Understanding the Emergence and Recombination of Distant Knowledge on Crowdsourcing Platforms

Completed Research Paper

Marcel Rhyn[1]
marcel.rhyn@unisg.ch

Ivo Blohm[1]
ivo.blohm@unisg.ch

Jan Marco Leimeister[1,2]
janmarco.leimeister@unisg.ch

[1]University of St. Gallen
Institute of Information Management
Müller-Friedberg-Strasse 8, CH-9000 St. Gallen, Switzerland

[2]University of Kassel
Research Center for Information Systems Design
Pfannkuchstr. 1, D-34121 Kassel, Germany

Abstract

Crowdsourcing represents a powerful approach for organizations to engage in distant search and mobilize knowledge distributed amongst a diverse network of people. While organizations generally succeed in generating large amounts of knowledge, they frequently fail to identify useful ideas that have the potential to solve problems or serve as innovation. We combine text mining and network analysis to examine how such contributions emerge on crowdsourcing platforms and how organizations may identify them. We find that useful ideas typically originate from members in a crowd with only few network ties and that these contributions become especially useful when they are enriched with local knowledge provided by experienced members on the platform. We extend existing research by examining the effects of network relationships and knowledge (re)combination in crowdsourcing. We also discuss the potential of network analysis and text mining to support organizations in tracking the origin of contributions and analyzing their content.

Keywords: Crowdsourcing, Local Search, Distant Search, Innovation

Introduction

Over the past decades, crowdsourcing has attracted much attention for its competitive advantages over traditional work structures in mobilizing distributed workforce and leveraging innovation (Thuan et al. 2016). In crowdsourcing, an organization uses an open call to outsource tasks that have previously been performed by dedicated employees or contractors to an independent network of people. Compared to traditional sourcing mechanisms that rely on only few designated agents, crowdsourcing deliberately seeks to harness the collective knowledge or creativity of the masses (Schenk and Guittard 2011). It allows organizations to move away from predefined routines and facilitates the search for distant knowledge outside their existing boundaries (Afuah and Tucci 2012). In consequence, crowdsourcing is currently being applied in a variety of different domains, including innovation management (e.g., Leimeister et al. 2009; Poetz and Schreier 2012), software development (e.g., Leicht et al. 2017; Stol et al. 2017), and the humanitarian aid sector (e.g., Barbier et al. 2012; Rogstadius et al. 2013).

While crowdsourcing offers the potential to search for distant knowledge in a very efficient and effective manner (Afuah and Tucci 2012), the quantity and complexity of information impinging on organizations is exceptionally high with this approach. IBM, for example, faced more than 46’000 ideas submitted by 150’000 contributors during its Innovation Jam (Bjelland and Wood 2008). Similarly, Dell’s innovation platform IdeaStorm has yielded more than 26’000 ideas with over 100’000 comments since its inception (Dell 2017). Given the limited ability of organizations to process information, they frequently fail to harness the full potential of crowdsourcing when searching for new and useful knowledge in such large pools of contributions. More specifically, Blohm et al. (2013) emphasize that crowdsourcing typically yields a large number of contributions that only have limited value for organizations and that the search for useful suggestions often represents a great challenge on crowdsourcing platforms. Similarly, Piezunka and Dahlander (2015) analyzed how 922 organizations responded to contributions submitted by crowds and found that organizations often miss valuable information because they are exposed to an overload of worthless information. They argue that organizations in crowdsourcing may “succeed in generating a particularly large amount of new knowledge, but that they fail to pay attention to the knowledge that has the most potential for innovation” (p. 875).

In existing literature, little is known about the emergence and evolution of new knowledge on crowdsourcing platforms and how organizations may identify contributions that capture such knowledge. Instead, much research has focused on well-approved contributions generated by experienced lead users in a crowd. Li et al. (2016) found that popular ideas submitted by contributors with prior experience on crowdsourcing platforms are more likely to be implemented than less popular ideas submitted by unknown contributors. Similarly, Schemmann et al. (2016) show that the chance of an idea being implemented by an organization increases when the contributor of the idea has previously examined other crowdsourced ideas and when the idea is popular within the crowd. In earlier studies, Huang et al. (2014) observed that contributors on crowdsourcing platforms learn how to come up with promising ideas over time through increased participation and peer voting. Bayus (2013) provides evidence that serial contributors are more likely to generate an idea that will be implemented than contributors with few ideas. However, popular ideas or solutions generated by experienced members of a crowd may not be the most useful contributions for organizations seeking to span their boundaries and gain new knowledge. Lüthje et al. (2005), for example, show that user-innovators in crowds almost always use local information to determine the need for new solutions and develop them. Thus, in the words of Piezunka and Dahlander (2015), organizations that engage in crowdsourcing face the risk of being lured and lulled by their crowds – “lured into wasting attention on the process of discerning good ideas from bad, and lulled into believing that the ideas expressed most often or most loudly are also the best” (p. 876).

We argue that, in order to find useful contributions in their search for new knowledge on crowdsourcing platforms, organizations must analyze the network structure of their crowds and monitor the topics that are being discussed among its members. We analyze cross-sectional data from a large crowdsourcing platform in Europe and combine statistical approaches from the fields of information retrieval, text mining, and network analysis to answer the following research question: How do useful contributions emerge and evolve on crowdsourcing platforms? We find evidence that such contributions are more likely to originate from members with only few effective network ties in the crowd. These contributions introduce new information and distant perspectives to the crowd, which are then further enriched and combined with local knowledge provided by experienced members on the platform. We argue that organizations searching for new and useful knowledge through crowdsourcing should pay attention to these types of contributions.

With these findings, our contribution is twofold. For research, we provide and extended understanding on how crowdsourcing may be employed for distant search in organizations. We examine the effects of network relationships and knowledge collaboration on the usefulness of crowdsourced contributions and find that simply engaging in crowdsourcing in an attempt to span organizational boundaries may not suffice to find distant knowledge. We show that, when searching for distant knowledge, it is important to differentiate the submitted contributions with regard to their origin in the crowd and their similarity to existing knowledge. We also show that useful contributions not only emerge from isolated contributions alone but from a combination of different topics brought together by members of a crowd. From a practical perspective, our results provide guidance for crowdsourcing intermediaries or organizations that host their own crowdsourcing platforms on how to identify useful contributions in the vast pool of data generated by their crowds. Based on our findings, we urge organization not to rely exclusively on common rating scales or voting systems for assessing crowdsourced contributions. Instead, we see great potential in network analysis and text mining to support organizations in tracking the origin of contributions in crowdsourcing, analyzing their content, and ultimately identifying the most useful ones.

The remainder of this paper is structured as follows: First, we present the theoretical background of our study and review existing research on both search theory and crowdsourcing. Second, we develop a set of hypotheses regarding the emergence and recombination of knowledge on crowdsourcing platform based on prior work in these fields. Third, we explain the methodology for our study by describing the underlying dataset, the variables, and the model that was used to test the hypotheses. Finally, we discuss our findings, outline their implications, and shed light on potential avenues for future research.

Theoretical Background

Local and Distant Search

In organizational theory (e.g., Dosi 1982; Nelson and Winter 1982), search refers to "the controlled and proactive process of attending to, examining, and evaluating new knowledge and information" in order to solve organizational problems or drive innovation (Li et al. 2013, p. 893). Existing research generally distinguishes between two notions of search: local search and distant search (Katila and Ahuja 2002; March 1991). In local search, an organization relates to knowledge that is close to its existing knowledge base and addresses problems by building upon established capabilities and routines (Stuart and Podolny 1996). Research indicates that this is the predominant search strategy used by organizations (Martin and Mitchell 1998; Tripsas and Gavetti 2000). Local solutions are familiar and can be found at relatively low costs or communication efforts (Carlile 2002; Helfat 1994; Rosenkopf and Almeida 2003). This makes local search efficient and reliable for organizations. However, while local knowledge allows for exploitation and facilitates learning, it often suffers from bounded rationalities and lacks the required diversity for effective problem solving (Laursen 2012; Rosenkopf and Nerkar 2001; Rothaermel and Alexandre 2009).

In distant search, organizations move away from predefined routines and reach beyond their boundaries to access unfamiliar knowledge and incorporate new information (Katila and Ahuja 2002). A large body of literature suggests that gaining access to distant knowledge greatly benefits organizations in adapting, diversifying, or reinventing themselves (Katila et al. 2012; Katila and Ahuja 2002). In this sense, distant knowledge has been found to inherit a particularly high potential for developing breakthrough innovation (Fleming 2001; Fleming and Sorenson 2004). The search for distant knowledge can either span technological boundaries or organizational boundaries. Searching beyond organizational boundaries is argued to be especially impactful for exploration (Rosenkopf and Nerkar 2001). There is large body of literature suggesting that the interaction with external sources of knowledge is essential to innovating or solving problems in organizations (e.g., Chesbrough 2003; Von Hippel 2005; Powell et al. 1996).

However, Katila and Ahuja (2002) outline that search efforts not only vary with regard to their scope (i.e., local versus distant search) but also with regard to their depth. Differences in depth of search can lead to varying degrees of familiarity with the acquired knowledge and, in turn, affect the organizations ability to generate new solutions from it (Katila and Ahuja 2002). In this regard, much research emphasizes the role of knowledge recombination (e.g., Fleming 2001; Fleming and Sorenson 2004; Hargadon and Sutton 1997). It is argued that, "by combining firm-specific accumulated understanding of certain knowledge elements (depth) with new solutions (scope), firms are more likely to create new, unique combinations that can be commercialized" (Katila and Ahuja 2002, p. 1180; Winter 1984).

As outlined by Piezunka and Dahlander (2015), organizations may rely on different means to access distant knowledge and combine it with local knowledge, for example, by hiring employees (e.g., Rosenkopf and Nerkar 2001), by acquiring new organizational units (e.g., Ahuja and Lampert 2001), or by forming alliances (e.g., Stuart and Podolny 1996). In recent years, crowdsourcing has emerged as a powerful, IT-facilitated approach for organization to gain access to distant knowledge by broadcasting tasks or value creation activities to a large and diverse network of people (Afuah and Tucci 2012).

Crowdsourcing

The fundamental principle of crowdsourcing revolves around the use of an open call through which an organization engages an independent network of people and leverages their collective workforce or knowledge in order to resolve a predefined problem or task (Blohm et al. 2013; Zhao and Zhu 2014). While crowdsourcing can be seen as an innovative way of organizing work (e.g., Durward et al. 2016) or engaging with potential customers (e.g., Schulten and Schaefer 2015), it has gained particular interest in search theory as a potential solution to distant search in organizations (e.g., Afuah and Tucci 2012; Piezunka and Dahlander 2015). The approach specifically seeks to mobilize resources distributed amongst a large number of individuals (Schenk and Guittard 2011). Compared to traditional search mechanisms that target only few dedicated employees or contractors, participation in crowdsourcing is generally non-discriminatory (Zogaj et al. 2014) and facilitates the self-selection of potential contributors to a problem (Afuah and Tucci 2012). This is based on the tenet that individuals who are not bound to the current thinking in the field of a particular problem are capable of offering "perspectives and heuristics that are novel and thus useful for generating solutions to these problems" (Jeppesen and Lakhani 2010, p. 1019). While it is more difficult for these individuals to assess the feasibility of a solution or an idea (Poetz and Schreier 2012), existing research provides empirical evidence that individuals distant to a domain are able produce more original and radical ideas than experts in the field (Kristensson et al. 2004; Magnusson 2009). Crowdsourcing allows organizations to gain access to such distant knowledge and collect a high number of diverse solutions from outside their boundaries in a very efficient and effective way (Afuah and Tucci 2012; Chesbrough 2003).

Given the decentralized nature of crowdsourcing, the interaction between the organizations and their crowds generally unfolds on IT-based platforms (Doan et al. 2011). These IT-based platforms represent the interface between organizations seeking to broadcast a task and contributors willing to perform the task. They also serve as focal points in this distributed network at which the contributions of the crowd are submitted, aggregated, and retrieved. In general, literature distinguishes between two types of approaches to crowdsourcing on these platforms: competition-based crowdsourcing and collaboration-based crowdsourcing (Blohm et al. 2013; Zhao and Zhu 2014). Competition-based crowdsourcing seeks to efficiently match organizations facing a particular problem with individuals possessing the relevant knowledge for its resolution (Felin and Zenger 2014). It is especially well suited for technical problems or design projects (Bourreau et al. 2012). However, as outlined by Majchrzak and Malhotra (2013) a problem with an approach to innovation that uses the crowd for the sole purpose gathering isolated contributions "is the lack of collaborative discourse that leads to generative co-creation, a foundational requirement for innovation from diverse sources" (p. 263). Thus, while research on crowdsourcing initially focused on temporary ideas competitions, organizations are increasingly interested in issuing more long-term calls and using collaborative crowdsourcing platforms as an integral part of their search activity – both internally and externally (Schemmann et al. 2016; Zuchowski et al. 2016). Collaborative crowdsourcing focuses on the recombination of knowledge and works best when members of the crowd can share information freely and accumulate or alter ideas (Boudreau and Lakhani 2013). In these collaborative settings with interactions unfolding on the platform, crowds can be regarded as connected networks of people that form around a focal organization to jointly generate new ideas or solutions (Simula and Ahola 2014). Such networks of people may provide organizations with "access to diverse, and otherwise hidden knowledge, while at the same time providing in some circumstances support for rich forms of knowledge exchange" (Felin and Zenger 2014, p. 922). In this way, platforms that foster interaction are especially powerful for sourcing new knowledge as they allow crowds to engage in a discourse and jointly develop alternatives, share ideas, or modify problem observations "to co-create solutions that would not have been suggested if only a single perspective had been represented" (Majchrzak and Malhotra 2013, p. 263). A number of studies have also shown that such co-created ideas in crowdsourcing are generally of higher quality than those autonomously submitted by individuals (e.g., Blohm et al. 2010; Majchrzak and Malhotra 2013).

Although crowdsourcing offers the potential to tap new knowledge provided by diverse networks of people in a very efficient and effective manner (Afuah and Tucci 2012), it represent a latent challenge for organizations to cope with the magnitude and complexity of information that is generated and discussed on these platforms. Organizations often struggle in assessing the large number of contributions on crowdsourcing platforms due to a lack of manpower or systematic evaluation processes (Li et al. 2016; Schulze et al. 2012). Thus, attention cannot be paid in detail to all information provided by the crowd (Ocasio 1997; Sullivan 2010). While much research focuses on popular contributions in crowdsourcing and discusses peer-review mechanisms, rating scales, or voting mechanisms (e.g., Blohm et al. 2016), little attention has been paid to the actual emergence and evolution of new knowledge on crowdsourcing platforms. In current literature, it remains mostly unclear where useful contributions originate, how this type of knowledge is being developed and (re)combined within a crowd, and how organizations may identify contributions that capture useful knowledge in the vast pools of information generated by crowds.

Development of Hypotheses

Organizations can use or host crowdsourcing platforms in an effort to span their organizational boundaries and engage in distant search. The platforms grant them access (i.e., provide an interface) to a large and diverse network of people who are willing to contribute their ideas or solutions to a particular problem. However, when searching for actually useful contributions and new knowledge on these crowdsourcing platforms, it is important to understand how networks of people form around organizations and how they create and share knowledge. Thus, for the development of our hypotheses, we draw upon theoretical findings from two different streams of research. First, we ground our study on prior work in the fields of network theory and problem solving (e.g., Perry-Smith and Shalley 2003). That is, we focus on crowdsourcing settings that allow knowledge to be shared and jointly developed amongst member of a crowd. A number of studies have already employed a network perspective to analyze such connected crowds that form around focal organizations or topics (e.g., Lu et al. 2017; Simula and Ahola 2014; Stephen et al. 2016). This stream of research provides valuable insights on how individuals with different positions in the network (i.e., the crowd) introduce new ideas or problem-solving approaches that may represent distant knowledge for organizations. A second, relevant stream of research is concerned with knowledge collaboration and offers insights on how crowds discuss and enrich these initial ideas or solutions to make them useful for organizations (e.g., Faraj et al. 2011).

First, we refer to research in the field of problem-solving and network theory. Existing literature in this field generally suggests that individuals have different perspectives on problems and employ different heuristics to derive solutions based on their prior experiences and their domain of expertise. This is based on findings that human problem solving involves the construction of an internal representation of the problem and the application of an appropriate heuristic to search for a potential solution (Dunbar 1998; Jeppesen and Lakhani 2010). It has been found that individuals overwhelmingly use familiar knowledge and prior experience in developing solutions to problems they encounter (Lovett and Anderson 1996; Lüthje et al. 2005). Although expertise in a domain and familiarity with existing knowledge may be useful, the most innovative or useful solutions may not necessarily be brought up by individuals affiliated with the actual domain of the problem. As shown by Jeppesen and Lakhani (2010) individuals that are technically and structurally distant from the problem domain “can offer perspectives and heuristics that are novel and thus useful for generating solutions to these problems” (p. 1019). They are often naïve with regard to the prevailing assumptions or theories in a particular domain (Gieryn and Hirsh 1983) and have access to differing knowledge and perspectives compared to individuals that are local to the domain. Research also shows that psychological distance greatly benefits creativity (Förster et al. 2004; Trope and Liberman 2010). Being remote to a problem and thinking in abstract terms may lead to more diverse and original solutions whereas thinking in concrete, technical terms often impedes innovation (Förster et al. 2004).

These arguments are also supported from a network perspective. Perry-Smith and Shalley (2003) illustrate that mental representations tend to converge in local networks (such as crowds) due to common experiences and an increased sharing of redundant information. As individuals immerse in a particular network, it becomes more difficult for them to see beyond their direct ties which provide mostly conformant information. Similarly, as the ties within a network or domain become stronger, conformity will hamper creativity. In consequence, it is argued that perspectives drawn from distant positions in a network will likely be more novel relative to existing standards within the domain (Perry-Smith and Shalley 2003).

Following this stream of research, we hypothesize that individuals with few network ties who are not already immersed in the crowd are more likely to provide new paradigms or problem-solving approaches to the platform than experienced individuals with many network ties in a crowd. Ultimately, these contributions should be more useful for organizations that search for new and distant knowledge through crowdsourcing. Furthermore, in line with the previous argument, we expect contributions that offer novel information to be more useful for organizations than contributions that refer to already available information.

H1. *Contributions that are created by members of a crowd with few network ties are more likely to be useful for organizations than contributions created by members of a crowd with many network ties.*

H2. *Contributions that contain novel information are more likely to be useful for organizations than contributions that refer to already available information.*

A second stream of research that provides valuable insights into the development of knowledge in crowdsourcing revolves around knowledge collaboration (Faraj et al. 2011). Existing literature in this field suggests that knowledge usually emerges not from a single contribution alone but from a recombination, modification, and integration of knowledge provided by different individuals in a network (Ye et al. 2016). Given that individuals employ different perspectives and problem-solving heuristics, information to develop innovations or solve problems have generally been found to be widely distributed among many people rather than concentrated among only few prolific individuals (Von Hippel 2005). When perspectives from distinct fields are brought together and combined, the proposed solutions or ideas are argued to have a high potential to be novel and deviate from established mindsets (Perry-Smith and Shalley 2003). Mumford and Gustafson (1988) in particular suggest that high levels of creativity usually emerge from very different schemata or cognitive structures being combined. Even in design science, it is argued that innovation and new knowledge are typically developed from an iterative expansion or revision of existing concepts and knowledge during the design process (e.g., Braha and Reich 2003; Hatchuel and Weil 2009).

A large number of studies indicate that collectively recombining knowledge is essential for developing innovative ideas or problem-solving approaches (e.g., Faraj et al. 2011; Ye et al. 2016). Recombination enables crowds to further enrich and develop ideas or solutions and aggregate the content for a more in-depth and comprehensive understanding. Furthermore, crowdsourcing greatly benefits from different topics or perspectives coming together on the platforms. Exposure to different alternatives or new perspectives has been found to trigger a process of using wider categorizations and generating more divergent solutions (Kanter 1988; Perry-Smith and Shalley 2003). Especially for tasks that revolve around the development of alternative ideas, access to diverse knowledge and perspectives lead to greater quantities of non-redundant solutions (Chatman et al. 2007; De Dreu and West 2001; Riedl and Woolley 2016). In crowdsourcing specifically, Riedl and Woolley (2016) observed that crowds coming up with useful solutions draw on diverse sets of topics and exhibit less redundancy regarding the exchanged information. The diversity of information in their discussions has been found to be much higher. Results presented in a related study conducted by Bayus (2013) points in a similar direction, showing that an individual's likelihood of generating promising ideas in crowdsourcing is positively affected by the diversity of his or her commenting activity on other ideas. In general, these findings suggest that useful knowledge in crowdsourcing may especially emerge from collective contributions provided by individual members of a crowd that collaborated and became creative (Geiger and Schader 2014; Ye et al. 2016).

These insights resonate strongly with the previously discussed (re)combination of knowledge in search theory (Fleming 2001). As assumed in H1 and H2, members with few network ties in the crowd are likely to introduce novel information, perspectives, or problem-solving approaches. Members who have established many network ties and who are immersed in the network may combine or recombine this new information with local knowledge that has already matured on the platform. They add more in-depth information to the initial contribution and combine it with different topics or insights that have already been accumulated. Thus, we assume that a contribution becomes more useful for organizations that search for new knowledge when the crowd combines it with novel information and adds depth to it. We also argue that contributions that offer a broader scope of perspectives and draw upon a diverse set of topics on the crowdsourcing platform are more likely to be useful for organizations than contributions that draw upon a less diverse set of perspectives.

H3. *Contributions whose discussion adds novel information are more likely to be useful for organizations than contributions whose discussion refers to available information.*

H4. *Contributions that combine information from a broad set of topics on the platform are more likely to be useful for organizations than contributions that refer to a narrow set of topics.*

Figure 1 depicts the underlying research model of this study. For organizations that search for new knowledge on crowdsourcing platforms, we expect the number of network ties a contributor has established in the crowd (H1) to be negatively related to the usefulness of a contribution and the novelty of the contribution (H2) to be positively related to the usefulness of a contribution. Furthermore, we assume that the usefulness of a contribution is positively affected when novel information is added by other members of the crowd (H3) and when a broad set of topics are combined (H4). We also control for alternative effects that are explained in more detail in the subsequent section below.

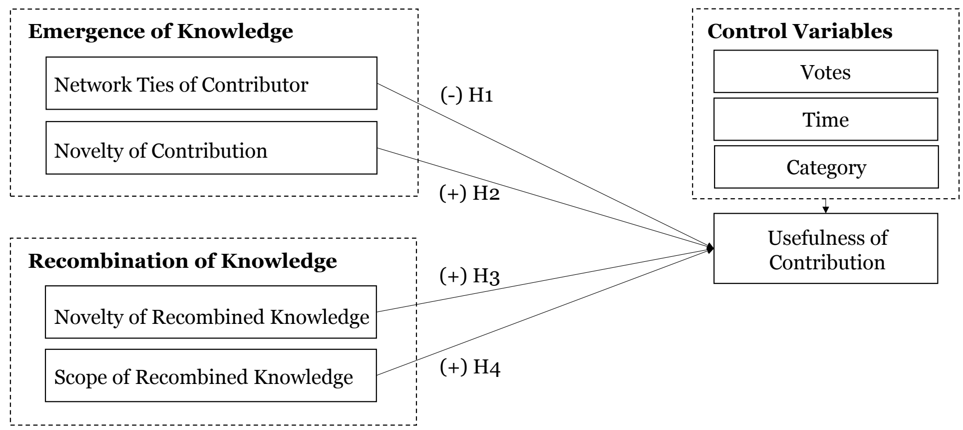

Figure 1. Research Model

Methodology

In order to address our research question and empirically test the outlined hypotheses, we combine statistical approaches from the fields of information retrieval, text mining, and network analysis to examine cross-sectional data retrieved from a crowdsourcing platform of a large Swiss organization. All calculations were performed with the R Language for Statistical Computing. For processing the textual data and analyzing their content, we employed algorithms provided by the *tm, openNLP,* and *topicmodels* packages. The network structure of the crowd was mapped and analyzed using the *igraph* package.

Dataset

For our study, we use a unique data set retrieved from a German-speaking crowdsourcing platform. The platform is operated by a Swiss transportation and logistics organization. Most notably, it revolves around web services and smartphone applications developed by the organization that include, for example, a ticket purchasing system, times schedules, or itineraries. On the platform, members of the crowd may either contribute their own ideas or comment and vote on the contributions of others. In addition, the organization frequently issues open calls on the platform asking the crowd for their feedback or ideas on new product releases. In this way, the platform aims to foster a collaborative discourse amongst its users and ensure interaction in the crowd. Thus, the platform hosts a rather stable, slowly growing crowd and does not rely on temporary teams or contests with time constraints. The ideas on the platform are continuously reviewed and evaluated by a team of 29 administrators. The administrators work for the organization and are responsible for identifying the most useful contributions for implementation.

As of August 2016, the platform has a total of 11'408 registered members in the crowd. While it is hosted by the organization and predominantly appeals to its customers, there are no prerequisites for joining the crowd. Hence, it represents an open crowd (cf. Corney et al. 2010; Zwass 2010) that includes potentially everyone who's interested and willing to contribute ideas or comments to the organization. 1700 (or 14.9%) of all registered members are active and made at least one contribution (i.e., idea, solution, or comment). On average, active members of the crowd made 2.86 (SD: 13.5) contributions on the platform. They receive no monetary compensation for their efforts. However, the organization occasionally awards the most active and valuable contributors in the crowd by inviting them to sponsored events or dinners.

With regard to the contributions, the original dataset retrieved from the platform contained a total of 2'304 ideas and 2'564 comments that discuss them. In preparation for the analysis, we cleaned the dataset. First, we removed all ideas and comments that were not written in German (i.e., only 2.3% of all contributions).

Since we are analyzing the content of the contributions with text mining algorithms, this step was necessary in order to ensure that the language of our text corpus is uniform and does not confound the results. Second, we excluded all responses by the administrators and moderators for our analysis. This was done because we are only interested in contributions generated by actual members of the crowd and not the organization. The administrators' responses are mostly "standard" responses, for example, to thank the crowd for their contributions. Third, we followed the commonly used bag-of-words approach and applied standard preprocessing steps in order to make the unstructured, user-generated text in our dataset compatible for text mining algorithms (cf. Feldman and Sanger 2007). That is, we tokenized the contributions and broke them up into individual terms. We applied standard transformations to the terms, including normalization (i.e., transforming all characters to lower-case) and stop word filtering (i.e., removing terms such as prepositions or articles that bear no value for our analysis). We also excluded extremely scarce terms (e.g., terms that were misspelled and, thus, treated as a separate, "new", terms). Our final dataset comprises a total of 1'927 ideas or suggestions for improvements with 1'859 comments. Of these 1'927 ideas or suggestions for improvements, 258 contributions have been implemented by the organization. The dataset has been retrieved in August 2016 and contains all information since the initial launch of the crowdsourcing platform in October 2015.

The dataset is suitable for our analysis for several reasons. First, the crowdsourcing platform was specifically created by the transportation and logistics provider for the purpose of spanning its organizational boundaries and eliciting new ideas or feedback from an independent crowd of users on how to improve its products and services. Thus, the explicit purpose of the platform is to access distant knowledge. Second, as representatives of the organization use the platform themselves to evaluate and select useful contributions, we have reliable expert labels and statuses on which our analysis can be grounded (see subsequent section below). Third, we note that the design of the platform and the characteristics of our dataset are comparable to those reported in related studies (e.g., Li et al. 2016; Schemmann et al. 2016). This should benefit the generalizability of our results.

Variables and Measures

Usefulness of the Contribution. We use the ***decision to implement*** a contribution as our dependent variable. Ultimately, the implementation indicates whether a contribution has been deemed useful by representatives of the organization for solving a problem or serving as innovation. In this way, we follow a large number of related studies that have already used this rationale to address similar research questions in crowdsourcing (e.g., Li et al. 2016; Piezunka and Dahlander 2015; Schemmann et al. 2016) and the notion Levitt (1963) who states that "ideas are useless unless used" (p. 79). The decision to implement an idea is made by representatives of the organization who serve as administrators on the platforms. These representatives are responsible for identifying the most useful contributions submitted by the crowd. They manually review and evaluate the contributions on the platform and assign a status to them depending on their decision. Thus, we have reliable labels on whether a contribution has been deemed useful or not. The status of a contribution is a binary variable (0 = not implemented, 1 = implemented) and has been retrieved from the platform.

Network Ties of Contributor. To measure the network ties of a contributor, we constructed a social network of the crowd on the platform. Each node in the network represents a member of the crowd. An edge or tie (i.e., a relationship) between two members of the crowd was established when two members exchanged information (i.e., when they commented or voted on their ideas or solutions). For simplicity, we constructed a non-directional network. In this way, we followed Hautz et al. (2010) and adopted a weak notion of network relationships. We calculate the ***effective network size*** (Borgatti 1997; Burt 1995) for each member of the crowd. The effective network size measures the number of ties that a member of the crowd has established in the network while discounting redundant ties. Accounting for redundant ties is important in our case as it can affect the type of knowledge that is exchanged between users (i.e., redundant information or diverse information from different network relationships). As each contribution on the platform was created by a member of the crowd, we are able to examine the origin of the contribution in the network. Contributions that are created by members of the crowd with many effective network ties are characterized by a large effective network size while contributions that are created by members of the crowd with only few effective network ties are characterized by a small effective network size.

Novelty of Contribution. In order to measure the novelty of information in a contribution, we process the content of the contribution with text mining algorithms and calculate an aggregated ***TF-IDF-index***. TF-IDF refers to a term weighting scheme in information retrieval (Salton and Buckley 1988) that accounts for the importance of a particular term (i.e., word) in a document (i.e., crowdsourced contribution). It measures the frequency of a term in a document normalized by the document length (TF) and multiplies this value with the inverse document frequency of the term (IDF). Generally speaking, a term that is frequently used in a contribution but rarely used in other contributions on the crowdsourcing platform will receive a high TF-IDF-value. In this way, it is possible to measure the novelty of the words in a contribution. We aggregated the TF-IDF-values for all terms in a contribution using the sum. Thus, contributions with a high TF-IDF-index include more novel information than contributions with a low TF-IDF-index. Related studies have already used aggregated TF-IDF-indices to analyze textual contributions in crowdsourcing (Rhyn and Blohm 2017; Zhang et al. 2016).

Novelty of Recombined Knowledge. In order to measure the novelty of the recombined knowledge, we analyzed the content of the discussions that developed around a contribution. For each contribution, the platform offered the possibility for other members of the crowd to add comments and provide their own perspectives or experiences. Thus, not only the initial contribution provides information for the organization but also the discussion that potentially combines the initial contribution with other perspectives and adds novel insights. Consistent with our previous measure, we calculated the average ***TF-IDF-indices of all comments*** for a contribution in order to measure the novelty of this recombined knowledge. Thus, a high TF-IDF-index suggests that, on average, the comments in the discussion added more novel information to a contribution than comments in a discussion with a low TF-IDF-index.

Scope of Recombined Knowledge. In order to measure the scope of the recombined knowledge, we analyze the topics that are being addressed in the contributions and their discussions. Topics can be interpreted as a distribution of words. A "design" topic, for example, will likely include words referring to colors or shapes and less likely include words referring to cars or trains. A "transportation" topic, on the other hand, will likely include words referring to cars or trains and less likely words referring to colors or shapes. We measure the distribution of the topics in a crowdsourced contribution and compare it to the average topic distribution on the crowdsourcing platform. Thus, contributions whose topic distribution is similar to the average topic distribution on the platform will combine a broad set of topics (i.e., large scope). Contributions whose topic distribution is less similar to the average topic distribution on the platform focus on only one or few specific domains (i.e., narrow scope). In order to calculate this measure, we use topic modeling based on the Latent Dirichlet Allocation (LDA) with a Gibbs sampler (Blei et al. 2003). LDA refers to a generative probabilistic model that can be used automatically detect topics that are underlying a collection of text documents (i.e., contributions on the platform). The process for detecting and analyzing the topics in text documents includes two essential steps. First, we used all contributions and comments created by the crowd to uncover the topics that are present on the crowdsourcing platform. Based on the approaches proposed by Griffiths and Steyvers (2004) and Arun et al. (2010), we found 36 topics on our platform, which are outlined in more detail in the discussion section below. In a second step, it is possible to assign each individual contribution with probabilities for addressing each topic. The distribution of topics can be represented by a vector. We used the ***cosine similarity*** to calculate the similarity between the topic vector of a contribution with its discussion and the mean topic vector on the platform. The cosine similarity has been found to be a valid measure for the similarity of posterior distributions as retrieved in topic modeling (Niekler and Jähnichen 2012). In this vein, the measure is also akin to the concept of information diversity as used in a related study by Riedl and Woolley (2016).

Control Variables. We use several additional variables to control for alternative effects that could influence the likelihood of a contribution being implemented by the organization. First, as shown by related literature (e.g., Li et al. 2016; Schemmann et al. 2016), popular ideas have a higher chance of being implemented by organizations than less popular ideas. Thus, on our platform, the voting behavior of the crowd might have influenced the decision of the administrators to implement a contribution. We account for this effect by including the number of ***votes*** per contribution as a control variable. Second, since we are analyzing cross-sectional data, older contributions have had more time to being discussed and noticed by the crowd or the administrators than contributions that have been submitted just recently. We control for this effect by measuring the ***time*** (in number of days) a contribution has been on the platform. Third, it is possible that a certain type of contribution is deemed more important and thus prioritized by the organization for implementation. On our platform specifically, the crowd was able to submit contributions that address two categories: ideas and problems (e.g., with using the organization's smartphone application or ticketing system). We control for the ***category*** in which the contribution was submitted by using dummy variable (0 = idea, 1 = problem).

The descriptive statistics and the correlation matrix for the variables in our study are listed below in Table 1 and Table 2 respectively. Most importantly, the correlations and the variance inflation factors for our independent variables, which range from 1.010 (for the novelty of the contribution) to 1.220 (for the votes), raise no concerns for multicollinearity.

Table 1. Statistics for the Dependent, Independent, and Control Variables				
	Mean	*Std. Dev.*	*Min.*	*Max.*
Dependent Variable				
Usefulness of Contribution	0.13	0.34	0	1
Independent Variables				
Network Ties of Contributor	5.13	23.71	0	178.75
Novelty of Contribution	2.65	0.88	0.27	6.87
Novelty of Recombined Knowledge	0.27	0.52	0	4.56
Scope of Recombined Knowledge	0.92	0.08	0.28	0.99
Control Variable				
Votes	2.77	9.79	0	218
Time	91.38	84.70	0	274
Category	0.36	0.48	0	1

Table 1. Descriptive Statistics

Table 2. Correlation Matrix								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1) UC	1							
2) NTC	-0.04	1						
3) NC	0.07***	-0.11***	1					
4) NRK	0.07***	0.00	0.00	1				
5) SRK	-0.01	-0.08***	0.04*	-0.25***	1			
6) Votes	0.15***	-0.01	0.04	0.14***	-0.19***	1		
7) Time	-0.29***	-0.03	-0.01	-0.04*	0.06**	-0.21***	1	
8) Category	0.08***	0.01	0.02	0.12***	-0.08***	-0.16***	0.04*	1

Note: ***p<0.01; **p<0.05; *p<0.10; [1] UC = Usefulness of Contribution, NTC = Network Ties of Contributor, NC = Novelty of Contribution, NRK = Novelty of Recombined Knowledge, SRK = Scope of Recombined Knowledge

Table 2. Correlations of Variables

Models and Results

Given that the dependent variable is binary, we use binary logistic regression to analyze the dataset and test our hypotheses. Our full model reads as follows:

$$P_{implemented} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_3 + \beta_4 \cdot x_4 + \beta_5 \cdot x_5 + \beta_6 \cdot x_6 + \beta_7 \cdot x_7 + \varepsilon)}}$$

where β_0 represents the constant term, x_1 represents the effective network ties of the contributor, x_2 represents the novelty of the contribution, x_3 represents the novelty of the recombined knowledge, x_4 represents the scope of the recombined knowledge, x_5 represents the votes, x_6 represents the time, and x_7 represents the category. The β-coefficients can be interpreted as the change in log odds for a one unit change in the variables. The error term is represented by ε. The results of the logistic regression are listed in Table 3. We report the coefficients, the standard errors, and the Wald statistics to assess the significance of the coefficients. The Maximum-Likelihood-Estimation (MLE) method was used to estimate the coefficients. Since some contributions were created by the same person in the crowd, we ran our analysis with clustered robust standard errors to control for potential dependencies. This approach has been widely used by related studies to control for similar effects (e.g., Li et al. 2016; Schemmann et al. 2016).

Table 3. Logistic Regression Model				
		Est.	*S.E.*	*Wald Z*
Intercept		-3.526	1.030	-3.42***
H1:	Network Ties of Contributor	-0.009	0.003	-3.45***
H2:	Novelty of Contribution	0.205	0.082	2.49**
H3:	Novelty of Recombined Knowledge	0.310	0.144	2.15**
H4:	Scope of Recombined Knowledge	1.737	1.042	1.67*
Control:	Votes	0.019	0.012	1.66*
Control:	Time	-0.016	0.002	-9.76***
Control:	Category	0.735	0.162	4.53***
Pseudo R-Sq.[1]		0.228		
Chi-Sq.		255.91*** (df = 7)		

Note: ***p<0.01; **p<0.05; *p<0.10; [1] Nagelkerke (1991)

Table 3. Results of the Logistic Regression

As shown in Table 3, the test of our logistic regression model against the constant-only model is highly statistically significant ($Chi^2(7) = 255.91$, $p < 0.00$). The Nagelkerke's Pseudo R^2 amounts to 0.228 (Nagelkerke 1991). All hypotheses with the exception of H4 are supported at highly significant levels of $p < 0.05$. For H4, we only find support at a confidence level of $p < 0.10$ but not at a confidence level of $p < 0.05$. The coefficients listed in Table 3 can be exponentiated and interpreted as odds-ratios. The following paragraphs discuss these odds-rations and summarize our findings with regard to our hypotheses.

In H1, we hypothesized that contributions that are created by members of a crowd with few network ties are more likely to be useful for organizations than contributions created by members of a crowd with many network ties. Indeed, we find that the odds of a contribution being implemented decreases by 0.90% ($\beta = -0.009$; $p = 0.001$) for every one unit increase in a user's effective network size on the platform. In essence, this suggests that, as contributors become more immersed in the crowd and exchange a lot of information with other users, they become familiar with dominant perspectives on the platform are less likely to come up with new contributions that are useful for organizations. The effect is statistically significant ($p < 0.05$).

In H2, we hypothesized that contributions that contain novel information are more likely to be useful for organizations than contributions that refer to already available information. Our results reveal that, for every one unit increase in the TF-IDF-index, the odds of a contribution being implemented increase by

22.75% ($\beta = 0.205$; $p = 0.013$). This shows that contributions which include a lot of terms that are rarely used in other contributions (i.e., contributions that offer novel information) are more likely to be useful for organizations than contributions that include rather common terms (i.e., contributions that refer to already existing information). The effect is statistically significant as well ($p < 0.05$).

In H3, we hypothesized contributions whose discussion adds novel information are more likely to be useful for organizations than contributions whose discussion refers to available information. We find that, for every one unit increase in the average TF-IDF-index for a discussion, the odds of a contribution being implemented increase by 36.34% ($\beta = 0.310$; $p = 0.031$). This goes to show that not only the initial contributions are important, but also the discussions that develop around them. A contribution becomes more useful for organizations when the crowd combines it with novel insights and further enriches the content. The effect is statistically significant ($p < 0.05$).

In H4, we hypothesized that contributions that combine information from a broad set of topics on the platform are more likely to be useful for organizations than contributions that refer to a narrow set of topics. On a confidence level of $p < 0.10$, it shows that if the distribution of the topics in a contribution perfectly matches the broad distribution of topics on the whole platform (i.e., reach a perfect value of 1), the contribution would be 5.68 times more likely to be implemented than contributions whose topic distribution refers to a more narrow set of topics ($\beta = 1.737$; $p = 0.096$).

The results presented in Table 3 also illustrate the importance of the control variables. Both the control variable for the time effect and the control variable for the category are highly statistically significant. The control variable for the votes is statistically significant at a confidence level of 0.10. Thus, contributions with a higher number of votes are more likely to be implemented than contribution with a lower number of votes. Consistent with prior work (e.g., Li et al. 2016; Schemmann et al. 2016), this indicates that the voting behavior of the crowd influences the decision of organizations to implement ideas. Second, recently submitted contributions are less likely to be implemented by the organization than older contributions. This can be explained by the fact that older contributions have had more time to be discussed, enriched, and selected by the administrators than recently submitted contributions. Furthermore, it takes time to implement contributions. These findings are also consistent with prior studies (e.g., Schemmann et al. 2016). Table 3 also shows that it is important to control for the category of the contribution. Contributions that address problems (e.g., ticket purchasing system not working, crashes of the organization's mobile application) are more likely to be implemented by the organization than ideas. Finally, we also conducted an additional robustness check and reran our model using centrality measures (i.e., the degree centrality) instead of the effective network size as an indication of an individual's connections in the crowd. The results remain consistent and show that contributions created by members of a crowd with a low degree centrality are more likely to be useful for organizations than contributions created by members of a crowd with a high degree centrality.

Discussion

The results of our study yield several important findings. First, they indicate that useful contributions typically originate from individuals with only few network ties in the crowd and that these contributions generally introduce new information to the platform. Both effects are statistically significant and correspond to theoretical insights from the fields of problem solving and network theory. These effects suggest that contributions created by members of a crowd who are not already immersed in the network offer novel insights and perspectives. Thus, they are especially useful for organizations in their search for new knowledge on crowdsourcing platforms. Similar findings have already been documented by Jeppesen and Lakhani (2010) who note that that technical and social marginality are positively related to problem-solving success in crowdsourcing. However, the results of our study also emphasize the effects of collaboration and knowledge sharing on crowdsourcing platforms that foster interaction between members of crowd. As members of a crowd become more familiar with both their peers and the organization, an increased exchange of mostly local information seems to lead to a convergence of perspectives or cognitive schemata that ultimately stifles their creativity and problem-solving capabilities (Perry-Smith and Shalley 2003). It is also likely that the problem representations of members in the crowd become more confined and concrete as their knowledge and expertise in the crowd grows. Such effects have typically been found to impede diverse solutions and innovation (Förster et al. 2004; Marsh et al. 1999). Another potential explanation is given by related research suggesting that users typically have only a few truly innovative ideas or solutions to offer before they begin to generate less innovative contributions and provide redundant information (Bayus 2013; Von Hippel 2005). Our findings hint that such effects can also be observed on crowdsourcing platforms, which are often specifically used or created by organizations for the purpose of searching for distant knowledge (Afuah and Tucci 2012). Organizations should be aware that, especially when hosting their own crowdsourcing platforms with dedicated crowds, increased participation, feedback, and peer-voting systems may quickly lead to bounded rationalities and local knowledge bases. Thus, it might not always be recommended to rely exclusively on rating scales or preference markets in crowdsourcing when searching for distant knowledge (cf. Blohm et al. 2016), as they might be biased by these bounded rationalities on the platforms. As shown by our results, both the structural origin of a contribution in a crowd as well as the textual characteristics of its content can be used as additional measures to identify potentially useful contributions in crowdsourcing. We argue that organizations searching for useful contributions should pay special attention to novel information provided by new members in a crowd who have not yet made strong connections to representatives of the organization or peers within the crowd.

Second, our study also reveals important insights on how knowledge is absorbed, recombined, and enriched by crowds on crowdsourcing platforms. In traditional group settings, some studies suggest that information diversity and the integration of different perspectives may derail discussions or lead to coordination problems. Cronin and Weingart (2013), for example, argue that diversity bears the risk of making it more difficult for groups to develop a shared understanding and communicate efficiently and effectively. Similarly, van Knippenberg et al. (2004) show how group diversity can lead to social categorizations. In the case of crowdsourcing, however, we found organizations to greatly benefit from different topics or perspectives coming together on the platforms. Consistent with related research (e.g., Faraj et al. 2011; Ye et al. 2016), our findings suggest that (re)combining different perspectives enables crowds to enrich and further develop ideas or solutions for organizations. On our platform specifically, we detected 36 topics that are underlying the contributions and comments created by the crowd. As the platform is hosted by an organization in the transportation and logistics sector, the topics relate to rather particular aspects in this domain, such as “tickets”, “connections”, “calendar”, or “time tables”. Across the platform, we note a relatively even distribution of the topics with around 3% per topic. Our results suggest that contributions whose discussion resembles this distribution and combines a diverse set of topics are more likely to be implemented than contributions whose discussion focuses on only one or few specific topics. For example, in one of the most commented, implemented idea, a member of the crowd suggested that the smartphone application, which the transportation and logistics provider offers to its customers, should automatically track the time and location when travelling by train. The user argued that this would make it easier to check for connections and platforms of departure and arrival. The idea was then taken up by other members of the crowd who linked the suggestion to different topics and became creative, for example, by proposing how the design should look like or how push notifications may be used for even more convenience. Furthermore, it was also discussed whether the idea is technically feasible and how it could be implemented. In addition, we do find statistically significant evidence that contributions whose discussion fosters new information and make the crowd become creative are more likely to be useful for organizations than contributions whose discussion is only based on existing information.

Taken together, this suggests that organizations searching for new knowledge on crowdsourcing platforms should not only pay attention to the contributions themselves, but also how these contributions are being discussed and developed by the crowd. Contributions provided by individuals with few network ties are likely to introduce new information and perspectives which can then be combined or recombined in a creative process with local knowledge by already experienced and immersed members in the crowd. Our findings suggest that when information or insights from different perspectives and topics are brought together and combined on crowdsourcing platforms, the related contributions and the information exchange around these contributions are especially useful for organizations. These findings also resonate with insights from search theory emphasizing the role of knowledge combination and recombination for organizations (e.g., Fleming 2001; Fleming and Sorenson 2004). Thus, our results entail a number of important theoretical and practical implications.

Theoretical Implications

From a theoretical perspective, we are able to contribute novel insights for both search theory and research on crowdsourcing. In recent years, a large body of literature on search theory has emphasized the importance of spanning organizational and technological boundaries in order to find distant knowledge (Katila et al. 2012; Katila and Ahuja 2002; Rosenkopf and Nerkar 2001). The interaction with external sources has been deemed especially promising for coming up with innovative ideas or solving existing problems in organizations (e.g., Chesbrough 2003). As a result, novel, IT-facilitated approaches for distant search, such as crowdsourcing, have emerged (Afuah and Tucci 2012). Crowdsourcing platforms offer organizations an interface to access knowledge distributed amongst large and diverse networks of people. They serve as focal points in this distributed network at which knowledge emerges and evolves.

Based on our findings, however, we argue that simply engaging in crowdsourcing or building crowdsourcing platforms in an attempt to span organizational boundaries may not suffice to find new and distant knowledge. The result presented in this study suggest that, even on crowdsourcing platforms, increased participation, feedback, and information sharing may lead to dominant schemata and mindsets which, ultimately, create local knowledge bases. Similar effects have already been described by Laursen (2012) and Christensen (1997) who argue that searching across organizational boundaries not always implies distant search, since existing customers require organizations to follow established trajectories – even when novel opportunities emerge. We find that it is important to differentiate the contributions with regard to their origin and their content when searching for distant knowledge on crowdsourcing platforms.

Furthermore, the results of our study underline the importance of knowledge recombination for distant search (cf. Fleming 2001; Fleming and Sorenson 2004; Hargadon and Sutton 1997) and extend its understanding in crowdsourcing. Our findings suggest that useful ideas and solutions not only emerge from isolated contributions alone but from a combination of different topics brought together by members of a crowd. Distant contributions from new users are likely to trigger such discussions and make the crowd mobilize its knowledge and become creative. As much as novel inputs and perspectives are important in crowdsourcing for the elicitation of distant knowledge, as much is the additional development and discussion of these inputs by experienced users in the crowd important. These individuals are familiar with the already accumulated knowledge base on the platform and may combine novel ideas or solutions with existing elements. Prior research has already discussed that combinations of existing knowledge elements (depth) with new ideas (scope) are likely to create unique solutions that can be commercialized (Katila and Ahuja 2002; Winter 1984). It shows that crowdsourcing can be employed as a mechanism to orchestrate these processes with knowledge distributed amongst large networks of people.

Finally, from a network perspective, our study also touches upon the classic problem of weak and strong ties in the context of innovation and knowledge sharing (Granovetter 1973, 1983). Our results suggest that novel information is likely to be solicited from members of the crowd who are not yet highly embedded and connected in the network but may bring diverse perspectives from outside the community to the platform. This is consistent with findings presented in earlier research conducted by Hargadon and Sutton (1997) and Lingo and Mahony (2010). Individuals who are not (yet) connected to redundant sources of information may serve as knowledge brokers and draw analogies or introduce new knowledge to a particular field or network. Hence, they have unique informational benefits compared to those who are structurally central and immersed in the network (Lingo and O'Mahony 2010). However, when complex or diverse knowledge needs to be shared and further developed, strong ties become more advantageous. Hansen (1999), for example, shows that weak interunit ties help project teams to search for useful knowledge in other subunits but impede the transfer of complex knowledge, which requires rather strong ties. In crowdsourcing, we found similar effects. New and useful knowledge initially emerges from ideas or solutions provided by members of the crowd who have not yet established a large number of connections in the network and are able to provide novel information to the crowd. Immersed users, on the hand, may add to these contributions by integrating and transferring complex and local knowledge through discussions.

Practical Implications

There are a number of practical implications that can be drawn from our results. First, our result show that useful ideas and solutions are often created by new members in a crowd with only few effective network ties. While it is common for organizations to pay special attention to lead users and follow suggestions that are popular in their established networks, we propose that organizations seeking to gain access to distant knowledge through crowdsourcing should rather focus on contributions that are generated by less immersed members of the crowd. Especially on well-established platforms with stable crowds, using rating scales and voting systems alone may not be the most appropriate mechanisms to identify such contributions. Instead, integrating crowdsourcing platforms with novel business analytics that offer capabilities for social network analysis, text mining, and topic modeling could bridge this gap. Based on our findings, we see great potential in these technologies to support organizations in tracking the origin of contributions in crowdsourcing, analyzing their textual characteristics, and ultimately identifying the most innovative ones. We propose that organizations should delve deeper into such possibilities and make use of novel business analytics or decision support systems when engaging in crowdsourcing.

Second, our results also offer valuable implications for managing crowds and collaboration on crowdsourcing platforms. The results of this study not only emphasize the importance of an initial idea but also the importance of the discussion that unfolds around these contributions. We show that contributions are more likely to be useful for organizations when they spur an exchange of diverse information and when they combine a broad set of topics. This is especially relevant for crowdsourcing intermediaries or organizations that host their own platform. Based on our findings, we urge organizations to encourage the exchange of information between different members of the crowds and integrate incentives for collaboration. We find crowdsourced contributions to be especially useful for organizations when they combine information from a broad range of topics and when the members of a crowd become creative. As unfamiliar and novel information may trigger this creative process, we also suggest that organizations should actively advertise ideas or solutions generated by newer users on the platform, for example, by using recommender systems that promote particularly innovative ideas for further discussion. In line with our first practical implication, not only the most popular and frequently voted contributions should be endorsed by organizations, but also contributions that deviate from well-established patterns.

Finally, our study may serve as the starting point for developers of such business analytics or recommender systems to design and customize related models on crowdsourcing platforms. We provide a set of variables that have been found to be statistically significant predictors for innovative contributions in our study. These variables are based on relatively simple measures, such as the users' effective network size in social network analysis or the TF-IDF-index in information retrieval. They may be used as a foundation or addition for predictive modeling. We thus encourage practitioners to build upon our findings and develop more sophisticated algorithms or models to facilitate the evaluation of large amounts of contributions on crowdsourcing platforms with related business analytics or decision support systems.

Limitations and Future Research

As with all research, the results and implications presented in this study should be regarded in light of its limitations. First, we analyze crowds from a network perspective in this study. The effects described and examined in our paper are inherently based on interactions that unfold on crowdsourcing platforms. Hence, an important boundary condition of this study is that there are connections between members of a crowd and that crowds can be treated as networks of people. As for platforms or crowdsourcing settings with little interaction between contributors, our findings might be less applicable. For future research, it would thus be interesting to further study how different forms of collaboration or interaction on crowdsourcing platforms affect the emergence and recombination of distant knowledge. Our findings should be viewed as initial insights from a collaborative crowdsourcing setting.

Second, this study examines crowdsourcing in an organizational setting. Thus, we focus on a context that leverages crowdsourcing as an approach for distant search in order to harness knowledge outside existing, organizational boundaries (Afuah and Tucci 2012). Furthermore, it must be noted that our results are based on data retrieved from a crowdsourcing platform hosted by a transportation and logistics provider. While the characteristics of the dataset and the platform are similar to those reported in related studies (e.g., Li et al. 2016; Schemmann et al. 2016), there is still the possibility that our findings may not apply to every other industry or application of crowdsourcing to the same extent (e.g., for local governments; see Masdeval and Veloso 2015). In this sense, additional data from more diverse crowdsourcing platforms would greatly benefit the generalizability of our findings. Especially with regard to topic distributions, we expect that platforms hosted by intermediaries who specifically target a broad and diverse crowd may benefit much more from a combination or recombination of knowledge brought together from different domains. In our study, this effect could not be empirically supported at a significance level of 0.05 but only the 0.10 level.

Third, our results are based on a cross-sectional study of the data. Thus, we measured the characteristics of the contributions at a particular point in time on the platform. While cross-sectional analyses are also commonly applied in related studies (e.g., Li et al. 2016; Schemmann et al. 2016), they provide only a static perspective on the dataset and the underlying effects. Another interesting perspective on where and how distant knowledge emerges in crowdsourcing or other distributed networks may be achieved by conducting longitudinal studies or using survival analysis. This dynamic perspective represents a promising avenue for future research to analyze in more detail how knowledge or topics emerge and evolve on crowdsourcing platforms over time.

Fourth, future research may examine in more detail how crowdsourcing platforms and evaluation processes for crowdsourced contributions should be designed. Based on this study, we argue that rating scales and voting systems may not be the most adequate mechanisms for identifying innovative contributions as their results are prone to being biased by already familiar perspectives and popular opinions amongst experienced members of the crowds. Thus, it would be interesting for future research to investigate how IT-supported processes with systems capable of tracking the origin of crowdsourced contributions and analyzing their textual characteristics may support organizations in their evaluation of large amounts of ideas and solutions on their platforms. From a theoretical perspective, we also urge future research to delve deeper into the possibilities of assessing and measuring innovative contributions in crowdsourcing. For example, studies may use textual features, similarity measures, or redundancy indices as proxies for the innovativeness of contributions on crowdsourcing platforms.

Conclusion

Crowdsourcing represents a powerful approach for organizations to engage in distant search and mobilize knowledge distributed amongst a diverse network of people. While organizations generally succeed in generating large amounts of new knowledge on crowdsourcing platforms, it represents a latent challenge to find useful contributions that have the potential to actually solve problems or serve as innovation. In existing literature, little attention has been paid to the emergence and evolution of new knowledge on crowdsourcing platforms and how organizations may identify contributions that capture such knowledge. In this study, we address this gap by analyzing cross-sectional data from a large crowdsourcing platform in Europe and combining statistical approaches from the fields of network analysis and information retrieval to empirically test a set of hypotheses. We find that new and useful contributions are typically created by individuals who have not (yet) established a large number of network ties in the crowd. They have unique informational benefits compared to those who are already immersed in the network. However, their contributions become especially useful when they are further enriched and combined with local knowledge provided by experienced members on the platforms. For researchers in the fields of distant search, knowledge collaboration, and crowdsourcing, we provide a more thorough understanding on how network relationships and information sharing affect the emergence and evolution of knowledge on crowdsourcing platforms. From a practical perspective, we offer guidance for crowdsourcing intermediaries or organizations that host their own crowdsourcing platforms on how to identify potentially useful contributions in the vast pool of data generated by their crowds. We see great opportunity for business analytics to support organizations in tracking the origin of contributions in crowdsourcing, analyzing their textual characteristics, and ultimately identifying the most useful ones. In light of these insights, organizations may leverage distant search on crowdsourcing platforms to its fullest extent.

References

Afuah, A., and Tucci, C. L. 2012. "Crowdsourcing as a Solution to Distance Search," *Academy of Management Review* (37:3), pp. 355–375.

Ahuja, G., and Lampert, C. M. 2001. "Entrepreneurship in the Large Corporation: A Longitudinal Study of How Established Firms Create Breakthrough Inventions," *Strategic Management Journal* (22:6–7), pp. 521–543.

Arun, R., Suresh, V., Veni Madhavan, C. E., and Narasimha Murty, M. 2010. "On Finding the Natural Number of Topics with Latent Dirichlet Allocation: Some Observations," in *Advances in Knowledge Discovery and Data Mining, PAKDD 2010*, Hyderabad: Springer.

Barbier, G., Zafarani, R., Gao, H., Fung, G., and Liu, H. 2012. "Maximizing Benefits from Crowdsourced Data," *Computational and Mathematical Organization Theory* (18:3), pp. 257–279.

Bayus, B. L. 2013. "Crowdsourcing New Product Ideas over Time: An Analysis of the Dell IdeaStorm Community," *Management Science* (59:1), pp. 226–244.

Bjelland, O. M., and Wood, R. C. 2008. "An Inside View of IBM's 'Innovation Jam,'" *MIT Sloan Management Review* (50:1), pp. 32–40.

Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. "Latent Dirichlet Allocation," *Journal of Machine Learning Research* (3:2003), pp. 993–1022.

Blohm, I., Bretschneider, U., Leimeister, J. M., and Krcmar, H. 2010. "Does Collaboration Among Participants Lead to Better Ideas in IT-based Idea Competitions? An Empirical Investigation," in *Proceedings of the 43rd Annual Hawaii International Conference on System Sciences, HICSS 2010*, Honolulu: IEEE, pp. 1–10.

Blohm, I., Leimeister, J. M., and Krcmar, H. 2013. "Crowdsourcing: How to Benefit from (Too) Many Great Ideas," *MIS Quarterly Executive* (12:4), pp. 199–211.

Blohm, I., Riedl, C., Füller, J., and Leimeister, J. M. 2016. "Rate or Trade? Identifying Winning Ideas in Open Idea Sourcing," *Information Systems Research* (27:1), pp. 27–48.

Boudreau, K. J., and Lakhani, K. R. 2013. "Using the Crowd as an Innovation Partner," *Harvard Business Review* (91:4), pp. 60–69.

Bourreau, M., Gensollen, M., and Moreau, F. 2012. "The Impact of a Radical Innovation on Business Models: Incremental Adjustments or Big Bang?," *Industry & Innovation* (19:5), pp. 415–435.

Borgatti, S. P. 1997. "Structural Holes: Unpacking Burt's Redundancy Measures," *Connections* (20:1), pp. 35–38.

Braha, D., and Reich, Y. 2003. "Topological Structures for Modeling Engineering Design Processes," *Research in Engineering Design* (14:4), pp. 185–199.

Burt, R. S. 1995. *Structural Holes: The Social Structure of Competition*, Cambridge, Massachusetts: Harvard University Press.

Carlile, P. R. 2002. "View of Knowledge and Boundaries: Boundary Objects in New Product Development," *Organization Science* (13:4), pp. 442–455.

Chatman, J. A., Polzer, J. T., Barsade, S. G., and Neale, M. A. 2007. "Being Different Yet Feeling Similar: The Influence of Demographic Composition and Organizational Culture on Work Processes and Outcomes," *Administrative Science Quarterly* (43:4), pp. 749–780.

Chesbrough, H. W. 2003. "The Era of Open Innovation," *MIT Sloan Management Review* (44:3), pp. 35–41.

Christensen, C. M. 1997. *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*, Cambridge, Massachusetts: Harvard Business School Press.

Corney, J. R., Torres-Sánchez, C., Jagadeesan, A. P., and Regli, W. C. 2010. "Outsourcing Labour to the Cloud," *International Journal of Innovation and Sustainable Development* (4:4), pp. 294–313.

Cronin, M. A., and Weingart, L. R. 2013. "Representational Gaps, Information and Conflict in Processing, Diverse Teams Functionally," *Academy of Management Review* (32:3), pp. 761–773.

Dell. 2017. "IdeaStorm," (available at http://www.ideastorm.com/; retrieved March 15, 2017).

Doan, A., Ramakrishnan, R., and Halevy, A. Y. 2011. "Crowdsourcing Systems on the World-Wide Web," *Communications of the ACM* (54:4), pp. 86–96.

Dosi, G. 1982. "Technological Paradigms and Technological Trajectories. A Suggested Interpretation of the Determinants and Directions of Technical Change," *Research Policy* (11:3), pp. 147–162.

De Dreu, C. K. W., and West, M. A. 2001. "Minority Dissent and Team Innovation: The Importance of Participation in Decision Making," *Journal of Applied Psychology* (86:6), pp. 1191–1201.

Dunbar, K. 1998. "Problem Solving," in *A Companion to Cognitive Science*, W. Bechtel and G. Graham (eds.), London: Blackwell, pp. 289–298.
Durward, D., Blohm, I., and Leimeister, J. M. 2016. "Crowd Work," *Business & Information Systems Engineering* (58:4), pp. 281–286.
Faraj, S., Jarvenpaa, S. L., and Majchrzak, A. 2011. "Knowledge Collaboration in Online Communities," *Organization Science* (22:5), pp. 1224–1239.
Feldman, R., and Sanger, J. 2007. *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, Cambridge: Cambridge University Press.
Felin, T., and Zenger, T. R. 2014. "Closed or Open Innovation? Problem Solving and the Governance Choice," *Research Policy* (43:5), Elsevier B.V., pp. 914–925.
Fleming, L. 2001. "Recombinant Uncertainty in Technological Search," *Management Science* (47:1), pp. 117–132.
Fleming, L., and Sorenson, O. 2004. "Science as a Map in Technological Search," *Strategic Management Journal* (25:8–9), pp. 909–928.
Förster, J., Friedman, R. S., and Liberman, N. 2004. "Temporal Construal Effects on Abstract and Concrete Thinking: Consequences for Insight and Creative Cognition," *Journal of Personality and Social Psychology* (87:2), pp. 177–189.
Geiger, D., and Schader, M. 2014. "Personalized Task Recommendation in Crowdsourcing Information Systems - Current State of the Art," *Decision Support Systems* (65), pp. 3–16.
Gieryn, T. F., and Hirsh, R. F. 1983. "Marginality and Innovation in Science," *Social Studies of Science* (13:1), pp. 87–106.
Granovetter, M. 1973. "The Strength of Weak Ties," *American Journal of Sociology*, pp. 1360–1380.
Granovetter, M. 1983. "The Strength of Weak Ties: A Network Theory Revisited," *Socialogical Theory* (1:1), pp. 201–233.
Griffiths, T. L., and Steyvers, M. 2004. "Finding Scientific Topics," *Proceedings of the National Academy of Sciences* (101:1), pp. 5228–5235.
Hansen, M. T. 1999. "The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge across Organization Subunits," *Administrative Science Quarterly* (44:1), pp. 82–111.
Hargadon, A., and Sutton, R. I. 1997. "Technology Brokering and Innovation in a Product Development Firm," *Administrative Science Quarterly* (42:4), p. 716-749.
Hatchuel, A., and Weil, B. 2009. "C-K Design Theory: An Advanced Formulation," *Research in Engineering Design* (19:4), pp. 181–192.
Hautz, J., Hutter, K., Füller, J., Matzler, K., and Rieger, M. 2010. "How to Establish an Online Innovation Community? The Role of Users and Their Innovative Content," in *Proceedings of the 43rd Hawaii International Conference on Systems Sciences, HICSS 2010*, Kauai: IEEE, pp. 1–11.
Helfat, C. E. 1994. "Evolutionary Trajectories Firm R & D in Petroleum," *Management Science* (40:12), pp. 1720–1747.
Von Hippel, E. 2005. *Democratizing Innovation*, Cambridge, Massachusetts: MIT Press.
Huang, Y., Singh, P. V., and Srinivasan, K. 2014. "Crowdsourcing New Product Ideas Under Consumer Learning," *Management Science* (60:90), pp. 2138–2159.
Jeppesen, L. B., and Lakhani, K. R. 2010. "Marginality and Problem-Solving Effectiveness in Broadcast Search," *Organization Science* (21:5), pp. 1016–1033.
Kanter, R. M. 1988. "When a Thousand Flowers Bloom: Structural, Collective, and Social Conditions for Innovation in Organization," *Research in Organizational Behavior* (10:1988), pp. 169–211.
Katila, R., and Ahuja, G. 2002. "Something Old, Something New: A Longitudinal Study of Search Behavior and New Product Introduction," *Academy of Management Journal* (45:6), pp. 1183–1194.
Katila, R., Chen, E. L., and Piezunka, H. 2012. "All the Right Moves: How Entrepreneurial Firms Compete Effectively," *Strategic Entrepreneurship Journal* (6:2012), pp. 116–132.
Kristensson, P., Gustafsson, A., and Archer, T. 2004. "Harnessing the Creative Potential Among Users," *Journal of Product Innovation Management* (21:1), pp. 4–14.
Laursen, K. 2012. "Keep Searching and You'll Find: What Do We Know About Variety Creation Through Firms' Search Activities for Innovation?," *Industrial and Corporate Change* (21:5), pp. 1181–1220.
Leicht, N., Blohm, I., and Leimeister, J. M. 2017. "Leveraging the Power of the Crowd for Software Testing," *IEEE Software* (34:2), pp. 62–69.
Leimeister, J. M., Huber, M., Bretschneider, U., and Krcmar, H. 2009. "Leveraging Crowdsourcing: Activation-Supporting Components for IT-Based Ideas Competition," *Journal of Management Information Systems* (26:1), pp. 197–224.

Levitt, T. 1963. "Creativity is Not Enough," *Harvard Business Review* (41:3), pp. 72–83.

Li, M., Kankanhalli, A., and Kim, S. H. 2016. "Which Ideas are More Likely to Be Implemented in Online User Innovation Communities? An Empirical Analysis," *Decision Support Systems* (84), pp. 28–40.

Li, Q., Magiitti, P. G., Smith, K. G., Tesluk, P. E., and Katila, R. 2013. "Top Management Team Attention to Innovation: The Role of Search Selection And Intentsity in New Product Introductions," *Academy of Management Journal* (56:3), pp. 893–916.

Lingo, E. L., and O'Mahony, S. 2010. "Nexus Work: Brokerage on Creative Projects," *Administrative Science Quarterly* (55:1), pp. 47–81.

Lovett, M. C., and Anderson, J. R. 1996. "History of Success and Current Context in Problem Solving: Combined Influences on Operator Selection," *Cognitive Psychology* (31:2), pp. 168–217.

Lu, Y., Singh, P. V., and Sun, B. 2017. "Is a Core-Periphery Network Good For Knowledge Sharing? A Structural Model of Endogenous Network Formation on a Crowdsourced Customer Support Forum," *MIS Quarterly* (41:2), pp. 607–628.

Lüthje, C., Herstatt, C., and Von Hippel, E. 2005. "User-Innovators and 'Local' Information: The Case of Mountain Biking," *Research Policy* (34:6), pp. 951–965.

Magnusson, P. R. 2009. "Exploring the Contributions of Involving Ordinary Users in Ideation of Technology-Based Services," *Journal of Product Innovation Management* (26:5), pp. 578–593.

Majchrzak, A., and Malhotra, A. 2013. "Towards an Information Systems Perspective and Research Agenda on Crowdsourcing for Innovation," *Journal of Strategic Information Systems* (22:4), Elsevier B.V., pp. 257–268.

March, J. G. 1991. "Exploration and Exploitation in Organizational Learning," *Organization Science* (2:1), pp. 71–87.

Marsh, R. L., Ward, T. B., and Landau, J. D. 1999. "The Inadvertent Use of Prior Knowledge in a Generative Cognitive Task," *Memory & Cognition* (27:1), pp. 94–105.

Martin, X., and Mitchell, W. 1998. "The Influence of Local Search and Performance Heuristics on New Design Introduction in a New Product Market," *Research Policy* (26:7–8), pp. 753–771.

Masdeval, C., and Veloso, A. 2015. "Mining Citizen Emotions to Estimate the Urgency of Urban Issues," *Information Systems* (54:2015), Elsevier, pp. 147–155.

Mumford, M. D., and Gustafson, S. B. 1988. "Creativity Syndrome: Integration, Application, and Innovation," *Psychological Bulletin* (103:1), pp. 27–43.

Nagelkerke, N. J. D. 1991. "A Note on a General Definition of the Coefficient of Determination," *Biometrika* (78:3), pp. 691–692.

Nelson, R., and Winter, S. G. 1982. *An Evolutionary Theory of Economic Change*, Cambridge, Massachusetts: Harvard University Press.

Niekler, A., and Jähnichen, P. 2012. "Matching Results of Latent Dirichlet Allocation for Text," *Proceedings of the 11th International Conference on Cognitive Modeling*, pp. 317–322.

Ocasio, W. 1997. "Towards an Attention-Based View of the Firm," *Strategic Management Journal* (18), pp. 187–206.

Perry-Smith, J. E., and Shalley, C. E. 2003. "The Social Side of Creativity: A Static and Dynamic Social Network Perspective," *The Academy of Management Review* (28:1), pp. 89–106.

Piezunka, H., and Dahlander, L. 2015. "Distant Search, Narrow Attention: How Crowding Alters Organizations' Filtering of Suggestions in Crowdsourcing," *Academy of Management Journal* (58:3), pp. 856–880.

Poetz, M. K., and Schreier, M. 2012. "The Value of Crowdsourcing: Can Users Really Compete with Professionals in Generating New Product Ideas?," *Journal of Product Innovation Management* (29:2), pp. 245–256.

Powell, W. W., Koput, K. W., and Smith-Doerr, L. 1996. "Interorganizational Collaboration and the Locus of Innovation: Networks of Learning in Biotechnology," *Administrative Science Quarterly* (41:1), pp. 116–145.

Rhyn, M., and Blohm, I. 2017. "A Machine Learning Approach for Classifying Textual Data in Crowdsourcing," in *Proceedings of the 13th International Conference on Wirtschaftsinformatik (WI)*, St. Gallen, pp. 1171–1185.

Riedl, C., and Woolley, A. W. 2016. "Teams vs. Crowds: A Field Test of the Relative Contribution of Incentives, Member Ability, and Collaboration to Crowd-Based Problem Solving Performance," *Academy of Management Discoveries* (doi: 10.5465/amd.2015.0097).

Rogstadius, J., Vukovic, M., Teixeira, C. A., Kostakos, V., Karapanos, E., and Laredo, J. A. 2013. "CrisisTracker: Crowdsourced Social Media Curation for Disaster Awareness," *IBM Journal of Research and Development* (57:5), pp. 1–13.

Rosenkopf, L., and Almeida, P. 2003. "Overcoming Local Search Through Alliances and Mobility," *Management Science* (49:6), pp. 751–766.

Rosenkopf, L., and Nerkar, A. 2001. "Beyond Local Search: Boundary-Spanning, Exploration, and Impact in the Optical Disk Industry," *Strategic Management Journal* (22:4), pp. 287–306.

Rothaermel, F. T., and Alexandre, M. T. 2009. "Ambidexterity in Technology Sourcing: The Moderating Role of Absorptive Capacity," *Organization Science* (20:4), pp. 759–780.

Salton, G., and Buckley, C. 1988. "Term Weighting Approaches in Automatic Text Retrieval," *Information Processing & Management* (24:5), pp. 513–523.

Schemmann, B., Herrmann, A. M., Chappin, M. M. H., and Heimeriks, G. J. 2016. "Crowdsourcing Ideas: Involving Ordinary Users in the Ideation Phase of New Product Development," *Research Policy* (45:6), Elsevier B.V., pp. 1145–1154.

Schenk, E., and Guittard, C. 2011. "Towards a Characterization of Crowdsourcing Practices," *Journal of Innovation Economics & Management* (7:1), pp. 93–107.

Schulten, M. B., and Schaefer, F. 2015. "Affective Commitment and Customer Loyalty in Crowdsourcing: Antecedents, Interdependencies, and Practical Implications," *The International Review of Retail, Distribution and Consumer Research* (25:5), Routledge, pp. 516–528.

Schulze, T., Indulska, M., Geiger, D., and Korthaus, A. 2012. "Idea Assessment in Open Innovation: A State of Practice," in *Proceedings of the 20th European Conference on Information Systems (ECIS)*, Barcelona, Spain: AIS, pp. 1–12.

Simula, H., and Ahola, T. 2014. "A Network Perspective on Idea and Innovation Crowdsourcing in Industrial Firms," *Industrial Marketing Management* (43:3), Elsevier Inc., pp. 400–408.

Stephen, A. T., Zubcsek, P. P., and Goldenberg, J. 2016. "Lower Connectivity Is Better: The Effects Of Network Structure On Redundancy Of Ideas And Customer Innovativeness In Interdependent Ideation Tasks," *Journal of Marketing Research* (53:2), pp. 263–279.

Stol, K.-J., LaToza, T. D., and Bird, C. 2017. "Crowdsourcing for Software Engineering," *IEEE Software* (34:2), pp. 30–36.

Stuart, T. E., and Podolny, J. M. 1996. "Local Search and the Evolution of Technological Capabilities," *Strategic Management Journal* (17:1996), pp. 21–38.

Sullivan, B. N. 2010. "Competition and Beyond: Problems and Attention Allocation in the Organizational Rulemaking Process," *Organization Science* (21:2), pp. 432–450.

Thuan, N. H., Antunes, P., and Johnstone, D. 2016. "Factors Influencing the Decision to Crowdsource: A Systematic Literature Review," *Information Systems Frontiers* (18:1), pp. 47–68.

Tripsas, M., and Gavetti, G. 2000. "Capabilities, Cognition, and Inertia: Evidence From Digital Imaging," *Strategic Management Journal* (21:10), pp. 1147–1161.

Trope, Y., and Liberman, N. 2010. "Construal-Level Theory of Psychological Distance," *Psychological Review* (117:2), pp. 440–463.

van Knippenberg, D., De Dreu, C. K. W., and Homan, A. C. 2004. "Work Group Diversity and Group Performance: An Integrative Model and Research Agenda," *Journal of Applied Psychology* (89:6), pp. 1008–1022.

Winter, S. G. 1984. "Schumpeterian Competition in Alternative Technological Regimes," *Journal of Economic Behavior and Organization* (5:3–4), pp. 287–320.

Ye, H. J., Blohm, I., Bretschneider, U., Goswami, S., Leimeister, J. M., and Krcmar, H. 2016. "Promoting the Quality of User Generated Ideas in Online Innovation Communities: A Knowledge Collaboration Perspective," in *Proceedings of the 37th International Conference on Information Systems (ICIS)*, Dublin, Ireland: AIS, pp. 1–16.

Zhang, Q., Zeng, D. D., Wang, F.-Y., Breiger, R., and Hendler, J. A. 2016. "Brokers or Bridges? Exploring Structural Holes in a Crowdsourcing System," *IEEE Computer* (49:6), pp. 56–64.

Zhao, Y., and Zhu, Q. 2014. "Evaluation on Crowdsourcing Research: Current Status and Future Direction," *Information Systems Frontiers* (16:3), pp. 417–434.

Zogaj, S., Bretschneider, U., and Leimeister, J. M. 2014. "Managing Crowdsourced Software Testing: A Case Study Based Insight on the Challenges of a Crowdsourcing Intermediary," *Journal of Business Economics* (84:3), pp. 375–405.

Zuchowski, O., Posegga, O., Schlagwein, D., and Fischbach, K. 2016. "Internal Crowdsourcing: Conceptual Framework, Structured Review, and Research Agenda," *Journal of Information Technology* (31:2), pp. 166–184.

Zwass, V. 2010. "Co-Creation: Toward a Taxonomy and an Integrated Research Perspective," *International Journal of Electronic Commerce* (15:1), pp. 11–48.