Ideas with Impact: How Connectivity Shapes Idea Diffusion

Dirk Deichmann
Erasmus University
Burgemeester Oudlaan 50, 3062 PA Rotterdam, The Netherlands
ddeichmann@rsm.nl

Christine Moser
VU Amsterdam
De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
c.moser@vu.nl

Julie M. Birkholz
Ghent University
Blandijnberg 2, 9000 Gent, Belgium
julie.birkholz@ugent.be

Adina Nerghes
Koninklijke Nederlandse Akademie van Wetenschappen
Oudezijds Achterburgwal 185, 1012 DK Amsterdam, The Netherlands
adina.nerghes@dh.huc.knaw.nl

Peter Groenewegen
VU Amsterdam
De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
pgroenew@gmail.com

Shenghui Wang
OCLC Research Leiden
Schipholweg 99, 2316 XA Leiden, The Netherlands
shenghui.wang@oclc.org

Ideas with Impact: How Connectivity Shapes Idea Diffusion

ABSTRACT

Despite a growing body of research on idea diffusion, there is a lack of knowledge on why some ideas successfully diffuse and stand out from the crowd while others do not surface or remain unnoticed. We address this question by looking into the characteristics of an idea, specifically its connectivity in a content network. In a content network, ideas connect to other ideas through their content—the words that the ideas have in common. We hypothesize that a high connectivity of an idea in a content network is beneficial for idea diffusion because this idea will more likely be conceived as novel yet at the same time also as more useful because it appears as more familiar to the audience. Moreover, we posit that a high social connectivity of the team working on the idea further enhances the effect of high content connectivity on idea diffusion. Our study focuses on academic conference publications and the co-authorship data of a community of computer science researchers from 2006 to 2012. We find confirmation for our hypotheses and discuss the implications of these findings.

Keywords: content network, social network, idea diffusion, research collaboration, scientific publication, citations

Ideas with Impact: How Connectivity Shapes Idea Diffusion

1. Introduction

Every day, new knowledge is being created and published in the form of scientific articles. When knowledge is steadily increasing, it becomes difficult for new ideas to gain traction and to get noticed (Katz, Levin, & Hamilton, 1963; Rogers, 2003). To navigate this complex and overloaded market of ideas, scientists often consider the citations that a publication receives. Independent of the alleged quality of a scholarly idea, citations indicate the recognition that it received in the past, evidencing the extent to which the idea diffused successfully in the community (Radicchi, Fortunato, & Castellano, 2008; Uzzi, Mukherjee, Stringer, & Jones, 2013; Wang, 2016). Research has shown that the network position of idea generators and their teams critically influences whether or not an idea diffuses successfully (Abrahamson & Rosenkopf, 1997; McFadyen & Cannella, 2004; Wang, 2016; Wejnert, 2002). For instance, a central network position is often associated with the possibility to access more resources and new knowledge (Perry-Smith, 2006; Tsai, 2001). This access provides teams with a higher awareness of what is going on in the field and thus might help the team to develop potentially more relevant ideas. A central network position could also be a signal of a team's status which could be leveraged to exert influence on others (Podolny, 2001).

While our knowledge about idea generators and their social network position has continuously increased, theorizing about the role of ideas has taken a backseat in recent years (Goldenberg, Lehmann, & Mazursky, 2001; Litchfield, Gilson, & Gilson, 2015). We suggest that it is necessary to not only focus on the social network, but also on the characteristics and the network of the idea itself (Goldenberg et al., 2001). We do so by studying how an idea's diffusion is shaped by its connectivity in a so-called "content network." A content network represents the semantic

connections between ideas. In this study, we specifically focus on how ideas are connected to other ideas through common title words. Similar to earlier studies on relatedness in the context of regional knowledge production (Boschma, Heimeriks, & Balland, 2014; Heimeriks & Balland, 2016) our approach is based on the analysis of co-occurrences of words. We adapt this approach and construct a connectivity measure for each scientific paper by examining the title words that different papers share with each other. Making these connections visible allows us to shed light on the organization of knowledge and how highly connected scientific ideas intermediate between different knowledge domains (Mannucci & Yong, 2018).

Ideas which have a high connectivity in the content network can indeed be considered as important "bridges" (Gloor, Krauss, Nann, Fischbach, & Schoder, 2009; Van der Hulst, 2009) and we therefore argue that they trigger more attention than ideas that are connected with fewer knowledge domains. When ideas have a high connectivity, they often combine and blend different knowledge into something novel (Hargadon & Sutton, 1997; Harvey, 2014). At the same time, individuals from different knowledge domains can relate more easily to at least certain aspects of such an idea and thus experience a higher familiarity, which often makes that particular idea more useful to them (Berg, 2016; Dailey & Mumford, 2006). We therefore argue that connectivity in a content network indicates novelty and familiarity—both of which are critical features of an idea and facilitate its diffusion process in the marketplace for ideas.

In addition to examining how the connectivity of an idea influences idea diffusion, we suggest that this effect is further enhanced by the social network of people and their teams. Bridging research on the influence of content connectivity on idea diffusion (Nerghes, Lee, Groenewegen, & Hellsten, 2015) and the influence of social network connectivity on idea diffusion (Abrahamson & Rosenkopf, 1997; McFadyen & Cannella, 2004; Wang, 2016; Wejnert, 2002), we show that the joint

effect of content and social network connectivity is highly beneficial for idea diffusion. This research thus contributes to our understanding of how an idea's connectivity across knowledge domains in a content network influences its diffusion and how social network connectivity of people and their teams can further assist in pushing forward an idea.

While others have studied, for example, how the relatedness of scientific topics shapes knowledge dynamics or specialization patterns on a regional level (Boschma et al., 2014; Heimeriks & Balland, 2016; Hidalgo et al., 2018), in our study, we focus on the diffusion of scholarly ideas (Horowitz, 2017). Prior work has studied idea diffusion in different ways, for example through looking at innovation adoption (Battisti & Iona, 2009) or sales (Garber, Goldenberg, Libai, & Muller, 2004). While recognizing the value of such approaches, we operationalize idea diffusion as the acceptance of an idea by groups of scientists and thus measure the total citation score of the scientific publication of an author team (Radicchi et al., 2008; Uzzi et al., 2013; Wang, 2016). Indeed, citations are often used to measure diffusion success—not only of scientific publications but also, for example, of patents (Balconi, Breschi, & Lissoni, 2004; Fischer & Henkel, 2012; Magerman, Van Looy, & Debackere, 2015; Sorenson & Fleming, 2004).

We investigate the Semantic Web research community, a sub-field of computer science, which develops structures and processes to organize, access, and share information on the Web. We use publication data from the Semantic Web Conference Corpus and complement this data with citation scores from Microsoft Academic Search. Our findings support our hypotheses that ideas that connect multiple knowledge domains are more successfully diffused, and that this effect is further enhanced when the author team is highly connected in the social network.

2. Theory and hypothesis development

2.1. Scholarly ideas

Ideas are conceptions in the mind; they are products of mental activity expressing a "thought or suggestion to a possible course of action" (Oxford English Dictionary, 2000). In this paper, we specifically examine scholarly ideas disseminated through scientific publications. An idea is represented most notably by a publication's title (Boschma et al., 2014). Besides a main idea that is communicated in the title, a publication may of course entail several other ideas, concepts, and contributions. In this paper, however, we entirely focus on the publication's title and thus on the main idea—the overall intellectual line—that is presented in the paper (Leydesdorff, 1989). The title is typically the first notion the reader has of a particular scientific idea. Therefore, the words that constitute the title carry meaning; they direct attention to an idea's structure (Milojević, Sugimoto, Yan, & Ding, 2011). In fact, title words "offer a means of making visible the internal cognitive structure" (Leydesdorff, 1989: 217) of a scholarly publication. This means that the title words represent the essence of a scientific idea. A publication represents a scientific idea that originates from one or several scholars. It is a concrete end-product documenting an idea, preceded by individual and collective activities such as designing the study, collecting data, writing, and revising the manuscript. Through publication, an idea is validated by and exposed to the larger scientific community (Balconi, Breschi, & Lissoni, 2004; Horowitz, 2017; McFadyen & Cannella, 2004). An impactful and, therefore, successful idea will be referred to more often by scholars in subsequent publications (Alvesson & Sandberg, 2013; Wouters, 1999); in other words, it will be cited more often.

2.2. Content connectivity shapes idea diffusion

Ideas and the knowledge that fuels these ideas are organized in knowledge domains (Mannucci & Yong, 2018). When an idea shares certain words with one idea and other words with another idea, it connects two or more knowledge domains (Leydesdorff, 1989). In order to identify ideas that bridge knowledge domains, we take a network perspective and assess an idea's connectivity in a "content network." Content networks have been defined as networks in which words are the nodes and the co-occurrences of those words form the ties between the nodes (Rice & Danowski, 1993). Through co-occurrence of, for example, title words of scientific articles, a content network illustrates how ideas, as represented by their title words, are connected. In this study, we focus on the connectivity of ideas in the larger content network. Ideas are highly connected as part of a content network when they share title words with many other ideas. These shared title words thus link different ideas to each other. We specifically operationalize connectivity by investigating an idea's betweenness centrality in a content network. Betweenness centrality reflects the extent to which an idea sits between all the other ideas present in a network (cf. Everett & Borgatti, 2005). Thus, if a focal idea exhibits a high betweenness centrality, it means that many other existing ideas in the network are connected to each other through the focal idea: the focal idea is on the shortest path between pairs of ideas in the network. We argue that ideas which have a high connectivity in the content network can be considered as important "bridges" (Gloor et al., 2009; Van der Hulst, 2009), and that these bridges facilitate people picking up on an idea. There are two interrelated reasons for why ideas with high connectivity can be considered as attractive candidates for successful diffusion: they signal novelty and at the same time they appear as more familiar and thus useful.

First, ideas which are central in the content network should exhibit a high degree of novelty. Serving as a bridge in the content network (Gloor et al., 2009; Van der Hulst, 2009), these highly connected ideas blend different knowledge domains (Hargadon & Sutton, 1997; Harvey, 2014). By connecting to a variety of knowledge domains, an idea expressed through a publication title will most likely establish new combinations of knowledge and thus signal to an audience that it takes a novel approach on a particular issue. Novelty is essential to scholarship (Wagner, Whetsell, & Mukherjee, 2019) and is even considered an indicator of quality (Trapido, 2015). Indeed, ample research has demonstrated that the novelty of an idea is an important factor which determines whether or not an idea diffuses successfully (e.g., Lee, Walsh, & Wang, 2015; Trapido, 2015; Uzzi et al., 2013; Wang, Veugelers, & Stephan, 2017). In turn, novel combinations of knowledge are more difficult to accomplish when the number of knowledge domains that an idea connects with is limited (Mannucci & Yong, 2018). In such a case, an idea will most likely extend the knowledge in a particular domain in an incremental way but will hardly offer a very original or new approach. Unlike highly connected ideas, those with low connectivity will therefore have a minor impact on the scientific community.

While the first reason for why ideas that have a high degree of connectivity should diffuse more successfully is concerned with the idea itself, the second reason addresses how people, in our case other scholars, may perceive these ideas. Diverse audiences can—at least partially—be familiar with the content of an idea when this idea is connected to different knowledge domains, such that ideas may appear as more familiar and thus are estimated as being more useful. In bridging different knowledge domains, ideas with high content connectivity span the boundaries between otherwise separated fields of expertise (cf. Cross & Prusak, 2002). When an idea connects to multiple knowledge domains, there is in fact great overlap in the vocabulary between the focal

idea and other, existing, ideas. This overlap should facilitate that more people in a particular knowledge domain pay attention to the idea in question because it appears familiar to them (Kaplan & Vakili, 2015). Familiarity also triggers that people evaluate an idea as useful (Berg, 2016; Dailey & Mumford, 2006). Contrary, if ideas have low connectivity, they most probably cover topics that are more specialized and that appear to be less known in other communities (Hill & Carley, 1999; Hooper, Marie, & Kalampokis, 2012; Nerghes et al., 2015). These ideas do not connect with many of the already existing ideas and people therefore have a harder time appreciating the usefulness of these ideas which are more distant to ideas in their own knowledge domain (Fleming, Mingo, & Chen, 2007).

At first sight, it might seem paradoxical that idea connectivity can signal *both* novelty and familiarity, because novelty is often conceived as the opposite of familiarity. That said, several studies have demonstrated that audiences which were tasked to evaluate ideas can indeed perceive ideas as both novel and familiar simultaneously. In fact, ideas which signal both novelty and familiarity often receive the most positive evaluations. For instance, Pieters, Warlop, and Wedel (2002) tested the effectiveness of advertisement originality and familiarity and demonstrated that advertisements which were evaluated as both original (i.e., novel) and familiar drew the most attention to the advertised brand. In a study about citations of scholarly papers, Uzzi and colleagues (2013) found that the most cited papers were those which seemed to combine novel and conventional (i.e., familiar) knowledge. Finally, in a study of patent citation data, Fleming (2001) argued that novel recombinations of familiar sets of technology components are, on average, more successful than novel recombinations of completely new technology components. While completely new component combinations might sometimes lead to an impactful breakthrough invention, they can also spark a higher level of uncertainty. More often than not, they therefore

turn into completely useless patent ideas which will fail to receive many citations. In their recent review on how people receive and evaluate creative ideas, Zhou and colleagues (2019: 2574) additionally concluded that "[h]ighly novel insights can achieve appreciation if they are grounded in strongly familiar knowledge." Thus, prior research provides evidence for our two-tiered argument that highly connected ideas are perceived as novel because they blend existing knowledge in new ways; while, at the same time, highly connected ideas are perceived as at least partially familiar by different audiences from these knowledge domains. Consequently, an idea which is perceived as novel and familiar should diffuse more successfully.

Hypothesis 1. There is a positive association between content connectivity and idea diffusion success.

2.3. Social connectivity enhances the effect of content connectivity on idea diffusion

Previously, we argued that ideas with high content connectivity attract attention because these ideas signal novelty and because they appear as more familiar and thus useful. Novelty and usefulness are also what creativity researchers find important when judging the creativity of an idea (Amabile, 1996; Woodman, Sawyer, & Griffin, 1993). While creativity is often desired, however, it is also associated with costs and risks (for a review, see Khessina, Goncalo, & Krause, 2018). Therefore, people might explicitly or implicitly resist creativity. For example, a creative idea might explicitly be rejected because it calls into question a person's long held assumptions and routines (Deichmann & Van den Ende, 2014). Adopting such an idea would therefore require some level of adjustment which may be perceived as too costly and time intensive (Kahneman, Knetsch, & Thaler, 1991). A creative idea might also be implicitly rejected. For example, research by Mueller and colleagues (2012) uncovered that people often have a covert bias for creativity: While they claim to value creativity, they nevertheless reject the most creative ideas. This

creativity bias is activated when evaluators of ideas experience uncertainty and when they are motivated to reduce this feeling of uncertainty.

We argue that a high social connectivity of the co-authorship team alleviates some of the costs and risks that could be associated with an idea that is highly connected in the content network (cf. Deichmann & Jensen, 2018). When teams are tasked with developing ideas, the position of the team in the larger social space (i.e., connections to other teams) may indeed influence important outcomes. Studies have suggested that the relationship of teams with other teams inside or outside an organization positively increases team performance (Oh, Chung, & Labianca, 2004; Perry-Smith & Shalley, 2014; Reagans, Zuckerman, & McEvily, 2004). Building on this literature, in this study we investigate the position of a co-authorship team in the larger social space. Within this social space, we specifically focus on the social connectivity of the team, operationalized here as a team's betweenness centrality. If a co-authorship team of an idea exhibits a high betweenness centrality it means that many other co-authorship teams in the social space are connected to each other through the focal team.

A high social connectivity should generally be important because highly connected co-authorship teams should have a greater possibility of gaining access to the wider expertise available within the network (Abbasi, Hossain, & Leydesdorff, 2012; Tsai, 2001). This access to expertise provides them with a higher awareness of what is going on in the field, which should help in the subsequent development of a promising idea. Beyond the direct effect that a high social network connectivity can have on the success of the idea's diffusion, we argue that it will also attenuate some of the costs and risks that might be linked to a creative idea (i.e., one with high content connectivity) and therefore influence the relationship between content connectivity and idea diffusion success.

Although we argued that highly connected ideas should generally diffuse more successfully, their perceived high degree of novelty might also trigger explicit or implicit resistance among some people in the target audience (Khessina et al., 2018). Indeed, novelty is often associated with uncertainty (Baer, 2012). Potential feelings of uncertainty about an idea, however, can be reduced when a co-authorship team is highly connected in the scientific community as this indicates that the team is able to produce high-quality work, which likely has spill-over effects for the evaluation of an idea (Zhou, Wang, Bavato, Tasselli, & Wu, 2019). For example, in the context of science, Trapido (2015) found that novel publications were cited more frequently when the authors had a reputation for having been able to produce novel work in the past. A team's connectedness in the social network similarly serves as a quality signal (Podolny, 2001). It highlights that a team has established an important role in the field by being an intermediary between different co-authorship teams (cf. Freeman, 1978). This role not only signals the team's standing in the field but also gives increased credibility to the team's novel ideas and thus is beneficial in diffusing these ideas.

We also argued that idea connectivity is beneficial for diffusion because highly connected ideas invoke a higher degree of familiarity and thus appear as more useful to an audience. A high social connectivity of the co-authorship team should further enhance this effect as it adds another layer of familiarity. In addition to larger and more diverse audiences being familiar with ideas that are connected to different knowledge domains, these audiences can also be familiar with the authors of an idea. This familiarity increases the more the co-authorship team occupies a central position in the community. This double-layered familiarity should not only help to attract more attention from the broader network of scientists (Sorenson & Fleming, 2004), but it should also help people who search the crowded marketplace for ideas for something that they find useful.

Past research showed that social connections play a vital role in the selection of ideas (Hofstetter, Aryobsei, & Herrmann, 2018; Reitzig & Sorenson, 2013). To that end, high social connectivity of a co-authorship team can serve as an additional filter that people might apply when searching for ideas. Research by Piezunka and Dahlander (2015) on how organizations filter and select ideas that external contributors suggest to them provides some initial evidence for this notion. They find that organizations pay less attention to ideas that come from contributors that they do not know—especially when these organizations have to review many different ideas.

To conclude, we argue that the effect of content connectivity on the success of idea diffusion is enhanced when the co-authorship team has a high connectivity in the social network. Social connectivity takes away some of the skepticism which could be linked to the development of new ideas and adds to the idea's familiarity.

Hypothesis 2. The effect of content connectivity on idea diffusion success is moderated by social connectivity; such that an idea with high content connectivity will have a higher likelihood of idea diffusion success when the team has a high, rather than a low, social connectivity.

3. Method

3.1. Sample and setting

In this study, we examine scholarly ideas in the form of scientific publications. Scientific publications document ideas that originate from one or several scholars. Publications codify knowledge and make these ideas explicit. Through the publication of ideas, the ideas become certified in the larger scientific community. We acknowledge that each publication can entail several ideas but we focus here on the main idea—the overall intellectual line—that is communicated in the title of the paper. The successful diffusion of a scholarly idea can be defined

in terms of the number of citations that a specific publication receives. The notion is that influential and impactful ideas will be referred to more often by scholars (Alvesson & Sandberg, 2013; Wouters, 1999).

We test our hypotheses in the setting of the Semantic Web research community. Semantic Web is a sub-field of Computer Science that develops structures and processes to organize, access, and share information on the Web. Techniques developed by the field provide a language (e.g., RDF, OWL, or XML) that expresses both data and rules for reasoning about the data and that allows rules from existing knowledge–representation systems to be exported onto the Web (Berners-Lee, Hendler, & Lassila, 2001). This facilitates interoperation between distributed data allowing metadata on the Web to be integrated.

The Semantic Web research community provides a suitable context to test our hypotheses for two reasons. First, in technically oriented fields such as Computer Science and the sub-field of Semantic Web, ideas are commonly disseminated first in the form of conference proceedings. Unlike, for example, Organizational Studies, where also book chapters and books count toward meaningful knowledge diffusion and conference proceedings are largely absent, in Computer Science, conferences and conference proceedings are an important documentation of scientific ideas. This means that most publications can be queried from conference proceedings databases, thus minimizing selection bias and increasing the likelihood of capturing most publications. Second, publishing in conference proceedings coincides with the speed of knowledge production and a rapid knowledge turnaround (Bar-Ilan, 2010). Again, unlike other fields such as Organizational Studies where it can take many years from initial submission to final publication, the turnaround time in the field of Computer Science can typically be expressed in months. Because publications receive citations relatively soon, we are able to closely observe how content

and social connectivity affect the diffusion of ideas. To conclude, for our purpose of investigating idea diffusion, it is beneficial to draw on data that is bounded in terms of publication outlets (i.e., conference proceedings) and publication speed (i.e., months instead of years).

We used publication records from the Semantic Web Conference Corpus (also known as the Semantic Web "Dog Food Corpus") to calculate content and social connectivity scores (our independent variables). The Dog Food Corpus is a community-maintained digital bibliometric database, with records on conferences, individual publications, and institutions involved in Semantic Web research (Möller, Heath, Handschuh, & Domingue, 2007). We extracted conference proceedings data from the Dog Food Corpus for all conferences related to Semantic Web research between 2006 (inception of the sub-field) and February 2012 (time of data collection), including articles, abstracts, posters, and demos.[1] This resulted in a total of 2,492 publication items with 5,456 unique authors. We used the Microsoft Academic Search (MAS) API to query our data. MAS, Microsoft's academic search engine which retired in 2012, to be renamed Microsoft Academic in 2016, was one of the most valid databases at the time for querying citation scores for Computer Science as it included conference proceedings which are often ignored by other large commercial citation indexes. MAS also provided a title similarity matching, through which we were able to match publications of the Semantic Web Conference Corpus, to ensure valid matching across the two databases. We matched 1,795 publications which have received citations. The remaining 697 publications had not yet received citations and therefore received a score of zero.

---

[1] For simplicity, we will refer to "publications" when we talk about the different items including proceedings, demos, and posters. Besides citing other proceedings, scholars in Computer Science may also cite demos or posters. For robustness, we ran additional analyses where we only focused on conference proceedings in article form. Results are similar to the ones reported in the paper.

3.2. Dependent variable: Idea diffusion success

To measure idea diffusion success, we follow prior studies (Magerman et al., 2015; Uzzi et al., 2013; Wang, 2016) and consider the success of a publication as acknowledged through citations. Citations measure how the publication was received in the scientific community; they show the "acknowledgment of [the idea's] original contributions to science from peers in the field" (Wang, 2016: 72). We queried citation scores per year from MAS.

Idea diffusion success, measured through citations, is a count variable. Due to the skewed nature of this variable, and to facilitate analysis and interpretation of the results, we used the natural logarithm of idea diffusion success. Before log-transforming the variable, we added a value of one to avoid taking the natural logarithm of zero.[2] Furthermore, we used a two-year moving window to see how many citations a particular publication had acquired. This means that we excluded publications from 2011 to 2012 because publications in one of these years would have less allocated time to acquire citations, thus we considered 2,096 publication items. The citation count includes self-citations. Shortly after publication, more self-citations are sometimes observed (Aksnes, 2003). However, we also ran analyses using longer (and even shorter) time windows, and results were similar to the ones we report in the main analyses. Therefore, we have no reason to believe that the inclusion of self-citations biased the results.

3.3. Independent variables

We derived both network connectivity measures from two-mode undirected binary networks created from the publication database. Two-mode networks, also known as affiliation or

---

[2] For robustness, we also ran mixed-effects negative binomial regressions to fit models with the original count variable. By and large, the results of these analyses are consistent with the results reported in this paper although we note that the interaction effect between content connectivity and social connectivity on idea diffusion success drops in significance ($p = 0.10$, two-tailed).

bipartite networks (Borgatti & Everett, 1997), discern two different sets of nodes, as well as the ties that link these different sets (Opsahl, 2013). Researchers often infer that one set plays a greater role in the generation of ties, and thus project the networks as one-mode. This causes information loss (Vernet, Kilduff, & Salter, 2014) and also assumes a primacy of the sets of nodes that we cannot infer (Opsahl, 2013). Instead, a two-mode data structure retains a more accurate representation of our data.

3.3.1. Content connectivity

Content networks have been defined as two-mode undirected and binary semantic networks in which publications are related when the words and concepts in the titles co-occur (Cucchiarelli, D'Antonio, & Velardi, 2012; Leydesdorff, 1989; Rice & Danowski, 1993). Semantic networks project text (in our case publication titles) into networks of concepts and the ties between them, where a concept can be a word or a phrase (Popping, 2003), and a tie a shared affiliation or co-occurrence. In generating the semantic networks, we first automatically identified possible phrases and concepts constituting bi-grams (bi-word expressions such as "big data") and n-grams (multi-word expressions such as "world wide web"). Bi-grams and n-grams were recoded by replacing the spaces between words with an underscore. The resulting list was manually reviewed by an expert to corroborate that the bi- and n-grams captured relevant terms for this specific field. Next, the data was cleaned by replacing punctuation, single letters, numbers, and stop words (e.g., the, a, in, and) with a placeholder to retain the structure of the title. Then stemming was applied (including capitalized words) to relate similar terms (such as plurals: computer and computers) and different verb forms (such as write, writing, wrote). In this process, we identified three entries in a language other than English: one entry in Italian, one entry in French, and one entry partly in English and partly in Spanish. We retained those entries in the dataset without translating them to

English. In addition, we used a spell checker to identify misspelled words and words that had been stemmed improperly (e.g., analysi, crisi, consensu) and created a thesaurus (i.e., dictionary) to identify these words correctly. The cleaning was completed in AutoMap (Carley, Columbus, & Landwehr, 2013) and R (R Development Core Team, 2013) and resulted in a refined dataset that allowed us to identify the co-occurrence of terms between publication titles. See Table 1 for an excerpt of the set with a list of the 50 most frequently used words.

Insert Table 1 about here

In this study, we use the proximity co-occurrence method (Rice & Danowski, 1993) to construct the content network. We generated a two-mode network for each year where publications were connected by shared title words or concepts. More specifically, a link between two words or concepts was created if these words or concepts occurred in the same title. Thus, a publication could be related to another publication multiple times. We analyzed this network by calculating the betweenness centrality score for each publication. Betweenness centrality measures the number of times a node is a bridge between other nodes' shortest paths to two other nodes (Freeman, 1977). Following Borgatti (2012: 2920), the betweenness centrality of node k is expressed as:

$$b_k = \frac{1}{2} \sum_{i \neq k}^{n} \sum_{j \neq k,i}^{n} \frac{g_{ikj}}{g_{ij}}$$

"where g_{ij} is the number of geodesic paths from node i to node j, and g_{ikj} is the number of geodesic paths from i to j that pass through k." Given our data structure, we used a two-mode version of this betweenness centrality measure in Ucinet 6 in which additional steps are taken to account for the two-mode data structure as described by Borgatti and Everett (1997).

18

The content connectivity score reflects the connection of a publication to other publications based on the frequency of shared words or concepts in the title. For our content networks, we therefore presume that a word with a high betweenness centrality has a higher likelihood to serve as a bridge between ideas. See Figure 1a for an exemplary content network visualization. In the example, Publication 1 is connected to Publications 3 and 5. Publication 3 shares the word "improve" and Publication 5 shares the word "structure" with Publication 1. Thus, Publication 1 is the bridge between Publication 3 and 5; therefore, it has a higher connectivity score than the two other publications.

Insert Figures 1a and 1b about here

3.3.2. Social connectivity

Our dataset includes 5,456 individual authors. We generated an undirected, binary, two-mode network for each year, where publications are connected by shared authors. To identify the social connectivity of the team across the different conferences, we used a measure of betweenness centrality in Ucinet 6 which accounts for the two-mode data structure in the way described by Borgatti and Everett (1997). A group of authors whose publication has a high social connectivity score is much more likely to reach, know, communicate, or interact with other authors. This is relative to the number of possible co-authorships allowed per conference regulations (typically the maximum is three). Thus, publications receive a higher social connectivity score the more they are connected to other publications and thus other co-authorship teams. See Figure 1b for an exemplary two-mode social network visualization. In the example, Publications 1 and 3 are linked, because Frank and Lucy are co-authors on both publications.

19

3.4. Controls

We also considered a number of control variables to exclude possible alternative explanations for idea diffusion success. Since all control variables were count variables, we log-transformed them to correct for skewness. First, we controlled for the *number of title words* (after data cleaning) in a given title. We did so because the number of title words may not only influence idea diffusion success (Van Wesel et al., 2014), but may also positively affect the betweenness centrality of a publication in the content network. Second, the makeup of a team influences team performance (Reagans & Zuckerman, 2001). Therefore, we controlled for the *number of authors*. Prior research indicates that the number of authors involved in a publication influences the publication's citations (Didegah & Thelwall, 2013; Van Wesel et al., 2014). Also, having more co-authors may positively influence the betweenness centrality of a co-authorship team.[3] We also controlled for the *average prior citation* count of the co-authorship team.[4] To construct this variable, we matched the publication to its author list. Then, we queried the individual names of the respective co-authors to return the author's complete publication record. We separated these publications by year using an automated script and then examined the author's citations in the years that followed until the year of the focal (team) publication. We then averaged the individual citation scores for the whole co-authorship team. Finally, we controlled for the number of

---

[3] Of all 2492 publications in our dataset, 281 are single-authored. In robustness checks, we excluded these single-authored publications and found similar results.

[4] We also calculated the average number of prior publications of the co-authorship team but this measure correlated strongly with our prior citations variable ($r = 0.78$). Results, however, are very similar whether or not we include prior publications in the models. In addition, we also controlled for the average scientific career age of the co-authorship team. This variable, too, correlated strongly with prior citations ($r = 0.72$). To construct this variable, we identified for each co-author the first noted publication in the MAS dataset and calculated the time in years until the focal publication. We then averaged individual values for the whole co-authorship team. Again, results remain robust, even when we do include this variable into our main analysis.

conferences that co-authors attended in a given year (*average conferences attended*). Here, we averaged again the individual scores across the whole co-authorship team. We included this control because it indicates the visibility of a team which, in turn, could influence the relationship between content and social network connectivity and idea diffusion success.

3.5. Analysis

As we have a nested data structure—several publications are submitted to any given conference—we used models with random intercept. These multilevel models allow us to analyze the data on two levels: publications (level 1) and conferences (level 2).[5] We fitted these multilevel models using the "mixed" command in Stata 15. We standardized both content connectivity and social connectivity before entering them into the regression models.

4. Results

Descriptive results and correlations of all the variables considered in the models can be found in Table 2.

Insert Table 2 about here

The results of our mixed-effects regression analyses are presented in Table 3. Model 1 includes the control variables that may influence the diffusion success of a publication. Results show that the number of authors and prior citations of the co-authorship team have a significant positive effect on idea diffusion success ($\beta = 0.56$, $p < 0.001$ and $\beta = 0.18$, $p < 0.001$, respectively).

---

[5] An alternative to this approach is to use a linear regression analyses and to control for conference fixed effects by including conference identifying dummies. The results of such an analysis are very similar to the results of the mixed-effects linear regression and thus provide additional evidence for the robustness of our findings.

In Models 2 and 3, we separately added content connectivity and social connectivity, and in Model 4 we added both simultaneously. We find confirmation for Hypothesis 1 because the effect of content connectivity on idea diffusion success is significant and positive in Models 2 and 4 ($\beta = 0.13$ $p < 0.001$ and $\beta = 0.13$, $p < 0.001$, respectively). In Model 5 we examine the interaction between content connectivity and social connectivity. The interaction term is significant and positive ($\beta = 0.18$, $p < 0.01$), confirming Hypothesis 2. In Figure 2, we plotted the interaction at high and low values (one standard deviation above and below the mean, respectively) of content connectivity and social connectivity. For this we used the estimates of Model 5; control variables are held constant at their mean value. The figure shows that the highest diffusion success can be attributed to publications with high content connectivity and high social connectivity. As further support for Hypothesis 2, a simple slopes test reveals that the slope between content connectivity and idea diffusion is significant ($\beta = 0.30$, $t = 4.69$, $p < 0.001$) at high levels of social connectivity (one standard deviation above the mean), whereas the slope is not significant ($\beta = -0.07$, $t = -0.94$, $p = 0.35$) at low levels of social connectivity (one standard deviation below the mean).[6]

---

[6] In our theory and analyses, we focus on two-mode betweenness centrality to assess the degree to which a publication is connected in a content network. High betweenness centrality reflects that a publication has many shared title words with other publications that themselves are not connected through shared title words. Such a publication, in other words, serves as a bridge between different topics and ideas. An alternative measure to capture a publication's connectivity is that of "constraint" (Burt, 1992). The advantage of the betweenness centrality measure we use is that it suits our two-mode network data (Borgatti & Everett, 1997) and that it is a more global measure, taking into account the entire network (Perry-Smith, 2006). Neither of these arguments applies to measuring constraint. In additional analyses, however, we also examined the role of constraint in both the content and the social network. To do so, we projected the networks as one-mode and calculated constraint scores for each title word and for each co-author. Next, we averaged these scores with the publication level. While the analysis shows no statistically significant effect of the interaction between content network constraint and social network constraint on idea diffusion, we find confirmation for our first hypothesis. Specifically, the results show that high constraint in the content network is negatively associated with idea diffusion success. This result is in line with our main analysis as it shows that publications that act as a bridge between several knowledge domains (i.e., have low constraint) receive more citations.

Insert Table 3 and Figure 2 about here

In Table 4 we report results of analyses for which we used different moving windows to investigate how many citations a particular publication has acquired (one-, three-, and five-year windows). Using these shorter and longer time windows (compared to the two-year window in our main analysis), by and large support our earlier findings. In particular, we find additional support for Hypothesis 2 across all models. Concerning the direct effect of content connectivity (Hypothesis 1), we observe similar findings in models using longer time windows, but no significant effects in a model in which we only include the citations a publication received in the year after publication. These findings show that, despite the Semantic Web being a highly dynamic community, highly connected ideas need about two years before they will be recognized in the form of citations.

Insert Table 4 about here

----

5. Discussion

In this paper, we studied how the connectivity of an idea in a content network influences idea diffusion. Our analysis of publications in the Semantic Web community shows that ideas which are highly connected in the content network perform better and receive more citations. Furthermore, we studied the moderating role of a team's social connectivity on the relationship between content connectivity and idea diffusion success. Our results show a positive interaction between content and social network connectivity. This means that ideas which bridge different knowledge domains in the

content network will amass even more citations when they are developed by teams that are highly connected in the social network of co-authorship teams.

5.1. Implications

Our findings advance earlier studies that investigated how social network structures and positions foster idea adoption and diffusion (Abrahamson & Rosenkopf, 1997; McFadyen & Cannella, 2004; Wejnert, 2002). We specifically complement these earlier studies by examining the connectivity of an idea in a content network. Scholars increasingly pay attention to how meaning is conveyed in networks (Ferguson, Groenewegen, Moser, Borgatti & Mohr, 2017) and in particular to the language and vocabulary that is used to express ideas (Moser, Groenewegen & Ferguson, 2017; Van Atteveldt, Moser, & Welbers, 2017). Whether an idea is novel or not is believed to be reflected in the description of that idea. For example, Kaplan and Vakili (2015) used a text-based, topic modeling approach to examine novel recombinations of ideas in patents. In this paper, we further extend this line of research with a network perspective. By using content networks to assess the connectivity of an idea relative to the other ideas in our dataset, we were able to uncover not only idea similarity (as we would have by using, for example, topic modeling), but also the underlying structure of ideas and how they bridge different knowledge domains (Cucchiarelli et al., 2012; Leydesdorff & Nerghes, 2015). Compared to other text-analytical methods, content networks are better able to reveal those ideas that bridge established knowledge domains by making use of the inherent meaning structures of ideas—in our case, publication titles.

We theorized that highly connected ideas diffuse more successfully because they are perceived as novel but also as familiar. While other studies have already highlighted that ideas which combine novelty with a dose of familiarity should, on average, be very successful on the marketplace for ideas (Fleming, 2001; Pieters et al., 2002; Uzzi et al., 2013), we further unpack how this balancing

act between the seemingly opposing forces of novelty and familiarity unfolds in practice. When ideas have a high connectivity in a content network, these ideas blend and recombine knowledge from different knowledge domains. As a consequence, such connected ideas also reach a larger and more diverse audience that occupies the different knowledge domains. While the audience in one domain might find a highly connected idea to be very novel as such an idea combines knowledge from another domain, the same audience is also more likely to recognize their own knowledge domain in the idea. This leads to an idea being perceived as—at least partially—familiar. Idea connectivity in a content network therefore enhances idea diffusion because connectivity not only signals novelty but also facilitates that these ideas are recognized as familiar and thus appear as more useful to a target audience.

Finally, looking simultaneously at both the connectivity of an idea in a content network and the connectivity of people in a social network allows us to provide important insights on the multiplicative effects of different network structures on idea diffusion success. Our results indicate that separately examining content and social connectivity only offers part of the explanation for why some idea diffuse more successfully than others. Our findings therefore confirm earlier notions that content and social networks should not be seen in isolation from each other (Mika et al., 2006; Roth & Cointet, 2010; Taramasco et al., 2010). We show that ideas diffuse more successfully when not only the idea is highly connected in the content network, but also when the team responsible for developing the idea is highly connected in the social network. This is because the connectivity of a co-authorship team can serve as a seal of approval for others who need to make a judgment about the idea. To that end, high social connectivity takes away uncertainty around a potentially novel idea and provides another layer of familiarity. This helps people who navigate through a large

marketplace for ideas and who need to decide about whether or not they will cite a particular idea in their own work.

5.2. Limitations and future research

Our study has some limitations and our findings should therefore be generalized with caution. We discuss here the three most important ones. First, we investigated an academic setting in which ideas are communicated through the dissemination of scientific publications. There are many benefits of the academic setting in terms of availability and objectivity of data. However, generalizability of our findings to other contexts might be questionable. Relatedly, in the current study we did not distinguish between regions or cities as loci of knowledge production. However, prior work has shown that space and location matter for scientific work and publications in that space and location influence path dependence and relatedness of knowledge (Boschma et al., 2014; Heimeriks & Balland, 2016; Hidalgo et al., 2018). Future research might therefore investigate how social and content networks relate to each other from a perspective of space.

Second, our study focused on conference publications, but we do not have any information on rejected publications. It is possible that ideas which scored low on content connectivity were not accepted to a conference because they appeared too unrelated to the conference theme. Moreover, co-authorship teams with high social connectivity might more easily be accepted to a conference than co-authorship teams with low social connectivity. Future research might address this issue by either collecting data about the rejected conference items or by studying a different setting in which information is available for both rejected and accepted ideas.

Finally, we propose to study more complete text when calculating the connectivity of an idea in a content network. In this study, we turned to the title words of a publication to operationalize content network connectivity. A title represents the cognitive content of a

publication (Leydesdorff, 1989) and thus should reflect what the publication is about. Examining the abstract of a publication could provide more fine-grained information about the content of the publication. For example, an abstract might not only contain information about the topic of the article but also the methodology used. Future research could then study connectivity of a publication in terms of topic as well as methodology. This could provide a better understanding of the different types of content networks.

5.3. Conclusion

Our study contributes to a better understanding of the idea diffusion process by highlighting the importance of an idea's connectivity in a content network. We show that an idea which links to several knowledge domains is more likely to get diffused. This successful diffusion is grounded on a higher level of perceived novelty and familiarity and can be further strengthened when the team of people involved has an equally high degree of connectivity in the social network.

References

Abbasi, A., Hossain, L., & Leydesdorff, L. 2012. Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks. Journal of Informetrics, 6(3): 403–412.

Abrahamson, E., & Rosenkopf, L. 1997. Social network effects on the extent of innovation diffusion: A computer simulation. Organization Science, 8(3): 289–309.

Aksnes, D. W. 2003. A macro-study of self-citations. Scientometrics, 56(2): 235–246.

Alvesson, M., & Sandberg, J. 2013. Has management studies lost its way? Ideas for more imaginative and innovative research. Journal of Management Studies, 50(1): 128–152.

Amabile, T. M. 1996. Creativity in context. Boulder, CO: Westview Press.

Baer, M. 2012. Putting creativity to work: The implementation of creative ideas in organizations. Academy of Management Journal, 55(5): 1102–1119.

Balconi, M., Breschi, S., & Lissoni, F. 2004. Networks of inventors and the role of academia: An exploration of Italian patent data. Research Policy, 33(1): 127–145.

Bar-Ilan, J. 2010. Web of science with the conference proceedings citation indexes: The case of computer science. Scientometrics, 83(3): 809–824.

Battisti, G., & Iona, A. 2009. The intra-firm diffusion of complementary innovations: Evidence from the adoption of management practices by British establishments. Research Policy, 38(8): 1326–1339.

Berg, J. M. 2016. Balancing on the creative highwire: Forecasting the success of novel ideas in organizations. Administrative Science Quarterly, 61(3): 433–468.

Berners-Lee, T., Hendler, J., & Lassila, O. 2001. The semantic web. Scientific American, 284(5): 34–43.

Borgatti, S. P. 2012. Aocial network analysis, two-mode concepts in. In R. Meyers (Ed.), Computational Complexity: 2912–2924. New York: Springer.

Borgatti, S. P., & Everett, M. G. 1997. Network analysis of 2-mode data. Social Networks, 19(3): 243–269.

Boschma, R., Heimeriks, G., & Balland, P. A. 2014. Scientific knowledge dynamics and relatedness in biotech cities. Research Policy, 43(1): 107–114.

Burt, R. S. 1992. Structural holes. Cambridge, MA: Harvard University Press.

Carley, K. M., Columbus, D., & Landwehr, P. 2013. Automap user's guide 2013. Pittsburgh, PA.

Cross, R., & Prusak, L. 2002. The people who make organizations go--or stop. Harvard Business Review, 80(6): 104–112.

Cucchiarelli, A., D'Antonio, F., & Velardi, P. 2012. Semantically interconnected social networks. Social Network Analysis and Mining, 2(1): 69–95.

Dailey, L., & Mumford, M. D. 2006. Evaluative aspects of creative thought: Errors in appraising the implications of new ideas. Creativity Research Journal, 18(3): 367–384.

Deichmann, D., & Jensen, M. 2018. I can do that alone…or not? How idea generators juggle between the pros and cons of teamwork. Strategic Management Journal, 39(2): 458–475.

Deichmann, D., & Van den Ende, J. 2014. Rising from failure and learning from success: The role of past experience in radical initiative taking. Organization Science, 25(3): 670–690.

Didegah, F., & Thelwall, M. 2013. Which factors help authors produce the highest impact research? Collaboration, journal and document properties. Journal of Informetrics, 7(4): 861–873.

Everett, M., & Borgatti, S. P. 2005. Ego network betweenness. Social Networks, 27(1): 31–38.

Ferguson, J. E., Groenewegen, P., Moser, C., Borgatti, S. P., & Mohr, J. W. 2017. Structure, content, and meaning of organizational networks: Extending network thinking. In P. Groenewegen, J. E. Ferguson, C. Moser, J. W. Mohr, & S. P. Borgatti (Eds.), Research in the Sociology of Organizations: 1–15. Bingley, UK: Emerald.

Fischer, T., & Henkel, J. 2012. Patent trolls on markets for technology - An empirical analysis of NPEs' patent acquisitions. Research Policy, 41(9): 1519–1533.

Fleming, L. 2001. Recombinant uncertainty in technological search. Management Science, 47(1): 117–132.

Fleming, L., Mingo, S., & Chen, D. 2007. Collaborative brokerage, generative creativity, and creative success. Administrative Science Quarterly, 52(3): 443–475.

Freeman, L. C. 1977. A set of measures of centrality based on betweenness. Sociometry, 40(1): 35–41.

Freeman, L. C. 1978. Centrality in social networks conceptual clarification. Social Networks, 1(3): 215–239.

Garber, T., Goldenberg, J., Libai, B., & Muller, E. 2004. From density to destiny: Using spatial dimension of sales data for early prediction of new product success. Marketing Science, 23(3): 419–428.

Gloor, P. A., Krauss, J., Nann, S., Fischbach, K., & Schoder, D. 2009. Web science 2.0: Identifying trends through semantic social network analysis. Proceedings - 12th IEEE International Conference on Computational Science and Engineering, CSE 2009, 4: 215–222.

Goldenberg, J., Lehmann, D. R., & Mazursky, D. 2001. The idea itself and the circumstances of its emergence as predictors of new product success. Management Science, 47(1): 69–84.

Hargadon, A. B., & Sutton, R. I. 1997. Technology brokering and innovation in a product development firm. Administrative Science Quarterly, 42(4): 716–749.

Harvey, S. 2014. Creative synthesis: Exploring the process of extraordinary group creativity. Academy of Management Review, 39(3): 324–343.

Heimeriks, G., & Balland, P. A. 2016. How smart is specialisation? An analysis of specialisation

patterns in knowledge production. Science and Public Policy, 43(4): 562–574.

Hidalgo, C., Balland, P., Boschma, R., Glaeser, E., He, C., et al. 2018. The principle of relatedness. ICCS, 451–457.

Hill, V., & Carley, K. M. 1999. An approach to identifying consensus in a subfield: The case of organizational culture. Poetics, 27(1): 1–30.

Hofstetter, R., Aryobsei, S., & Herrmann, A. 2018. Should you really produce what consumers like online? Empirical evidence for reciprocal voting in open innovation contests. Journal of Product Innovation Management, 35(2):209–229.

Hooper, C. J., Marie, N., & Kalampokis, E. 2012. Dissecting the butterfly: Representation of disciplines publishing at the web science conference series. ACM Web Science 2012: Conference Proceedings, 197–200.

Horowitz, I. L. 2017. Communicating ideas: The politics of scholarly publishing. New York: Routledge.

Kahneman, D., Knetsch, J. L., & Thaler, R. H. 1991. Anomalies: The endowment effect, loss aversion, and status quo bias. Journal of Economic Perspectives, 5(1): 193–206.

Kaplan, S., & Vakili, K. 2015. The double-edged sword of recombination in breakthrough innovation. Strategic Management Journal, 36(10): 1435–1457.

Katz, E., Levin, M., & Hamilton, H. 1963. Traditions of research on the diffusion of innovation. American Sociological Review, 28(2): 237–252.

Khessina, O. M., Goncalo, J. A., & Krause, V. 2018. It's time to sober up: The direct costs, side effects and long-term consequences of creativity and innovation. Research in Organizational Behavior, 38: 107–135.

Lee, Y. N., Walsh, J. P., & Wang, J. 2015. creativity in scientific teams: unpacking novelty and impact. Research Policy, 44(3): 684–697.

Leydesdorff, L. 1989. Words and co-words as indicators of intellectual organization. Research Policy, 18(4): 209–223.

Leydesdorff, L., & Nerghes, A. 2017. Co-word maps and topic modeling: A comparison using small and medium-sized corpora (N < 1,000). Journal of the Association for Information Science and Technology, 68(4):1024–1035.

Litchfield, R. C., Gilson, L. L., & Gilson, P. W. 2015. Defining creative ideas: Toward a more nuanced approach. Group & Organization Management, 40(2): 238–265.

Magerman, T., Van Looy, B., & Debackere, K. 2015. Does involvement in patenting jeopardize one's academic footprint? An analysis of patent-paper pairs in biotechnology. Research Policy, 44(9): 1702–1713.

Mannucci, P. V., & Yong, K. 2018. The differential impact of knowledge depth and knowledge breadth on creativity over individual careers. Academy of Management Journal, 61(5): 1741–1663.

McFadyen, M. A., & Cannella Jr., A. A. 2004. Social capital and knowledge creation: Diminishing returns of the number and strength of exchange relationships. Academy of Management Journal, 47(5): 735–746.

Mika, P., Elfring, T., & Groenewegen, P. 2006. Application of semantic technology for social network analysis in the sciences. Scientometrics, 68(1): 3–27.

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.

Möller, K., Heath, T., Handschuh, S., & Domingue, J. 2007. Recipes for Semantic Web Dog Food: the ESWC and ISWC metadata projects. In K. Aberer, K.-S. Choi, N. Noy, D. Allemang, K.-I. Lee, et al. (Eds.), The Semantic Web: 802–815. Heidelberg, Germany: Springer.

Moser, C., Groenewegen, P., & Ferguson, J. 2017. Meaning in organizational networks - From social to digital and back. In P. Groenewegen, J. E. Ferguson, C. Moser, J. W. Mohr, & S. P. Borgatti (Eds.), Research in the Sociology of Organizations: 211–229. Bingley, UK: Emerald.

Mueller, J. S., Melwani, S., & Goncalo, J. A. 2012. The bias against creativity: Why people desire but reject creative ideas. Psychological Science, 23(1): 13–17.

Nerghes, A., Lee, J.-S., Groenewegen, P., & Hellsten, I. 2015. Mapping discursive dynamics of the financial crisis: A structural perspective of concept roles in semantic networks. Computational Social Networks, 2(15): 1–29.

Oh, H., Chung, M. H. O., & Labianca, G. 2004. Group social capital and group effectiveness: The role of informal socializing ties. Academy of Management Journal, 47(6): 860–875.

Opsahl, T. 2013. Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. Social Networks, 35(2): 159–167.

Oxford English Dictionary. 2000. Ideas. Oxford: United Kingdom: Oxford University Press.

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.

Perry-Smith, J., & Shalley, C. E. 2014. A social composition view of team creativity: The role of member nationality-heterogeneous ties outside of the team. Organization Science, 25(5): 1434–1452.

Pieters, R., Warlop, L., & Wedel, M. 2002. Breaking through the clutter: Benefits of advertisement originality and familiarity for brand attention and memory. Management Science, 48(6): 765–781.

Piezunka, H., & Dahlander, L. 2015. Distant search, narrow attention: how competition for attention alters organizations' filtering of suggestions in crowdsourcing. Academy of Management Journal, 58(3): 856–880.

Podolny, J. M. 2001. Networks as the pipes and prisms of the market. American Journal of Sociology, 107(1): 33–60.

Popping, R. 2003. Knowledge graphs and network text analysis. Social Science Information, 42(1): 91–106.

R Development Core Team. 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing Vienna Austria.

Radicchi, F., Fortunato, S., & Castellano, C. 2008. Universality of citation distributions: Toward an objective measure of scientific impact. Proceedings of the National Academy of Sciences of the United States of America, 105(45): 17268 –17272.

Reagans, R., Zuckerman, E., & McEvily, B. 2004. How to make the team: Social networks vs. demography as criteria for designing effective teams. Administrative Science Quarterly, 49(1): 101–133.

Reagans, R., & Zuckerman, E. W. 2001. Networks, diversity, and productivity: The social capital of corporate R&D teams. Organization Science, 12(4): 502–517.

Reitzig, M., & Sorenson, O. 2013. Biases in the selection stage of bottom-up strategy formulation. Strategic Management Journal, 34(7): 782–799.

Rice, R. E., & Danowski, J. A. 1993. Is it really just like a fancy answering machine? Comparing semantic networks of different types of voice mail users. Journal of Business Communication, 30(4): 369–397.

Rogers, E. M. 2003. Diffusion of innovations (5th ed.). New York: Free Press.

Roth, C., & Cointet, J.-P. 2010. Social and semantic coevolution in knowledge networks. Social Networks, 32(1): 16–29.

Sorenson, O., & Fleming, L. 2004. Science and the diffusion of knowledge. Research Policy, 33(10): 1615–1634.

Taramasco, C., Cointet, J.-P., & Roth, C. 2010. Academic team formation as evolving hypergraphs. Scientometrics, 85(3): 721–740.

Trapido, D. 2015. How novelty in knowledge earns recognition: The role of consistent identities. Research Policy, 44(8): 1488–1500.

Tsai, W. 2001. Knowledge transfer in intraorganizational networks: Effects of network position and absorptive capacity on business unit innovation and performance. Academy of Management Journal, 44(5): 996–1004.

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. 2013. Atypical combinations and scientific impact. Science, 342(6157): 468–72.

Van Atteveldt, W., Moser, C., & Welbers, K. 2017. Being apart together: Convergence and divergence in the field of Dutch politics. In P. Groenewegen, J. E. Ferguson, C. Moser, J. W. Mohr, & S. P. Borgatti (Eds.), Research in the Sociology of Organizations: 49–64. Bingley, UK: Emerald.

Van der Hulst, R. C. 2009. Introduction to social network analysis (SNA) as an investigative tool. Trends in Organized Crime, 12(2): 101–121.

Van Wesel, M., Wyatt, S., Ten Haaf, J., Wesel, M., Wyatt, S., et al. 2014. What a difference a colon makes: How superficial factors influence subsequent citation. Scientometrics, 98(3): 1601–1615.

Vernet, A., Kilduff, M., & Salter, A. 2014. The two-pipe problem: Analysing and theorizing about 2-mode networks. Contemporary Perspectives on Organizational Social Networks, 40: 337–354.

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.

Wang, J. 2016. Knowledge creation in collaboration networks: Effects of tie configuration. Research Policy, 45(1): 68–80.

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.

Wejnert, B. 2002. Integrating models of diffusion of innovation: A conceptual framework. Annual Review of Sociology, 28(1): 297–326.

Woodman, R. W., Sawyer, J. E., & Griffin, R. W. 1993. Toward a theory of organizational creativity. Academy of Management Review, 18(2): 293–321.

Wouters, P. 1999. The citation culture. Ph.D. Thesis. University of Amsterdam.

Zhou, J., Wang, X. M., Bavato, D., Tasselli, S., & Wu, J. 2019. Understanding the receiving side of creativity: A multidisciplinary review and implications for management research. Journal of Management, 45(6): 2570–2595.

Table 1
Word frequency list.

Word	Frequency	Type
ontology	302	single
semantic	261	single
semantic web	188	bi-gram
web	183	single
query	151	single
data	133	single
evaluation	106	single
search	104	single
approach	102	single
model	100	single
corpus	98	single
rdf	96	single
system	95	single
service	91	single
annotation	89	single
network	88	single
resource	88	single
language	83	single
social	79	single
knowledge	72	single
tool	72	single
text	70	single
automatic	68	single
owl	68	single
base	64	single
user	63	single
reason	63	single
learn	59	single
framework	59	single
application	56	single
extraction	56	single
database	54	single
speech	53	single
analysis	49	single
map	49	single
translation	48	single
answer	46	single
tag	46	single
linked data	46	bi-gram
alignment	45	single
sparql	45	single
graph	44	single
online	42	single
process	40	single
technology	40	single
document	39	single
corpora	39	single
build	39	single
relation	39	single
entity	38	single

Table 2
Descriptive statistics and correlation matrix.

Variable	N	Mean	S.D.	Min.	Max.	1	2	3	4	5	6
1. Idea diffusion success[1]	2,096	1.31	1.51	0	6.54						
2. Content connectivity	2,096	0.01	0.01	0	0.18	0.09					
3. Social connectivity	2,096	0.00	0.00	0	0.01	0.05	0.02				
4. Number of title words[1]	2,096	1.85	0.33	0	2.94	-0.02	0.19	0.02			
5. Number of authors[1]	2,096	1.05	0.53	0	2.71	0.28	0.01	0.22	0.08		
6. Average prior citations[1]	2,096	2.80	1.11	0	7.06	0.30	0.04	0.04	-0.02	0.26	
7. Average conferences attended[1]	2,096	0.20	0.36	0	1.79	0.06	0.04	-0.01	-0.08	0.06	0.20

[1] Natural logarithm.

Table 3
Mixed-effects regression of idea diffusion success.[a]

Variables	Idea diffusion success[1]				
	Model 1	Model 2	Model 3	Model 4	Model 5
Constant	-0.17	-0.01	-0.17	-0.01	-0.03
	(0.22)	(0.22)	(0.22)	(0.22)	(0.22)
Number of title words[1]	-0.05	-0.13	-0.05	-0.13	-0.12
	(0.08)	(0.09)	(0.08)	(0.09)	(0.09)
Number of authors[1]	0.56***	0.56***	0.56***	0.56***	0.55***
	(0.06)	(0.05)	(0.06)	(0.06)	(0.06)
Average prior citations[1]	0.18***	0.17***	0.18***	0.17***	0.18***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Average conferences attended[1]	-0.04	-0.02	-0.04	-0.02	-0.01
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
Content connectivity		0.13***		0.13***	0.11***
		(0.03)		(0.03)	(0.03)
Social connectivity			0.01	0.01	0.03
			(0.03)	(0.03)	(0.03)
Content connectivity x					0.18**
Social connectivity					(0.06)
Variance of constant	0.36	0.36	0.36	0.36	0.36
Variance of residual	1.58	1.57	1.58	1.57	1.56
Log likelihood	-3479.72	-3470.86	-3479.63	-3470.78	-3466.19
Publications	2,096	2,096	2,096	2,096	2,096
Conferences	26	26	26	26	26

[a] Standard errors are in parentheses.
+ $p < 0.10$
* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$
Two-tailed tests.
[1] Natural logarithm.

Table 4

Mixed-effects regression of idea diffusion success with different year windows.[a]

Variables	Idea diffusion success[1]: 1-year window		Idea diffusion success[1]: 3-year window		Idea diffusion success[1]: 5-year window	
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Constant	-0.10	-0.13	-0.10	-0.11	-0.26	-0.16
	(0.19)	(0.19)	(0.27)	(0.27)	(0.54)	(0.54)
Number of title words[1]	-0.07	-0.05	-0.06	-0.05	0.05	0.02
	(0.08)	(0.08)	(0.10)	(0.10)	(0.23)	(0.23)
Number of authors[1]	0.42***	0.41***	0.61***	0.60***	0.64***	0.59***
	(0.05)	(0.05)	(0.07)	(0.07)	(0.17)	(0.17)
Average prior citations[1]	0.12***	0.13***	0.24***	0.24***	0.47***	0.48***
	(0.02)	(0.02)	(0.03)	(0.03)	(0.08)	(0.07)
Average conferences attended[1]	0.04	0.05	0.02	0.02	0.62	0.62
	(0.08)	(0.08)	(0.10)	(0.10)	(0.47)	(0.47)
Content connectivity	-0.02	-0.03	0.17***	0.15***	0.20***	0.16***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.05)	(0.05)
Social connectivity	0.01	0.07*	-0.00	0.03	-0.08+	-0.06
	(0.03)	(0.03)	(0.03)	(0.03)	(0.05)	(0.05)
Content connectivity x		0.26***		0.17**		0.21*
Social connectivity		(0.05)		(0.06)		(0.09)
Variance of constant	0.16	0.16	0.54	0.54	0.53	0.52
Variance of residual	1.44	1.43	1.73	1.72	2.51	2.47
Log likelihood	-3767.57	-3756.46	-2944.50	-2940.72	-871.98	-868.89
Publications	2,341	2,341	1,726	1,726	460	460
Conferences	26	26	23	23	9	9

[a] Standard errors are in parentheses.
+ $p < 0.10$
* $p < 0.05$
** $p < 0.01$
*** $p < 0.001$
Two-tailed tests.
[1] Natural logarithm.

Figure 1a
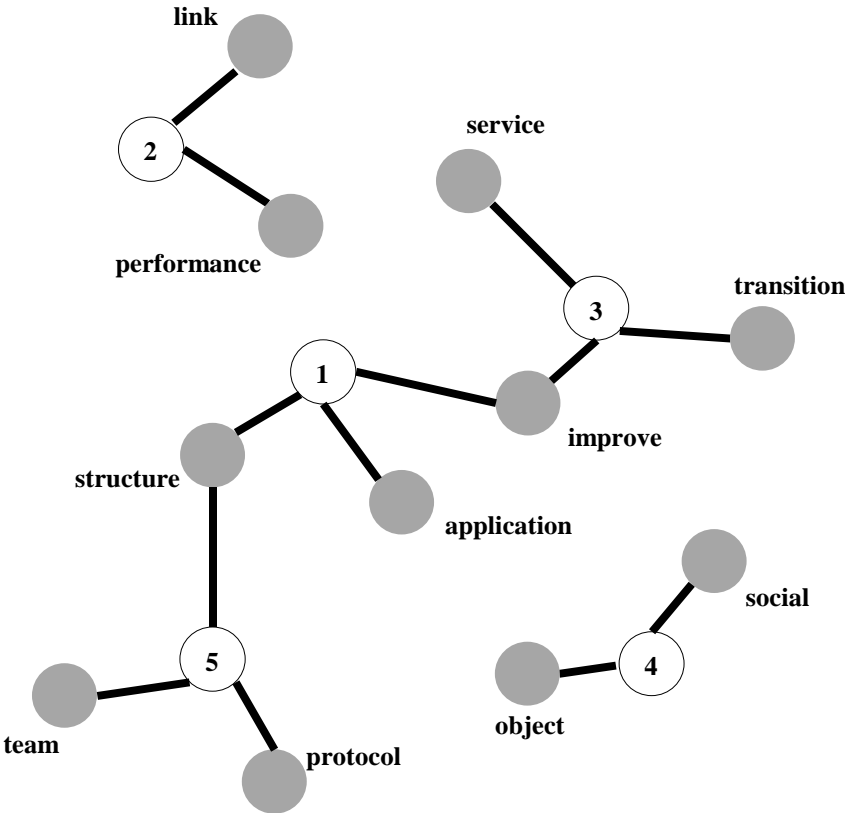Exemplary content network visualization.

Figure 1b
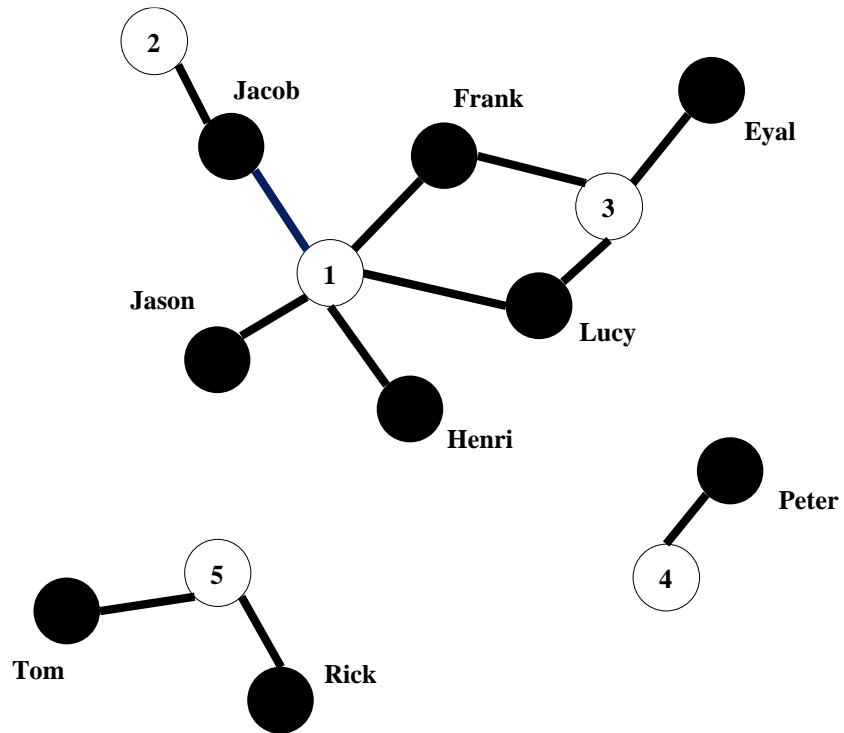Exemplary social network visualization.

Figure 2

Interaction of content connectivity and social connectivity on idea diffusion success.