computer Science. The discipline of Computer Science exists for 30 plus years and is an intellectually unified field with a number of mature sub-fields such as bioinformatics, artificial intelligence/cognitive science, cybernetics, quantum computing, distributed systems as well as business applications. In this paper, a population of Semantic Web scientists is investigated. Semantic Web is a sub-field of Computer Science that has emerged in the last 15 years and develops structures and processes to organize, access, and share information on the Web. It specializes in decentralized technologies of knowledge representation by providing ‘a language that expresses both data and rules for reasoning about the data and that allows rules from any existing knowledge –representation system to be exported onto the Web [42]. These languages, RDF, OWL and XML are referred to as ontologies (domain-specific vocabularies) which facilitate interoperation, so that metadata on the Web can be integrated.

3.2 Data

We used publication data, including title and author information, from the Semantic Web Conference Corpus (as known as the Semantic Web Dog Food Corpus) as a source for the network analyses. The Dog Food Corpus is an electronic bibliometric database with records on conferences, individual publications and institutions involved in Semantic Web research [43]. In the field of Semantic Web, two annual conferences comprise the knowledge dissemination moments: the International Semantic Web Conference (ISWC) and the European Semantic Web Conference (ESWC), both launched in 2006. The majority of knowledge dissemination in Computer Science is realized through conferences and conference proceedings [44], a trend seen in many other technically oriented fields. This allows the field to address research in a timely manner, as opposed to the traditional fashion of publishing in books and print journals. Conference proceeding publication data was extracted from the Dog Food Corpus on the International and European Semantic Web Conferences from 2006–2010.

From within the total larger set of queried data, a selection of ‘active scientists’ was made. Scientists were considered active if they were present, by way of having an accepted paper in the proceedings, in one of the two conferences per year at least 4 of the 5 years and worked with at least one author during this 5 year period (no single authored papers exist in this set). Ninety-two active scientists were identified. We then analyzed title and author information of their accepted papers. From this set of publication data co-authorship is used as a measure of a social network among scientists [45-46], and title words as a measurement for topics of the semantic networks. Our data set included 242 publications, with 516 unique authors.

3.3 Analyses

As a first step, we removed stop-words from the publication titles (e.g. ‘a’, ‘that’, etc.). All other title words were considered as individual topics, phrases were not considered in this analysis. We then derived the network structures of author and semantic networks. Semantic networks were compiled from an affiliation matrix of the data: author information was entered in rows and title words in columns, with the value indicating the frequency of one author including a specific word in the paper title. These languages, RDF, OWL and XML are referred to as ontologies (domain-specific vocabularies) which facilitate interoperation, so that metadata on the Web can be integrated.
assessed for both social and semantic networks to explore connections between the two in describing ideation.

3.3.1 Network measures. As explained in the theoretical framework, density and transitivity shed light on the structural properties of the networks, whereas degree centrality informs about positions of individual nodes in the network. We calculated the density, transitivity and degree centrality measures of both social and semantic networks for each year using Ucinet 6 [48] and used Netdraw [49] to visualize these networks. Density assesses the degree of dyadic connection in a given network [50]. It is measured by the ratio of ties divided by the number of pairs between the ties, thus calculating a measure of all possible dyadic relations. Transitivity explains the clustering in the networks, allowing us to reflect on the tendency of groups of nodes to interact. When measuring transitivity, the triad level is considered to assess the number of possible relations within triads (that is, ties between three nodes). Transitivity is calculated using a routine which counts the relative prevalence of four possible types of relations, which are no tie, one tie, two ties or three ties [50] between three actors. We divided the number of transitive triads by the number of all present triads. Finally, degree centrality refers to the total number of ties that actors have in a given network. More formally speaking, it entails ‘the number of vertices adjacent to a given vertex in a symmetric graph is the degree of that vertex’[48]. Degree centrality is calculated using the underlying formula expressed as

$$\Sigma(c_{\text{max}} - c(vi))$$

divided by the maximum value, where $c(vi)$ is the degree centrality of vertex vi, given vertices v1... vn and maximum degree centrality $c_{\text{max}}$.

These three network measures give insight into the characteristics of both the social and semantic networks at the various time periods, allowing us to characterize the situations under which knowledge topics evolve.

3.3.2 Statistics. To assess the statistical difference between the mean degree centrality of authors and topics over time, we performed Paired Sample Tests. The overall difference in degree centrality was calculated through One-way ANOVA (analysis of variance) that allowed us to test statistical differences between groups. Both routines were performed using Stata.

4. Findings

First, we discuss three structural characteristics of the networks through centrality, density and transitivity. Then we focus on the individual measure of degree centrality and trace the evolution of both author and topic centrality over time. Finally, we explore how knowledge topics are organized using a number of topics as illustration.

Figure 1 shows the network density of both the social and semantic networks over time. The density of both the author networks and topic networks remains relatively stable. In Figure 2, the two networks’ transitivity is depicted, showing shifts in transitivity tendencies in the author networks, and a steadier pattern in the topic networks.

![Figure 1: Author and topic networks density 2006-2010](image1)

![Figure 2: Author and topic networks transitivity 2006-2010](image2)

![Figure 3: Author and topic mean degree centrality 2006-2010](image3)
The topic degree centrality stays stable, author degree centrality shifts (Figure 3). A statistical comparison of these graphs further corroborates the abovementioned findings, mean author degree centrality does not significantly differ from one year to the other (see Table 2 for details); however there is a significant overall effect of time on the degree centrality of authors (F(4,1689)=3.12, p< .05, \(\omega^2=1.98\)). In the topic networks a two-sample t-test illustrates that the mean topic degree centrality does not significantly differ from one year to the other, except for the difference from year 2007 to 2008 (see Table 2 for details), but again that there is a significant overall effect of time on the degree centrality of topics (F(4,2135)= 3.21, p< .05, \(\omega^2=.06\)).

Table 1: Network measures

<table>
<thead>
<tr>
<th></th>
<th>Author density</th>
<th>Topic density</th>
<th>Author Transitivity</th>
<th>Topic Transitivity</th>
<th>Author degree mean</th>
<th>Topic degree mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>0.012</td>
<td>0.1476</td>
<td>74.182</td>
<td>13.76</td>
<td>2.78</td>
<td>.05343</td>
</tr>
<tr>
<td>2007</td>
<td>0.008</td>
<td>0.1625</td>
<td>54.420</td>
<td>17.34</td>
<td>3.62</td>
<td>.05762</td>
</tr>
<tr>
<td>2008</td>
<td>0.0086</td>
<td>0.1471</td>
<td>52.240</td>
<td>14.96</td>
<td>3.462</td>
<td>.04960</td>
</tr>
<tr>
<td>2009</td>
<td>0.0088</td>
<td>0.1237</td>
<td>61.463</td>
<td>15.91</td>
<td>3.243</td>
<td>.04574</td>
</tr>
<tr>
<td>2010</td>
<td>0.0088</td>
<td>0.1382</td>
<td>56.516</td>
<td>14.33</td>
<td>3.359</td>
<td>.04669</td>
</tr>
</tbody>
</table>

Table 2: Paired Sample Tests of mean degree centrality

<table>
<thead>
<tr>
<th></th>
<th>Diff. Mean</th>
<th>Diff. Std. Error</th>
<th>t</th>
<th>df</th>
<th>Sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author networks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006 – 2007</td>
<td>.168</td>
<td>.237</td>
<td>.710</td>
<td>533</td>
<td>.478</td>
</tr>
<tr>
<td>2008 - 2009</td>
<td>.218</td>
<td>.132</td>
<td>1.661</td>
<td>773</td>
<td>.097</td>
</tr>
<tr>
<td>2009 - 2010</td>
<td>.116</td>
<td>.132</td>
<td>.881</td>
<td>752</td>
<td>.379</td>
</tr>
<tr>
<td><strong>Topic networks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006 - 2007</td>
<td>-.004</td>
<td>.005</td>
<td>.0872</td>
<td>761</td>
<td>.383</td>
</tr>
<tr>
<td>2007 - 2008</td>
<td>.008</td>
<td>.004</td>
<td>1.976</td>
<td>969</td>
<td>.048</td>
</tr>
<tr>
<td>2008 - 2009</td>
<td>.004</td>
<td>.004</td>
<td>1.026</td>
<td>901</td>
<td>.305</td>
</tr>
<tr>
<td>2009 - 2010</td>
<td>-.001</td>
<td>.004</td>
<td>.262</td>
<td>900</td>
<td>.794</td>
</tr>
</tbody>
</table>

In Figures 4a-e, the social networks are visualized. The nodes represent authors, ties through co-authorship are indicated by lines. The size of nodes represents authors’ degree centrality. A visual inspection of these graphs shows that the distribution of the author networks stays relatively similar after the initial year of 2006. Thus, the mean degree centrality is fairly consistent over time.

Figure 4a: Author network 2006 (degree centrality)

Figure 4b: Author network 2007 (degree centrality)

Figure 4c: Author network 2008 (degree centrality)

Figure 4d: Author network 2009 (degree centrality)

Figure 4e: Author network 2010 (degree centrality)
In order to provide a more detailed explanation of the ideation process we further zoom in on the specific topics of the semantic networks. Figures 5a-e show the topic networks over time based on degree centrality. Nodes represent title words, ties are indicated by lines. The size of nodes represents topics’ degree centrality. In order to retain clear graphs, we chose to fix a cutoff point of >4 links, that means that the topics represented in these graphs have to have been present at least 4 times in the semantic network. We also remove both the topic terms semantic and web from these graphs. These two topic terms were reflected consistently in title words over the years, but do not reflect new knowledge.

A study of these graphs shows that the topics do change over time, despite the earlier indicated stability in the structure of the semantic network. Whereas in 2006 ontology and OWL were most central in the topic network, this changed to data and linked in 2010. The graphs also show that the clustering of knowledge topics changed: 2006 and 2007 were mostly technically oriented, that is ontology, system, OWL and RDF are central. 2008 functions as a pivotal year, with a relatively dispersed network that does not feature any distinct clusters. In 2009 and 2010, however, we see two clusters emerging. On the upper left hand of the 2009 graph, the cluster evolves around ontology, RDF, mapping and linked. The other cluster centers on knowledge, data, and query. In 2010, this distinction becomes more pronounced: the left side of the graph is dominated by ontology, query, source and RDF,
whereas the right side features knowledge, linked, and data.

To illustrate these dynamics, we traced these topics that we mentioned in the previous paragraph over time: OWL, data, linked, system, RDF, knowledge, query and source. Figure 6 shows how their degree centrality changes over time. We see that system and ontology decrease (the solid lines), and the group of data, knowledge, linked, and source increases (the dashed lines). OWL and RDF stay relatively stable (the dotted lines).

Figure 6: Change in degree centrality of particular topics

5. Discussion and conclusion

The aim of this paper was to explore how semantic and social networks emerge over time, fostering ideation. Our findings indicate that in an emerging field, new knowledge emerges from a specific social network. In particular, the author networks remain sparse but transitive, with mean degree centrality changing over time. Investigation of the semantic networks shows that the initial focus on technically oriented topics changed toward a dual focus on technically oriented and more application oriented topics, with 2008 acting as a pivotal year. Further inspection of a number of illustrative topics corroborates this finding, showing that the topics change in-line with shifts in the centrality of actors in the social network.

Our findings suggest that ideation - the process of generation of new knowledge - is aided namely by the ‘slack’ in the author networks. By slack we mean the leeway or elasticity that is provided in a network. In this case, new knowledge topics are organized in a social network that features low density, high transitivity and shifting degree centrality over time. Consequently, this leads to networks which facilitate the organization of new knowledge topics. This is in line with literature arguing that a balance is necessary between too dense and transitive networks and achieving innovation [16, 26]. However, we want to stress that this consensus is temporal: it changed from an initial consensus in 2006 to a different one in 2010. As such, we argue that consensus should indeed be regarded as temporal. Striving for too much consensus might inhibit future ideation and innovation.

Our results show that in a network that has enough ‘slack’ it is possible for new topics to emerge within a relatively short time span. The networks apparently provide for just enough slack to foster a fruitful exchange of knowledge in order for a new topic cluster to emerge [15, 17], allowing a temporary form of consensus between clusters [12]. We propose that future research on ideation and especially the process of consensus-reaching in ideation needs to take into account its temporal aspects.

This study is not without limitations. We aimed to explore how networks emerge over time, and therefore did not aim for testing causality. However, future research might take up this challenge and establish the interrelatedness of both semantic and social networks: does a slack network lead to the emergence of new knowledge topic clusters, or is it the other way around? Additionally, this would allow researchers to address questions of network inertia. In order to take up this challenge, tools measuring bi-directional influence [39-40] need to be further developed. Along these lines, the obvious affordances of automation of co-word occurrence over the manual coding to identify key topics in communities should be weighted further by researchers of semantic network methods, to assess differences in validity of methods in analyzing semantic networks.

Furthermore, future work might take into account a sub-set of the broader field of Computer Science, or other fields. As such, we performed a case study with a selection of active authors, which suited our aim to conduct exploratory research. However, future studies might be able to expand this scope and include other fields of science, as well as a larger sample size.

The current study emphasizes that more research is needed to separate fluctuations in ideation. In particular, it might be interesting to investigate differences between topic fluctuations: what would count as revolution (much stronger emergence of a new knowledge), and what would rather be conceptualized as a more incremental kind of ideation. Hence, taking a look at novel scientific innovations and
tracing how their topic and author centrality develops over time might be a future point of departure, following salient events [51] instead of ‘regular’ periods in science.

These results provide a potential new perspective on how networks foster the organization of knowledge topics which requires a combination of consensus on topics and a social network structure which allows slack. It is relevant for understanding any community of actors working in a technical domain from scientists to R&D groups.

6. References

27. McFayden, M., M. Semadeni, and A. Cannella Jr., Value of strong ties to disconnected others:
Examining knowledge creation in biomedicine.


