Leveraging the Wisdom of Crowd to Address Societal Challenges: A Revisit to the Knowledge Reuse Process for Innovation through Analytics*

Yue Han[1], Pinar Ozturk[2], Jeffrey V. Nickerson[3]

[1]Le Moyne College, USA, hany@lemoyne.edu, [2]Duquesne University, USA, ozturkp@duq.edu, [3]Stevens Institute of Technology, USA, jnickerson@stevens.edu

Abstract

Societal challenges can be addressed not only by experts, but also by crowds. Crowdsourcing provides a way to engage the general crowd to contribute to the solutions of the biggest challenges of our times: how to cut our carbon footprint, how to address worldwide epidemic of chronic disease, and how to achieve sustainable development. Isolated crowd-based solutions in online communities are not always creative and innovative. Hence, remixing has been developed as a way to enable idea evolution and integration, and to harness reusable innovative solutions. Understanding the generativity of remixing is essential to leveraging the wisdom of the crowd to solve societal challenges. At its best, remixing can promote online community engagement, as well as support comprehensive and innovative solution generation. Organizers can maintain an active online community; community members can collectively innovate and learn; and as a result, society may find new ways to solve important problems. What affects the generativity of a remix? We address this by revisiting the knowledge reuse process for innovation model. We analyze the reuse of proposals in an online innovation community which aims to address global climate change issues, Climate CoLab. We apply several analytical methods to study factors that may contribute to the generativity of a remix and uncover that remixes that include prevalent topics and integration metaknowledge are more generative. Our findings suggest strategies and tools

* Han, Y., Ozturk, P., and Nickerson, J.V. (2020) Leveraging the Wisdom of Crowd to Address Societal Challenges: A Revisit to the Knowledge Reuse Process for Innovation through Analytics, *Journal of the Association for Information Systems.*

that can help online communities to better harness collective intelligence for addressing societal challenges.

Keywords: Societal challenges, climate change, innovation, knowledge reuse, remixing, online communities, collective intelligence

1 Introduction

Large-scale societal issues have been framed as wicked problems (Head & Alford, 2015; Rittel & Weber, 1973) and grand challenges (George, Howard-Grenville, Joshi, & Tihanyi, 2016). These challenges are difficult to solve. On the one hand, many of these challenges are urgent, yet there is no central authority to solve them; different stakeholders do not even agree on what these problems really are. On the other hand, these challenges are composed of complex dilemmas and emergent issues – all of which are dynamic, contextually bound, and require changes in individual and societal behaviors (George et al., 2016).

Traditionally, the policies and proposals to address societal challenges have been created by legislators, policy-makers, and experts within organizations and businesses (Margolis & Walsh, 2003; Scherer & Palazzo, 2011; Callaghan, 2014). However, organizations and businesses are fundamentally unable to deal with these challenges on their own because their innovation pipeline is inherently full of inefficiencies, delays, dictates of the market mechanisms, and political decision-making (Chesbrough, 2006; Grimm, Fox, Baines, & Albertson, 2013), which have raised many discussions in different disciplines – including the IS community. One stream of IS research has focused on applying different information and communication technologies to help organizations address societal challenges (Leong, Pan, Newell, & Cui, 2016; Srivastava, Teo, & Devaraj, 2016; Venkatesh, Rai, Sykes, & Aljafari, 2016). Another stream aligns with the open call on "consider emergent digital designing as a replacement for organizations" (Majchrzak, Markus, & Wareham, 2016): these IS researchers seek to solve societal challenges with the wisdom of the crowd that is external to the organizations (Brabham, 2008; Mergel & Desouza, 2013; Malone, 2018).

By nature, societal challenges reflect the issues and problems the majority of the society are facing, so crowd members are likely to be affected. Because these problems are general, organizations and businesses have turned to the crowd to tackle them. Crowds are perceived as capable of providing innovative solutions that internal teams might not be (Brabham, Ribisl, Kirchner, & Bernhardt, 2014; Certoma, Corsini, & Rizzi, 2015; Muller et al., 2015; Schlagwein & Bjørn-Andersen, 2014). Thus, in recent years, crowdsourcing has been adopted to aid organizations and businesses to address societal challenges (Mergel & Desouza, 2013; Malone et al., 2017). One of the foremost examples is the Challenge.gov platform built following the principles of President Barack Obama's Open Government initiative (White House, 2009). As a crowdsourcing open innovation initiative of the US government, Challenge.gov adopts a crowdsourcing approach aiming to engage previously disenfranchised stakeholders in solving particular problems of government agencies (Mergel & Desouza, 2013). Other examples include Climate CoLab platform, where people from all over the world create proposals to solve climate change challenges (Malone et al., 2017), and Foldit platform, where users view and build on each other's models of protein structures, leading to a solution in as few as three weeks, when years of medical research had been unsuccessful up to that point (Cooper et al., 2010; Khatib et al., 2011). Table A1 in the Appendix provides a summary of the main existing usage of crowdsourcing for solving societal challenges, together with their pros and cons.

A major strength of crowdsourcing is that it generates a large number of ideas within a short amount of time, which is difficult to achieve within organizations (Chiu, Liang, & Turban, 2014; Majchrzak & Malhotra, 2013). Involving crowd members in solving societal challenges can also expand the exploration of the solution space because the crowd by its nature engages in divergent thinking (Afuah & Tucci, 2012; Ward, 2001). In addition, inviting a general crowd to address a societal challenge increases the public's awareness of the issues and the potential solutions. Although crowdsourcing has many benefits, prior research has also expressed concerns regarding the potential issues (Bayus, 2013; Saxton, Oh, & Kishore, 2013). One main critique is that crowdsourcing is not efficient because many of the ideas generated are superficial or redundant (Bjelland & Wood, 2008). Another critique is that crowd tends to fail to incorporate multiple perspectives when generating ideas (Schenk & Guittard, 2011). These two issues have raised general discussions surrounding how to better use crowds to address complex problems, including societal challenges. One solution is remixing, which comes with a structure that allows task division and integration: crowd members can build on, reuse, and recombine previous works done by not only themselves but also others to generate new ideas (Hill & Monroy-Hernández, 2013; Kyriakou, Nickerson, & Sabnis, 2017; Malone et al. 2017). Unlike the initial implementation of crowdsourcing that asks crowd members to work individually on ideas, remixing allows access to others' work, which has the potential to not only reduce the redundancy but also deepen ideas, which in turn can lead to innovation at the collective level (Malone, 2018; Wisdom & Goldstone 2011). By integrating multiple prior works, the crowd can also develop more comprehensive ideas. In addition, remixing brings a learning opportunity for all community members (Dasgupta, Hale, Monroy-Hernández, & Hill, 2016): They can deepen their understanding of the domain knowledge through remixing each other's work. Thus, many online innovation communities have incorporated remixing into their platform design to harness collective intelligence (Cheliotis & Yew, 2009; Kyriakou et al. 2017; Resnick et al. 2009; Malone et al. 2017).

Remixing can be seen as a process that supports knowledge reuse in online communities. While knowledge is sometimes reused for convenience, as when one acquires a technology product, we focus on knowledge being reused in order to build deeper knowledge, a process called knowledge reuse for innovation (Armbrecht et al., 2001; Majchrzak, Cooper, & Neece, 2004). Previous studies on knowledge reuse for innovation (KRI) in offline settings mostly focus on studying factors that affect the quality of the innovation (Boh 2008; Cheung, Chau, & Au, 2008), while studies on KRI in online settings tend to focus more on the generativity of the innovation (Hill & Monroy-Hernández, 2013; Kyriakou et al., 2017; Stanko, 2016). In general, a reusable innovation can trigger more contributions from other community members, which promotes online social engagement and supports innovative idea generation. In addition, generativity is essential to tackle societal challenges: Incentivizing reusable innovations could help to increase the collective awareness of a societal challenge and expand the coverage of the solution space by inspiring more comprehensive solutions.

Although prior studies have examined multiple artifacts that affect the generativity of a newly created innovation (Hill & Monroy-Hernández, 2013; Kyriakou et al. 2017), these studies have not addressed remixing in the societal challenge domain. Solving a societal challenge is very different from any of the tasks in the online innovation communities that have been studied in previous research: 1) the task has a specific objective while other remixing communities mostly seek for open-ended creations, 2) the sequential nature of the reuse process is critical for solving societal challenges as new creations need to integrate previous ideas, while the reuse in other online communities could happen at any point of the remix network, 3) in solving societal challenges, no single solution or formulation of the problem is sufficient because different stakeholders do not even agree on what the problem really is, and therefore, there are no right or wrong answers, only answers that are better or worse from different points of view. Keeping in mind the uniqueness and complexity of these issues, we explore how to encourage reusable innovative ideas to leverage the wisdom of the crowd to address societal challenges.

To better use remixing to help address societal challenges, it is important to understand the nature of remixing – the KRI process. Thus, we revisited the KRI process and made an analytical approach to understand how the knowledge reuse process affects the generativity of an innovation. The KRI model is a six-stage process model that involves three major actions—

reconceptualizing the problem, searching and evaluating ideas to reuse, and developing the selected idea (Majchrzak et al, 2004). Specifically, we aim to address the following research question in this paper: How do the three major actions in the knowledge reuse process for innovation affect the generativity of an innovation that addresses societal challenges? In other words, what processes may help someone reuse knowledge that can in turn generate more reuse for solving grand challenges?

To examine this research question, we collected and analyzed data from an online innovation community called Climate CoLab, which aims to address one type of societal challenge–global climate change (Malone et al., 2017). In the Climate CoLab website, community members are encouraged to participate in different contests by creating novel proposals that address global climate change. In these contests, proposal creators search for and integrate pre-existing proposals when creating their novel entries. The contests are designed to record traces of the knowledge reuse path: what content has been reused, when it was reused, and whether this content descended from previous content. Because of the complexity of societal challenges, we applied multiple text analytical methods including a specialized technique that uses the community-generated Wikipedia ontology for topic detection and text similarity comparison. We examined the effect of the three major actions on the generativity of the final creation by analyzing three important outcome features of these actions: proposal topic prevalence, the number of high-quality proposals reused, and the encoded metaknowledge about the rationale for the integration.

This paper aims to contribute to both the knowledge reuse literature and the usage of emerging digital designs to leverage the wisdom of the crowd for tackling societal challenges (Majchrzak et al. 2016; Markus, 2001). Our findings reveal how the three major actions in the KRI process affects the reusability of a remix and suggest that incorporating prevalent topics when reconceptualizing the problem and encoding metaknowledge about the integration of the ideas reused when developing the integrated idea can increase the generativity of a final creation that addresses societal challenges. These findings can be generalized to other online innovation communities–those not specifically designed to solve certain societal challenges but utilize collective intelligence to contribute to the solution of other complex problems. This paper also aims to provide practical implications. Our findings provide knowledge workers various strategies to increase the generativity of their creations, and suggest online innovation community designers different tools to better support knowledge reuse for innovation in an online community.

The structure of the rest of the paper is as follows. The following sections start with a brief review of related work, followed by our hypotheses. Then we describe our empirical study and present the analyses and results. Finally, we discuss the implications for theory and practice and suggest future research possibilities.

2 Theoretical Development

2.1 Remixing as a Method to Support Innovation

Traditionally, in-house experts within an organization work on creating new product and service innovations via a private-collective innovation model (von Hippel & von Krogh, 2003). Nowadays many organizations also generate new product and service ideas from customers via open innovation (Chesbrough, 2006; Eservel, 2014). Open innovation can bring organizations new perspectives since it recognizes customers as an important idea source that can actively contribute to product innovations besides passively providing valuable information for marketing and sales divisions of the organization (Chesbrough & Crowther, 2006; Elmquist, Fredberg, & Ollila, 2009; Kristensson, Magnusson, & Matthing, 2002). While open innovation has been widely discussed by researchers, another term – crowdsourcing – has been used to describe the phenomena in which strangers are recruited to accomplish tasks (Howe, 2006). Some researchers identify crowdsourcing as a process that can produce open innovation (Estellés-Arolas & González-Ladrón-de-Guevara, 2012; Phillips, 2010). Others suggest that these two concepts are at the same logical level and share an overlapping domain—crowd innovation (Howe, 2008).

As the development of Internet technology has promoted the reshaping of digital collaboration patterns in online communities (Hoegg, Martignoni, Meckel, & Stanoevska-Slabeva, 2006), knowledge reuse has been incorporated into the design of some online crowd innovation communities such as ccMixter, Climate CoLab, and Scratch (Cheliotis & Yew, 2009; Malone et al., 2017; Resnick et al., 2009). These sites allow users to search and repurpose user-generated content to generate creative outcomes. They are also designed to trace the knowledge reuse path: what content has been reused and where this content came from within the community. Many scholars use *remixing, a term* originated in the music industry to describe a process of modifying music by changing the attributes of its component tracks, to refer to this traceable knowledge reuse and to describe combinations in online communities (Cheliotis, Hu, Yew, & Huang, 2014; Faraj, Jarvenpaa, & Majchrzak, 2011; Lessig, 2008; Navas, 2012).

Remixing has a built-in feature of engagement in that it encourages people to build on each other's work. Because of this, remixing can be used as a tool to support crowd creativity: Community members build upon others' work to develop further innovations, and then share these improvements for others to reuse (Hill & Monroy-Hernández, 2013; Nickerson, 2015; Sojer & Henkel 2010). Typical examples of this kind of community include Wikipedia, where contributors collaborate in editing articles for an encyclopedia (Estellés-Arolas & González-Ladrón-de-Guevara, 2012); GitHub, where users build and reuse software code together (Dabbish, Stuart, Tsay, & Herbsleb, 2012); Scratch, where children create and remix projects using programming (Resnick et al., 2009); and Thingiverse, where participants design and recombine 3D printing ideas (Flath, Friesike, Wirth, & Thiesse, 2017). More importantly, remixing can also be used to harness collective intelligence for citizen science such as the Climate CoLab website (Malone et al., 2017). To better use remixing for solving societal challenges and other complex tasks, it is essential to understand knowledge reuse in online communities.

2.2 Knowledge Reuse for Innovation

Knowledge reuse is commonly interpreted as the process of locating and using shared knowledge (Alavi & Leidner, 2001). Researchers believe that knowledge reuse is important to study because it contributes to combinative capabilities (Grant, 1996; Kogut & Zander, 1992) and innovation in organizations (Armbrecht et al., 2001; Majchrzak et al., 2004). To understand knowledge reuse, researchers have created several frameworks, which have become the foundations for later studies. Grant (1996) developed a knowledge-based theory that focuses on the analysis of knowledge integration mechanisms. Szulanski (2000) created a four-stage knowledge reuse process with a "knowledge reuse as replication" focus. Markus (2001) developed a theory of successful knowledge reuse with an emphasis on knowledge management systems and repositories.

These models explain reuse for replication but not reuse for innovation. This is because reuse for replication at best contributes to incremental innovation, not radical innovation. The processes used in radical innovation are different (Argote, 2012; Grant, 1996; Leonard & Sensiper, 1998). Knowledge reuse for replication is a process that focuses on knowledge acquisition in solving a problem or increasing productivity; knowledge reuse for innovation, on the other hand, involves knowledge integration: Knowledge workers integrate others' knowledge with their own knowledge to generate innovation. Because of this, Majchrzak, Cooper, and Neece (2004) built a staged process model for knowledge reuse for innovation that explains how innovators search for and recombine knowledge in order to generate new knowledge. This model will be referred to as the knowledge reuse process (KRI) model in this paper.

The knowledge reuse process model has been used as the foundation for later studies. A few researchers extended the discussion and suggested enhancements (e.g. Chewar & McCrickard, 2005). Most of these studies focused on the discussion of what artifacts affect the quality of the innovation and how to further optimize these artifacts to improve the knowledge reuse process (Boh, 2008; Durcikova & Fadel, 2016; Faniel & Majchrzak, 2007; Kankanhalli, Lee, & Lim, 2011; Khedhaouria & Jamal, 2015; Majchrzak & Malhotra, 2013). For example, Faniel and Majchrzak (2007) suggested ways to optimize the knowledge management systems and technologies to help knowledge reuse for innovation. Durcikova and Fadel (2016) discussed how to better use knowledge electronic repositories at the search stage of knowledge reuse. In addition, a few papers have explored how adaption, metaknowledge, and other factors influence knowledge reuse (McGrath & Parkes, 2007; Zhang & Watts, 2008). Although there is a rich literature on knowledge reuse for innovation and the quality of the innovative outcome, few studies have explored the relationship between the process and the generativity/reusability of the innovative outcome.

2.3 Generativity in Remixing Research

The overarching goals of most online innovation communities are 1) attracting more participation and 2) generating creative work. Previous research suggests that remixing is an important form of online engagement (Banker, Bardhan, & Asdemir, 2006). Also, a reusable innovation may trigger more contributions from other community members, which could also contribute to the generation of innovative ideas. As both goals can be achieved by increasing the reusability of remixes, it is essential to study the reusability of creations in online innovation communities (Cheliotis et al., 2014; Hill & Monroy-Hernández, 2013). The reusability of creations in online remixing communities is described by these studies as generativity. Generativity, also called fecundity, represents to the number of times a work is remixed/reused (Hill & Monroy-Hernández, 2013).

Previous studies sought to determine factors that are correlated with generativity (Hill & Monroy-Hernández, 2013; Jarvenpaa & Standaert, 2018; Stanko, 2016). Some researchers have found that popularity, intertextuality, and derivativity, as well as the author's fecundity and social embeddedness all affect generativity (Cheliotis et al., 2014). To better understand generativity in a remixing community, Kyriakou, Nickerson, and Sabnis (2017) studied a 3D printing design community and discussed the relationship between reuse and metamodels—a kind of reuse for innovation. However, metamodels are very specific components of the knowledge reuse for innovation process. By contrast, we are interested in the effects of the major actions in the knowledge reuse process on the reuse of the resulting innovation and the sequence of steps taken to create an innovation. Thus, there is a need to revisit the process of knowledge reuse for innovation to understand how the process affects generativity.

2.4 Three Major Actions in Knowledge Reuse for Innovation

The knowledge reuse process for innovation in Majchrzak, Cooper, and Neece (2004) has six stages: reconceptualize the problem and the approach for innovation; decide to search for reusable ideas; scan for reusable ideas; briefly evaluate reusable ideas; conduct in-depth analysis on reusable ideas and select one; and fully develop the reused idea. This process consists of three major actions: 1) reconceptualize the problem, 2) search and evaluate ideas to reuse, and 3) develop the selected idea.

2.4.1 Major Action 1: Reconceptualize the Problem

The first major action in the knowledge reuse process model is to reconceptualize the problem. In this action, creators redefine the problem and determine the main theme of their creation. They need to find a balance between ambitious conceptualizations and the potential existence of an idea that they can reuse (Majchrzak et al., 2004). This leads to a tradeoff between novelty and prevalence. A prevalent idea can be an idea that includes commonly discussed fundamental topics; it can also be an idea that includes non-fundamental but popular topics that are trending within the community. A prevalent idea is more likely to be reused by creators either because of preferential attachment within the reuse network (Barabási & Albert, 1999), or because of familiarity (Brown & Duguid, 1991; Nonaka, 1994; Hutcheon, 2012). Therefore, we propose that ideas that include more prevalent topics are more likely to be reused. This leads to the following hypothesis related to the performance of the first action in the knowledge reuse process model—the problem reconceptualization hypothesis:

H1: A remix containing more prevalent topics is more likely to be reused.

2.4.2 Major Action 2: Search and Evaluate Ideas to Reuse

The second action in the knowledge reuse process model is searching for and evaluating ideas to reuse. In this action, creators select ideas that can be reused in their new idea. Both the quantity and quality of the ideas they select are indicators of the creator's performance. Therefore, we measure the performance of this action by counting the number of high-quality ideas creators decide to reuse.

Some researchers suggested that remixes tend to form chains; creations that are remixes themselves are more likely to generate future remixes in an online remixing community called Scratch (Hill & Monroy-Hernández, 2013). However, another study examined a music remixing community and suggested that a music remix that has reused more previous music works is less likely to be reused by others because users find it easier to reuse a single work than to recombine multiple sources (Cheliotis & Yew, 2009). Later, researchers found that the relationship between the number of previous works reused in a remix and the generativity of the remix is not linear. Instead, there exists a U-shaped relationship (Cheliotis et al., 2014). Since the type of artifact studied in that paper is a special media form—music—we want to test if the relationship observed in that study can be generalized to other remix communities. As a result, we propose the following hypothesis related to the performance of the second action in the knowledge reuse process for innovation—the idea search and evaluation hypothesis:

H2: The number of high-quality ideas reused in a remix has a U-shape relationship with the generativity of this remix.

2.4.3 Major Action 3: Develop the Selected Idea

The third action in the knowledge reuse process for innovation is idea development. In this action, creators incorporate the ideas reused to form a final creation. The key element in this action is the integration of the reused ideas. The performance of this action could be evaluated by the metaknowledge expressed about the integration: whether the creators have explicitly explained how they integrate the selected ideas and how well the aggregated information is related to the selected ideas.

Previous studies suggest that metaknowledge about an idea, such as describing the context and credibility of the source affects a creator's reuse decision, perhaps by reassuring the creator (Markus, 2001; Majchrzak et al., 2004). We extend this idea and hypothesize that including metaknowledge about how creators integrate the reused knowledge has a positive influence on the generativity of this creation. That is, societal challenge solutions such as proposals related to climate change are complex artifacts. They are composed from other artifacts. When composition takes place, the composed elements are in relation to each other and to an overall goal. The extent to which the rationale for picking a set of components and articulating how they contribute to each other is the extent to which the proposal designers have provided integration metaknowledge. Work on design rationale has articulated the importance of such rationales to help designers think through their problem (Carroll & Rosson, 2003; Wang, Farooq, & Carroll, 2013). Other work on metaknowledge has noted that metaknowledge helps others understand design artifacts (Choi, Lee, & Yoo, 2010; Leonardi, 2014; Rico, Sánchez-Manzanares, Gil, & Gibson, 2008). Thus, more integration metaknowledge might be associated with more reuse for two reasons. It might reflect a more thoughtful design process leading to a better design. And it might serve as a signal to others that the design is in fact well considered. In addition, metaknowledge about integration could serve as a boundary object: the more integration metaknowledge the more effective it could help the communication among users across different boundaries (Carlile, 2002; Mark, Lyytinen, Bergman, 2007; Nicolini, Mengis, & Swan, 2012; Star & Griesemer, 1989). This leads to our hypothesis related to the performance of the third action in the knowledge reuse process model—the idea development hypothesis:

H3: A remix that encodes more integration metaknowledge is more likely to be reused.

3 Research Design

To answer our research question and test our hypotheses, we conducted an empirical study using data from an online innovation community, Climate CoLab, which addresses an important societal challenge—global climate change (Malone, Laubacher, & Dellarocas, 2010). In Climate CoLab, members collaborate with each other to enter contests by creating proposals. These contests aim to solve multiple sub-problems for global climate change such as carbon pricing, energy supply, and transportation (Figure 1). So far, the website has nearly 75,000 registered members and over 500,000 visitors.

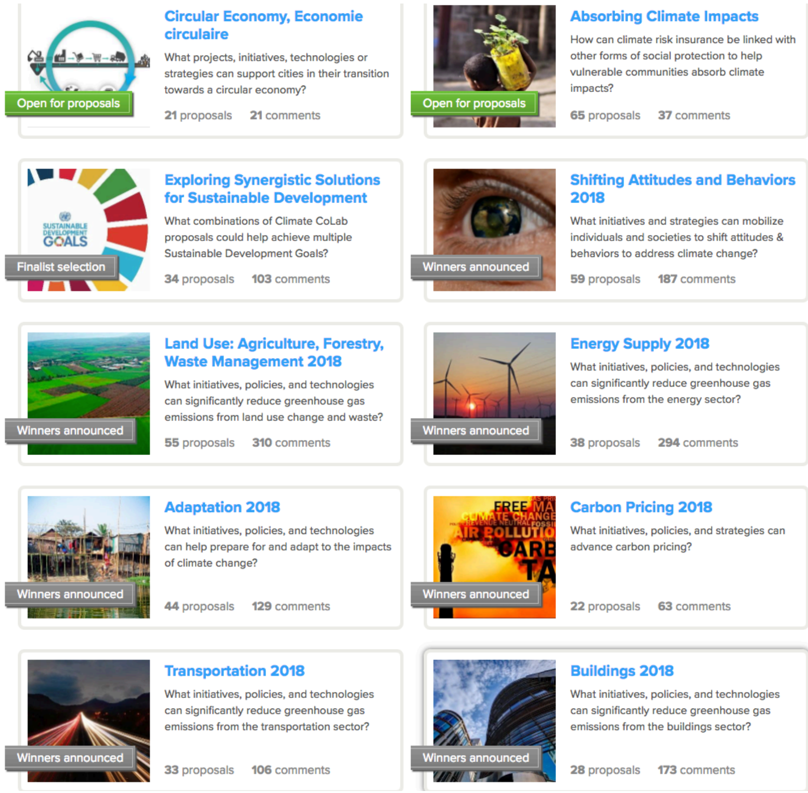

Figure 1. Contests on the Climate CoLab Website

We choose Climate CoLab for the following reasons: First, this online community aims to harness collective intelligence using the remixing mechanism to solve an important societal challenge. Second, the goal of this online community is to generate innovative proposals, which is a form of innovation. Therefore, each innovative proposal is considered as an innovation in this study. Third, members in this community have different backgrounds and geographic locations. Most community members generate diverse ideas that are previously unknown to each other. This provides the exploration condition for knowledge reuse for innovation (Armbrecht et al., 2001). Fourth, Climate CoLab encourages knowledge reuse for innovation and has incorporated this approach into its contest design (Malone et al., 2017). There are three main types of contests in the Climate CoLab website: basic, regional, and global. Proposals in a regional contest are encouraged to reuse proposals submitted to the basic contests, while proposals in a global contest are required to reuse proposals from the regional contests (Figure 2). In addition, proposal creators are required to provide links to the proposals they have used. This reuse information helps us identify all reuse relationships and build up the proposal reuse network (Figure 3). More importantly, we can quantify and examine the generativity of remixes.

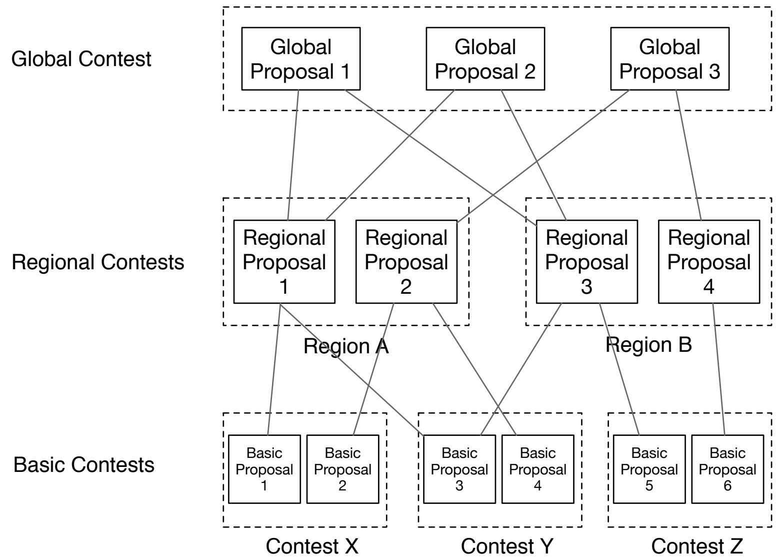

Figure 2. Climate CoLab Proposal Reuse Structure

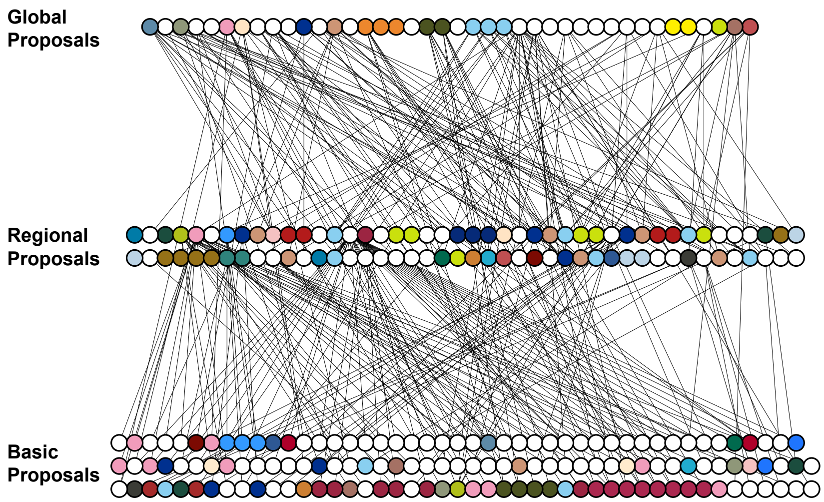

Figure 3. Proposal Reuse Network in the Climate CoLab Website

Note: Each node represents a proposal and is colored based on their owner. Proposals that share the same owner have the same color. If a proposal owner has only created one proposal, the proposal is colored in white.

3.1 Data Collection

In this empirical study, given our focus on the generativity of remixes, we analyzed proposals in the regional contests because they both reuse knowledge (proposals in the basic contests) and have been reused by others (proposals in the global contest). We collected all the proposals in the 2015 regional contests and global contest on the Climate CoLab website.

3.2 The Dependent Variable: Generativity

In this study, we evaluate the reuse for innovation by the generativity of a proposal. Therefore, our dependent variable for all hypotheses in this paper is the generativity of a remix: how many times a remix has been reused. We measure generativity by counting the number of times a regional proposal has been reused in global proposals. In each global proposal, there is a section where proposal creators provide the links to the regional proposals they have reused. As shown in Figure 4 global proposal creators explicitly state which proposal they have reused from each regional contest. Words in blue are hyperlinks to the listed regional proposal. We collected information in this section for all global proposals to calculate the generativity of each regional proposal. For example, if a regional proposal is reused in three different global proposals, the generativity of this regional proposal is recorded as 3.

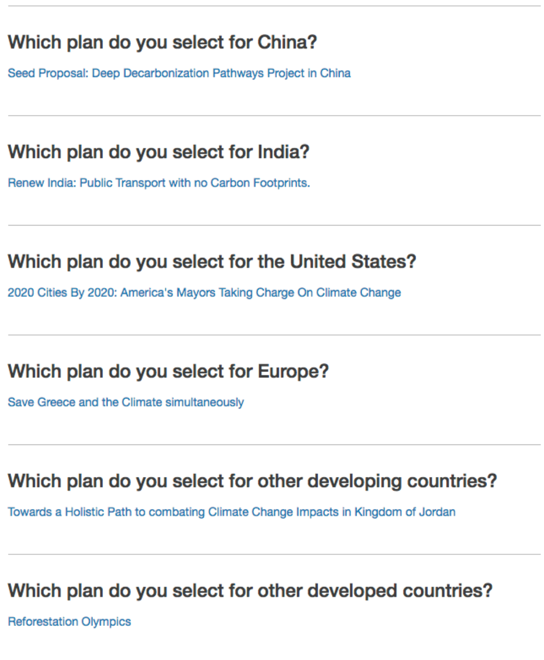

Figure 4. Links to Regional Proposals in a Section of Global Proposals

3.3 Independent Variables

3.3.1 Independent Variable for H1

Proposal Topic Prevalence is our independent variable for H1. To determine the prevalence of a proposal, we calculated the proposal topic prevalence for each regional proposal to see if the creator has included knowledge that is prevalent in the contest. A proposal with a high proposal topic prevalence score includes either fundamental topics that are commonly discussed or popular topics within the contest that are familiar to other community members, or both.

One of the most popular approaches in describing which topics are covered in a document (which is the proposal in our case) and what a document is about is to describe the document with relevant terms that represent semantic concepts important to the document. This is an ontology-based approach (Zouaq, Gasevic, & Hatala, 2011). Ontologies are defined as the explicit formal specifications of the terms in a domain and relations among them (Gruber, 1993), and hence, they tend to encompass only a single domain corpus (i.e. medicine, wine etc.). The domain corpus must have a good coverage of domain knowledge for generating a comprehensive ontology. Existing works have exploited different sources as corpus for ontologies. Some early works used manually established corpora by domain experts (Baker, Filmore, & Lowe, 1998) or corpora derived from books, magazines, and news organizations automatically or semi-automatically (Khan, Luo, & Yen, 2002). But these corpora are not so easy to extend because knowledge contained in the corpora is fixed in time and by region, and cannot be easily updated. Later works have utilized web-based corpora such as DBpedia and Wikipedia (Gabrilovich & Markovitch, 2007; Yu, Thom, & Tam, 2007) as vast amounts of highly organized human knowledge is encoded in those corpora and they undergo constant development (Kane, 2011; Keegan, Gergle & Contractor, 2013) so the breadth and depth steadily increase over time.

Thus, in order to identify the topics of each proposal, we opted to use Wikipedia because it is currently the largest knowledge corpus on the Web. Wikipedia is available in dozens of languages, while its English version is the largest of all with 3.6 billion words in over 3.6 million articles –which is 60 times as many as the next largest English-language encyclopedia, Encyclopedia Britannica ("Wikipedia: Size comparisons", 2019, para. 3). To automatically identify the topics covered in a proposal, we extracted the plain text of each proposal and employed a two-step process developed by Genc, Mason, and Nickerson (2013).

In the first step, we identified candidate concepts within the main text of a proposal and mapped them to corresponding Wikipedia pages. To extract these concepts, we first removed stop-words and punctuation marks, and segmented the main text into n-grams in a sliding window fashion. Then we searched for the n-grams in Wikipedia title search. In Wikipedia, all pages are tagged with categories that they belong to and these categories are linked to each other in a network graph structure. For each proposal, we recorded all categories listed in the corresponding Wiki pages.

In the second step, we used the category network to determine a common set of high-level topics based on the Wiki pages identified in the first step (Figure 5). At the time of our analysis, Wikipedia included 28 main topic categories (“Category: Main topic classifications”, 2019, para. 5). When we traversed five (or more) levels of the category graph, most of our initial topics hit one of those main topic categories and led all of the proposals to share a topic and be connected to each other. Thus, we stopped the traversal at level four for each proposal and recorded all identified categories as the topics it covers.

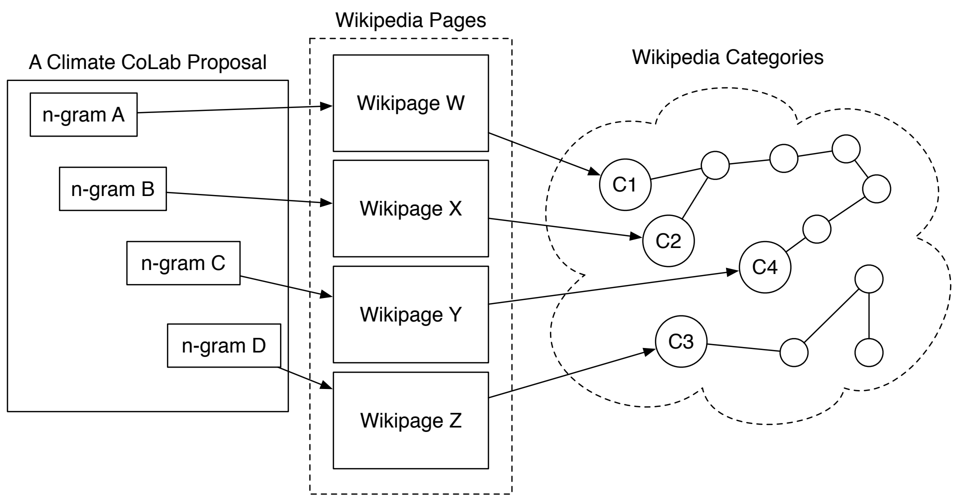

Figure 5. Identifying Topics in a Regional Proposal

Then we calculated the topic prevalence score for each topic within a contest. The topic prevalence score is the degree of topic node divided by the maximum possible topic degree. For example (Figure 6), in contest X, there are four proposals and topic 2 is covered in one proposal. The topic prevalence score for topic 2 is then calculated as 1 divided by 4. After this calculation, for each proposal, we computed a proposal topic prevalence score by summing up the topic prevalence score of all the topics presented in a proposal. For example, in proposal B, two topics are covered, topic 1 and topic 2. Thus, the proposal topic prevalence score of this proposal is 1.

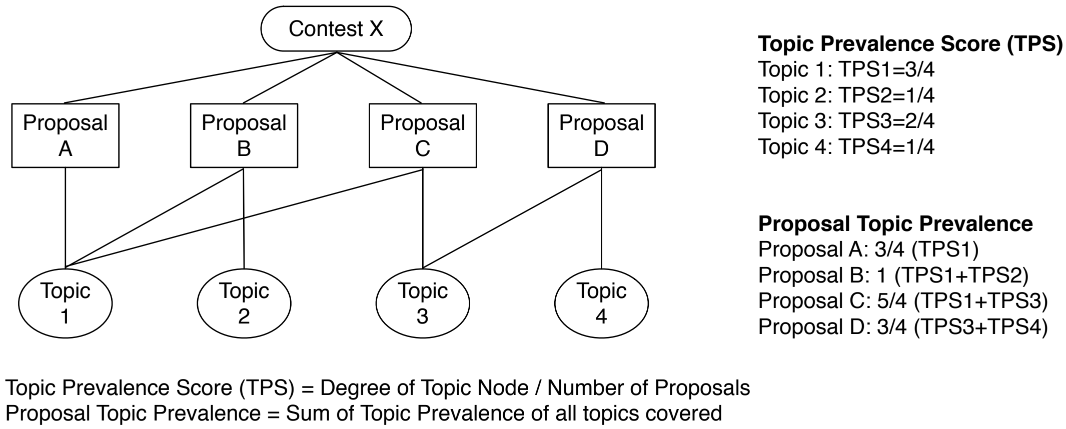

Figure 6. Calculating Proposal Topic Prevalence

3.3.2 Independent Variable for H2

Number of High-quality Proposals Reused is our independent variable for H2. In Climate CoLab contests, there is a special section in each regional proposal where proposal creators can create hyperlinks to the basic proposals they have reused and write down how they have incorporated these proposals (Figure 7). We analyzed the information in this section for each regional proposal to identify the basic proposals that have been reused. Then we checked each basic proposal's expert evaluation to determine its quality.

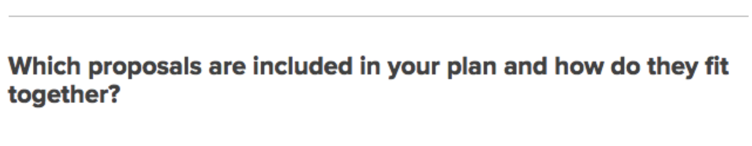

Figure 7. A Section in Regional Proposals that Provides Reuse Links and Integration Metaknowledge

In Climate CoLab, each proposal in a basic contest is rated by a group of experts. These experts evaluate proposals based on their quality and advance high-quality proposals to enter the semi-final phase for further development. If a basic proposal has been selected by CoLab experts as semi-finalist in that basic contest (Figure 8), we counted this proposal as a high-quality proposal. Then we calculated the total number of high-quality proposals reused by a regional proposal.

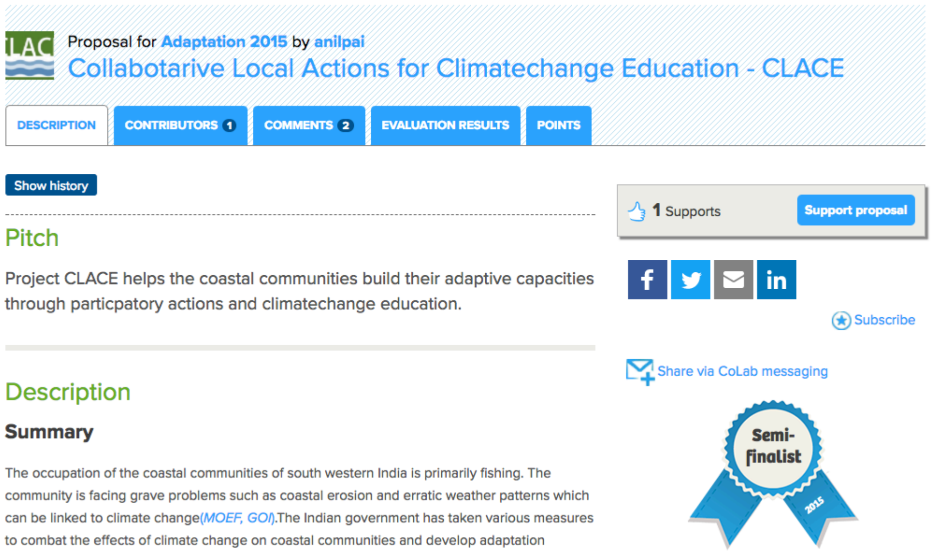

Figure 8. An Example of a High-Quality Basic Proposal

3.3.3 Independent Variable for H3

Integration metaknowledge is our independent variable for H3. To understand how metaknowledge affects generativity, we extracted information from the section shown in Figure 7 for each regional proposal and automatically coded all regional proposals. We extracted the content from this integration section (e.g. Figure 9) and analyzed the hyperlinks (words in blue) and plain text (words in black). Hyperlinks indicates whether a regional proposal has integrated basic subproposals and plain text contains information about the integration metaknowledge: whether metaknowledge has been included in a regional proposal; and if so, the content of the metaknowledge. In addition, we extracted the text from the summary section (e.g. the summary section shown in Figure 8) from all basic proposals that had been reused by a regional proposal and conducted a text similarity analysis using Jaccard similarity between the integration section of the regional proposal and the summary section of all basic proposals it reuses. Therefore, the integration metaknowledge score calculated for each regional proposal includes two parts: 1) whether the regional proposal includes direct links to the basic proposals it reuses, and 2) whether the regional proposal integration section covers similar text to the basic proposals it reuses as indicated by the Jaccard similarity. Thus, besides indicating the existence of metaknowledge, the integration metaknowledge score also reveals the amount of integration metaknowledge: the higher the score, the more coverage of the metaknowledge. The more coverage in the metaknowledge, we reason, the better the proposal authors have articulated how all the components of the proposal related to each other and the proposal as a whole.

Which proposals are included in your plan and how do they fit together?

We start right away with ClimateCoin, a new cryptocurrency which mints new coins to anyone who pays to offset carbon. It's implemented on a scriptable cryptocurrency platform called Ethereum, and can offset carbon via existing providers of voluntary carbon offsets.

In particular, America has vast farmlands, which if properly managed could sequester a large amount of carbon. One well-known method for reliably sequestering large amounts of carbon is to burn organic materials in low oxygen, producing charcoal. Once ground and worked into soil, it remains stable for centuries, and in many environments improves the fertility of soil. Carbon-Negative Biochar Economies suggests using cryptocurrency to fund biochar projects.

Figure 9. The Integration Section of an Example Regional Proposal

3.4 Control Variables

In this study, we controlled for the factors related to the proposal contributors and the CoLab reuse structure (Figure 10). The control variables associated with the proposal contributors are the *Number of Contributors*, the *Proposal Owner's Tenure, and Owner Network Control.* The number of contributors represents the number of participants who have edited the proposal, which might have an influence on the generativity of a remix because of preferential attachment within the user network (Barabási & Albert, 1999). Prior literature suggests that a creator's experience is important in generating reusable creations (Lim, 1994; Kyriakou et al., 2017). As there is no information about a creator's year of experience outside of the community, we measured proposal owners experience by their membership on the Climate CoLab website: The number of days they have been a CoLab member before creating the proposal. Such a community tenure variable has been used in many other studies of online communities (Bateman, Gray, & Butler, 2011; Faraj, Kudaravalli, & Wasko, 2015; Kyriakou et al., 2017; Mein Goh, Gao, & Agarwal, 2016). Some owners have created more than one regional proposal, which might lead to a user-network effect. Therefore, we also controlled for this factor by Owner Network Control variable: If the owner of a regional proposal has created more than one regional proposal, we mark the proposal as 1, otherwise 0.

The control variables associated with the CoLab reuse structure are the *Sequence of Proposal Creation* and the fixed effect of the *Regional Contest*. The sequence of proposal creation is a time-related control variable that indicates which proposals were created early and which were created later. Proposals that were created earlier have greater potential to be seen by other community members as they have been on the website for a longer time. Global proposals on the Climate CoLab website are required to reuse only one proposal from each regional contest. Since each regional contest varies in the number of entries, proposals in different regional contests may face different levels of competition. Thus, we also controlled for this fixed effect.

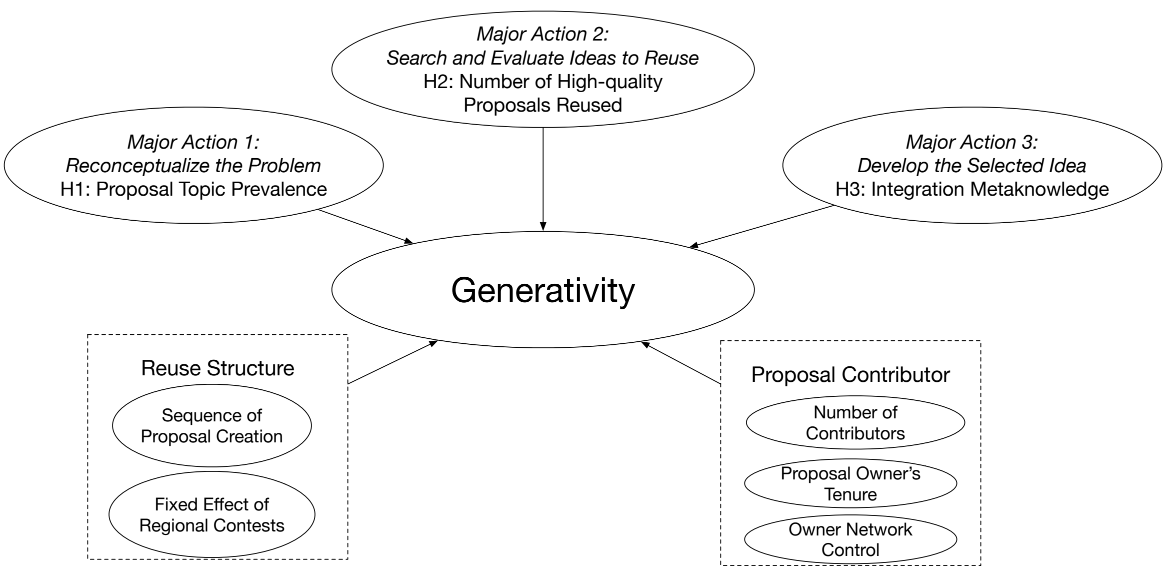

Figure 10. Research model

4 Analysis and Results

To test our hypotheses, we created a series of Poisson regression models. All regression models have the same dependent variable and control variables. We followed Green's (1991) formula to determine the number of observations for our regression models. The descriptive statistics of all variables are listed in Table 1. We have standardized all the independent variables and control variables in all regression models. We have also conducted a post-hoc power analysis for each model. The results for both the Poisson regression and power analysis are presented (Table 2). The correlation table, and multicollinearity check can be found in the Appendix (Table A2 & A3).

Table 1. Descriptive Statistics of All Variables

Variable	Obs	Mean	Std.Dev.	Min	Median	Max
Generativity	81	1.148	1.476	0	1	5
Proposal Topic Prevalence	81	170.704	155.304	4	126	889
Number of High-quality Proposals Reused	81	0.975	2.392	0	0	11
Integration Metaknowledge	81	0.312	0.480	0	0	1.118
Number of Contributors	81	1.593	2.072	1	1	14
Sequence of Proposal Creation	81	9.481	7.321	1	7	29
Proposal Owner's Tenure	81	260.815	407.472	0	52	1698
Owner Network Control	81	0.469	0.502	0	0	1

Table 2 shows five Poisson regression models. Model 1 is a basic regression model with all control variables. The number of contributors, the owner network control, and the sequence of proposal creation have no significant influence on generativity. The proposal owner's tenure is positively associated with the generativity of a remix: A regional proposal created by experienced users is more likely to be reused by global proposals.

Model 2 tests the relationship between the proposal topic prevalence with the generativity of a regional proposal. The result shows that proposal topic prevalence has a positive influence on the generativity of a remix, which suggests that proposals that include more prevalent topics are more likely to be reused in the future. Therefore, hypothesis 1 is supported.

In model 3, we study both the quantity and quality of proposals that have been reused. We examined if the number of high-quality basic proposals reused in a regional proposal has a U-shape relationship with the generativity of this proposal. The result shows that there is no significant curvilinear relationship between the two variables. Therefore, hypothesis 2 is not supported.

Model 4 is a test on the integration metaknowledge. In this model, we examined the influence of integration metaknowledge on the generativity of a regional proposal. The result of model 4 suggests that encoding more integration metaknowledge has a positive influence on the generativity of a remix: the higher the coverage of the metaknowledge, the more likely the regional proposal will be reused by global proposals. Therefore, hypothesis 3 is also supported. Based on our analysis and results, we summarize our findings in table 3.

Table 2. Poisson Regression Model for Generativity

		Model 1	Model 2	Model 3	Model 4
	Constant	-0.822*	-0.883*	-0.706*	-0.886*
Control	Number of Contributors	0.099	0.084	0.120	0.153
	Sequence of Proposal Creation	-0.099	-0.024	-0.023	-0.129
	Proposal Owner's Tenure	0.449***	0.487***	0.541***	0.450***
	Owner Network Control	-0.020	-0.096	-0.177	-0.041
Fixed Effect (Contest)	1303007	1.262**	0.864	0.596	0.665
	1302013	1.348**	0.368	-0.038	0.148
	1302019	0.549	0.827	0.651	0.908
	1302025	0.875*	0.954*	0.725	0.618
	1302031	1.177**	0.983*	0.718	0.923*
H1	Proposal Topic Prevalence		0.478***	0.557***	0.528***
H2	Number of High-quality Proposals Reused			0.226	-0.432
	Number of High-quality Proposals Reused (squared)			-0.373	0.131
H3	Integration Metaknowledge				0.366**
Number of Observations		81	81	81	81
McFadden's R-Square		0.208	0.277	0.285	0.310
Power Analysis: Effect Size		0.26	0.38	0.40	0.45
Power (sig.level=0.05)		0.88	0.97	0.97	0.98

***p<0.001; **p<0.01; *p<0.05

Table 3. Summary of Findings

Hypotheses	Results
H1: A remix containing more prevalent topics is more likely to be reused.	Supported
H2: The number of high-quality ideas reused in a remix has a U-shape relationship with the generativity of this remix.	Not Supported
H3: A remix that encodes more integration metaknowledge is more likely to be reused.	Supported

5 Discussion

This empirical study explored the relationship between the three major actions in the knowledge reuse process for innovation and the generativity of the innovative outcome created for addressing societal challenges. As shown in Table 3, our first hypothesis H1 is supported. This finding suggests that the decision a creator makes when reconceptualizing the problem is essential to the generativity of a remix. Addressing the problem with prevalent topics will lower the barrier for future adaptation and thus increases the reusability of a remix.

Previous studies suggested that the number of previous works reused in a remix and the generativity of this remix follows a U-shape relationship. However, in our study this hypothesis H2 is not supported. This might be related to the difference in the media form of the creations. Previous studies were conducted using data from either ccMixter or Scratch; the former generates music remixes and the latter generates projects using a drag and drop programming language. In both communities, the knowledge reuse is direct and explicit. Creators in these communities are allowed and encouraged to embed the reused work or part of the work to serve a specific need. For example, in ccMixter creators can directly incorporate a piece of drumbeat for the background in a music remix. Meanwhile, creators in Scratch can also fork a piece of code to achieve a function in their remixes. On the other hand, the knowledge reuse in Climate CoLab is quite different. Like citing literatures in academic writing, proposal creators wouldn't directly reuse sentences from the ideas they are reusing; instead, the reuse is more likely to happen on the idea level. This suggests that the number of high-quality proposals reused may be less important than the inter-relationship among reused ideas.

Integration is the key component when developing an idea by reusing knowledge. Our result supports the argument that encoding more integration metaknowledge increases the generativity of a remix (H3). Including integration metaknowledge and providing better coverage of topics in the component artifacts of the proposal signal the quality of integration as it shows that the creator has fully understood the reused content and developed a clear logic when integrating the knowledge. In addition, integration metaknowledge serves as an index that may help people better understand the structure of the idea and the connections between the knowledge reused in a remix, and hence increases the remix's potential for adaptation in the future. Especially for members in online innovation communities who mostly participate in their spare time and are limited in the time they can spend on a creation (Paulini, Maher, & Murty, 2014; Zhang, Hahn, & De, 2013), integration metaknowledge creates a quick access to knowledge and improves the efficiency of knowledge reuse.

5.1 Theoretical Contributions

The theoretical contributions of our study are twofold. First, it contributes to the better usage of crowdsourcing for tackling societal challenges by explicating the role of remixing in leveraging the wisdom of the crowd. For complex tasks like solving societal challenges, remixing can better harness collective intelligence and motivate more comprehensive creations as it encourages collaboration and integration which helps to break the knowledge boundary that exists in most crowdsourcing methods. Second, our study also contributes to the knowledge reuse literature as it is one of the first approaches to examine the relationship between the KRI process and the generativity of the innovative outcome in online settings. The analytical approach in our study deepens our understanding on the impact of the performance of the three major actions and shows that incorporating prevalent topics when reconceptualizing the problem and encoding more integration metaknowledge when developing the integrated idea can increase the generativity of the final creation.

5.2 Implications for Knowledge Creators and Platform Designers

Our study also has practical implications as it can help both knowledge workers and online innovation community designers better harness the wisdom of crowd through remixing to address societal challenges. Our findings suggest that creators can adopt certain strategies to increase the reusability of their creations when they build off previous artifacts. They can widely browse the previous creations and incorporate prevalent topics when reconceptualizing the problem; they can also think through their integration rationale carefully and provide an explicit and comprehensive summary through integration metaknowledge.

Increasing the generativity of remixes is beneficial to not only generating individual innovate ideas, but also maintaining an active community: It encourages more collaboration and communication among community members and potentially leads to more user activity. Thinking along these lines, another implication of our study is that designers of online innovation communities might consider introducing features and applying analytics to help creators perform better in each step of the knowledge reuse process for innovation.

One essential factor that influences the creators when they are reconceptualizing the problem is their knowledge of the solution space. It is almost impossible to adopt a good strategy if they don't know what knowledge is available. Due to the constantly growing number of artifacts in online communities, it can be very difficult to browse all submissions. Therefore, it might be helpful to conduct large scale text analytics and incorporate a design feature that automatically detects and summarizes the solution space for community members. For example, creating an idea heat map or an idea network may be a good way to help people create an overall picture of the current solution space.

When searching and evaluating ideas to reuse, creators face a different environment in online communities. These communities tend to provide a more open environment that allows all community members to see each other's creations. Creators in these communities can easily access many resources. However, this often leads to information overload. The way to support this action in online communities is not maximizing the number of available artifacts but streamlining search. Therefore, we conjecture that applying text analytics and similarity calculations to develop tools like recommender systems (e.g. Siangliulue, Chan, Huber, Dow, & Gajos, 2016) can improve the efficiency of search which will in turn lead to increased generativity.

When people move to the last action – developing the idea – they sometimes reach to the source of the knowledge reused to better understand the knowledge and thus better integrate the knowledge. Making this communication easier is essential for the performance of integration. Currently, many online communities have already incorporated a within-community email system. To help community members communicate in a timely fashion, it might also be worthwhile to consider including an instant messaging system. In addition, expression of integration metaknowledge might be motivated through templates that encourage short summaries of all artifacts; and encourage short rationales to explain why some sets of artifacts were reused in a particular work. The short summaries may encourage recombination, and the rationales may give confidence to others that the work has solid foundations. In addition, large-scale text analytics can be applied to help improve the quality of integration metaknowledge: the coverage calculation described above could be automatically calculated and provided to users to encourage revisions and improvements.

5.3 Limitations and Future Work

Adopting remixing does not guarantee the success of an online innovation community. Large-scale problem solving requires the task to be well divided and solutions to be well combined (Kittur et al., 2013; Malone, 2018). The platform studied in this paper, Climate CoLab, has fulfilled these requirements. The task was well divided by experts based on the topic and geographic location; and the remixing structure was also well designed as it allows multiple inheritance which encourages both diversity and integrity (Malone et al., 2017). These strengths of the platform design have greatly helped our analytical approach to address the research question. Yet, our analytical methods may not be applied to all online remixing communities. For platforms that do not provide such traceable reuse structures or section for integration metaknowledge, other analytical methods may be adopted for automating the analyses. And, given this study was observational, and given online communities often have feedback loops that create endogeneity issues, future research might use experiments to better understand the causal factors that drive quality.

Our study subject is an on online innovation community that aims to solve global climate change issues. We suggest that remixing can be used in a similar fashion to utilize crowdsourcing creativity for other societal challenges. But are our findings generalizable to other types of innovation communities? It is possible to look at other online open innovation communities such as GitHub (Dabbish et al., 2012) and Scratch (Resnick et al., 2009) to see if the reuse processes in these communities are similar. The proposals in our study are text-based creations. It is possible to perform future studies with sites that allow for remixing in different media forms.

Our study also suggests a few additional research questions that can be examined in the future: Are there any relational variables that influence the generativity of a remix in online communities? For example, does the creator's position in the user network affect the generativity of his/her creation? This can be examined via network analysis, in particular checking for network autocorrelation. In addition, the co-occurrence of proposals that are being reused can be analyzed: What kind of proposals and proposal topics are more likely to be reused together? And how does that affect the generativity and quality of the higher-level proposal? We briefly mentioned that online communities provide more open environments than organizations. What are the major differences between knowledge reuse processes in online communities and in organizations? Future studies might use qualitative analysis to explore and reveal these major differences.

6 Conclusion

Reusing knowledge to generate innovative ideas for societal challenge is a complex task. This study examined the relationship between the knowledge reuse process model and the generativity of the outcome. Our findings suggest that the major actions in this process model directly influence the generativity of a remix. Knowledge workers can adopt varied strategies to generate reusable artifacts, and designers of online communities can build tools that make exploration of artifacts easier in order to encourage recombination. They can also apply multiple analytical methods to build tools that help users think through their reasoning for reusing combinations of artifacts. Rationales may be helpful for both the integrator and the future creator who may be reusing the integrated package. Together, a knowledge repository can be built up for solving societal challenges. That is, creators communicate with themselves, and also with the prospective remixers of their work. They create a kind of structured memory for their future selves, and for their future community.

Acknowledgement

This material is based upon work supported by the National Science Foundation under grants 1909803, 1717473, 1442840, and 1422066.

References

Afuah, A., & Tucci, C. L. (2012). Crowdsourcing as a solution to distant search. *Academy of Management Review, 37*(3), 355-375.

Alavi, M., & Leidner, D. E. (2001). Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly*, *25*(1), 107-136.

Argote, L. (2012). *Organizational learning: Creating, retaining and transferring knowledge.* Springer Science & Business Media.

Armbrecht Jr, F. R., Chapas, R. B., Chappelow, C. C., Farris, G. F., Friga, P. N., Hartz, C. A., McIlvaine, E., Postle, S. R., & Whitwell, G. E. (2001). Knowledge management in research and development. *Research Technology Management, 44*(4), 28-48.

Baker, C. F., Fillmore, C. J., & Lowe, J. B. (1998). The berkeley framenet project. *Proceedings of the 17th International Conference on Computational Linguistics-Volume 1* (pp. 86-90). Stroudsburg, PA.

Banker, R. D., Bardhan, I., & Asdemir, O. (2006). Understanding the impact of collaboration software on product design and development. *Information Systems Research, 17*(4), 352-373.

Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*(5439), 509-512.

Bateman, P. J., Gray, P. H., & Butler, B. S. (2011). Research note—the impact of community commitment on participation in online communities. *Information Systems Research, 22*(4), 841-854.

Bayus, B. L. (2013). Crowdsourcing new product ideas over time: An analysis of the Dell IdeaStorm community. *Management Science, 59*(1), 226-244.

Bjelland, O. M., & Wood, R. C. (2008). An inside view of IBM's' Innovation Jam'. *MIT Sloan Management Review, 50*(1), 32.

Boh, W. F. (2008). Reuse of knowledge assets from repositories: A mixed methods study. *Information & Management, 45*(6), 365-375.

Brabham, D. C. (2008). Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence, 14*(1), 75-90.

Brabham, D. C. (2009). Crowdsourcing the public participation process for planning projects. *Planning Theory, 8*(3), 242-262.

Brabham, D. C., Ribisl, K. M., Kirchner, T. R., & Bernhardt, J. M. (2014). Crowdsourcing applications for public health. *American Journal of Preventive Medicine, 46*(2), 179-187.

Brown, J. S., & Duguid, P. (1991). Organizational Learning and Communities-of-Practice: Toward a Unified View of Working, Learning, and Innovation. *Organization Science, 2*(1), 40-57.

Callaghan, C. W. (2014). Crowdfunding to generate crowdsourced R&D: The alternative paradigm of societal problem solving offered by second generation innovation and R&D. *The International Business & Economics Research Journal (Online), 13*(6), 1599.

Carlile, P. R. (2002). A pragmatic view of knowledge and boundaries: Boundary objects in new product development. *Organization Science, 13*(4), 442-455.

Carroll, J. M., & Rosson, M. B. (2003). Design rationale as theory. In J. M. Carroll (Eds.), *HCI models, theories and frameworks: Toward a multidisciplinary science* (pp. 431-461). Burlington, MA: Morgan Kaufmann Publishers.

Category: Main topic classifications. (2019). In Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Category:Main_topic_classifications.

Certoma, C., Corsini, F., & Rizzi, F. (2015). Crowdsourcing urban sustainability. Data, people and technologies in participatory governance. *Futures, 74*, 93-106.

Cheliotis, G., & Yew, J. (2009). An analysis of the social structure of remix culture. *Proceedings of the Fourth ACM International Conference on Communities and Technologies* (pp. 165-174). University Park, PA.

Cheliotis, G., Hu, N., Yew, J., & Huang, J. (2014). The antecedents of remix. *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 1011-1022). Bartimore, MD.

Chesbrough, H. W. (2006). *Open innovation: The new imperative for creating and profiting from technology.* Harvard Business Press.

Chesbrough, H., & Crowther, A. K. (2006). Beyond high tech: early adopters of open innovation in other industries. *R&d Management, 36*(3), 229-236.

Cheung, P. K., Chau, P. Y., & Au, A. K. (2008). Does knowledge reuse make a creative person more creative? *Decision Support Systems, 45*(2), 219-227.

Chewar, C. M., & McCrickard, D. S. (2005). Links for a human-centered science of design: Integrated design knowledge environments for a software development process. *Proceedings of the 38th Annual Hawaii International Conference on Systems Sciences* (pp. 256c-256c). Big Island, HI.

Chiu, C. M., Liang, T. P., and Turban, E. (2014). What can crowdsourcing do for decision support? *Decision Support Systems, 65*, 40-49.

Choi, S. Y., Lee, H., & Yoo, Y. (2010). The impact of information technology and transactive memory systems on knowledge sharing, application, and team performance: a field study. *MIS Quarterly, 34(*4), 855-870.

Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., Leaver-Fay, A., Baker, D & Popović, Z. (2010). Predicting protein structures with a multiplayer online game. *Nature, 466*(7307), 756.

Dabbish, L., Stuart, C., Tsay, J., & Herbsleb, J. (2012). Social coding in GitHub: transparency and collaboration in an open software repository. *Proceedings of the 15th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 1277-1286). Seattle, WA.

Dasgupta, S., Hale, W., Monroy-Hernández, A., & Hill, B. M. (2016). Remixing as a pathway to computational thinking. *Proceedings of the ACM 2016 Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 1438-1449). San Francisco, CA.

Durcikova, A., & Fadel, K. J. (2016). Knowledge sourcing from repositories: The role of system characteristics and psychological climate. *Information & Management, 53*(1), 64-78.

Elmquist, M., Fredberg, T., & Ollila, S. (2009). Exploring the field of open innovation. *European Journal of Innovation Management, 12*(3), 326-345.

Eservel, U. Y. (2014). IT-enabled knowledge creation for open innovation. *Journal of the Association for Information Systems, 15*(11), 805.

Estellés-Arolas, E., & González-Ladrón-De-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science, 38*(2), 189-200.

Faniel, I. M., & Majchrzak, A. (2007). Innovating by accessing knowledge across departments. *Decision Support Systems, 43*(4), 1684-1691.

Faraj, S., Jarvenpaa, S. L., & Majchrzak, A. (2011). Knowledge collaboration in online communities. *Organization Science, 22*(5), 1224-1239.

Faraj, S., Kudaravalli, S., & Wasko, M. (2015). Leading Collaboration in Online Communities. *MIS Quarterly, 39*(2), 393-412.

Flath, C. M., Friesike, S., Wirth, M., & Thiesse, F. (2017). Copy, transform, combine: Exploring the remix as a form of innovation. *Journal of Information Technology,* 32(4), 306-325.

Gabrilovich, E., & Markovitch, S. (2007). Computing semantic relatedness using wikipedia-based explicit semantic analysis. *Proceedings of IJCAI* (pp. 1606-1611). Hyderabad, India.

Genc, Y., Mason, W. A., & Nickerson, J. V. (2013). Classifying Short Messages using Collaborative Knowledge Bases: Reading Wikipedia to Understand Twitter. *Workshop Concept Extraction Challenge Proceedings* (pp. 50-53). Rio de Janeiro, Brazil.

George, G., Howard-Grenville, J., Joshi, A., & Tihanyi, L. (2016). Understanding and tackling societal grand challenges through management research. *Academy of Management Journal, 59*(6), 1880.

Gossel, B. M., Brüntje, D., & Will, A. (2016). Crowd and society: outlining a research programme on the societal relevance and the potential of crowdfunding. In D. Brüntje & O. Gajda (Eds.), *Crowdfunding in Europe* (pp. 55-67). Switzerland: Springer International.

Grant, R. M. (1996). Prospering in Dynamically-Competitive Environments: Organizational Capability as Knowledge Integration. *Organization Science, 7*(4), 375-387.

Green, S. B. (1991). How many subjects does it take to do a regression analysis. *Multivariate Behavioral Research, 26*(3), 499-510.

Grimm, R., Fox, C., Baines, S., & Albertson, K. (2013). Social innovation, an answer to contemporary societal challenges? Locating the concept in theory and practice. *Innovation: The European Journal of Social Science Research, 26*(4), 436-455.

Gruber, T. (1993). What is an ontology? *WWW Site http://www-ksl.stanford.edu/kst/what-is-an-ontology.html*.

Head, B. W., & Alford, J. (2015). Wicked problems: Implications for public policy and management. *Administration & Society, 47*(6), 711-739.

Hill, B. M., & Monroy-Hernández, A. (2013). The remixing dilemma: The trade-off between generativity and originality. *American Behavioral Scientist, 57*(5), 643-663.

Hoegg, R., Martignoni, R., Meckel, M., & Stanoevska-Slabeva, K. (2006). Overview of business models for Web 2.0 communities. In K. Meißner & M. Engelien (Eds.), *Virtuelle Organisation und Neue Medien 2006* (pp. 33-50). Dresden: TUDpress.

Howe, J. (2006). The rise of crowdsourcing. *Wired magazine, 14*(6), 1-4.

Howe, J. (2008). *Crowdsourcing: Why the power of the crowd is driving the future of business*. New York: Crown Publishing Group.

Hutcheon, L. (2012). *A theory of adaptation*. Abingdon, OX: Routledge.

Jarmolowicz, D. P., Bickel, W. K., Carter, A. E., Franck, C. T., & Mueller, E. T. (2012). Using crowdsourcing to examine relations between delay and probability discounting. *Behavioural Processes, 91*(3), 308-312.

Jarvenpaa, S., & Standaert, W. (2018). Digital probes as opening possibilities of generativity. *Journal of the Association for Information Systems, 19*(10), 982-1000.

Kane, G. C. (2011). A multimethod study of information quality in wiki collaboration. *ACM Transactions on Management Information Systems (TMIS), 2*(1), 4.

Kankanhalli, A., Lee, O. K. D., & Lim, K. H. (2011). Knowledge reuse through electronic repositories: A study in the context of customer service support. *Information & Management, 48*(2-3), 106-113.

Keegan, B., Gergle, D., & Contractor, N. (2013). Hot off the wiki: Structures and dynamics of Wikipedia’s coverage of breaking news events. *American Behavioral Scientist, 57*(5), 595-622.

Khan, L., Luo, F., & Yen, I. L. (2002). Automatic ontology derivation from documents. *Proceedings of the 15th Conference on Advanced Information Systems Engineering (CAISE)*. Klagenfurt/Velden, Austria.

Khatib, F., DiMaio, F., Cooper, S., Kazmierczyk, M., Gilski, M., Krzywda, S., Zabranska, H., Pichova, I., Thompson, J., Popović, Z., & Jaskolski, M. (2011). Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nature Structural and Molecular Biology, 18*(10), 1175.

Khedhaouria, A., & Jamal, A. (2015). Sourcing knowledge for innovation: knowledge reuse and creation in project teams. *Journal of Knowledge Management, 19*(5), 932-948.

Kittur, A., Nickerson, J. V., Bernstein, M., Gerber, E., Shaw, A., Zimmerman, J., Lease, M., & Horton, J. (2013). The future of crowd work. *Proceedings of the 16th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 1301-1318). San Antonio, TX.

Kogut, B., & Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science, 3*(3), 383-397.

Kristensson, P., Magnusson, P. R., & Matthing, J. (2002). Users as a hidden resource for creativity: Findings from an experimental study on user involvement. *Creativity and Innovation Management, 11*(1), 55-61.

Kyriakou, H., Nickerson, J. V., & Sabnis, G. (2017). Knowledge reuse for customization: metamodels in an open design community for 3D printing. *MIS Quarterly, 41*(1), 315-332.

Lessig, L. (2008). *Remix: Making art and commerce thrive in the hybrid economy.* New York: Penguin.

Leonard, D., & Sensiper, S. (1998). The role of tacit knowledge in group innovation. *California Management Review, 40*(3), 112-132.

Leonardi, P. M. (2014). Social media, knowledge sharing, and innovation: Toward a theory of communication visibility. *Information Systems Research, 25*(4), 796-816.

Leong, C. M. L., Pan, S. L., Newell, S., & Cui, L. (2016). The emergence of self-organizing e-commerce ecosystems in remote villages of China: A tale of digital empowerment for rural development. *MIS Quarterly, 40*(2), 475-484.

Lim, W. C. (1994). Effects of reuse on quality, productivity, and economics. *IEEE Software, 11*(5), 23-30.

Majchrzak, A., Cooper, L. P., & Neece, O. E. (2004). Knowledge reuse for innovation. *Management Science, 50*(2), 174-188.

Majchrzak, A., & Malhotra, A. (2013). Towards an information systems perspective and research agenda on crowdsourcing for innovation. *The Journal of Strategic Information Systems, 22*(4), 257-268.

Majchrzak, A., Markus, M. L., & Wareham, J. (2016). Designing for digital transformation: Lessons for information systems research from the study of ICT and societal challenges. *MIS Quarterly, 40*(2), 267-277.

Malone, T. W., Laubacher, R., & Dellarocas, C. (2010). The collective intelligence genome. *MIT Sloan Management Review, 51*(3), 21.

Malone, T. W., Nickerson, J. V., Laubacher, R., Hesse Fisher, L., De Boer, P., Han, Y., & Towne, W. B. (2017). Putting the Pieces Back Together Again: Contest Webs for Large-Scale Problem Solving. *Proceedings of the 20th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 1661-1674). Portland, OR.

Malone, T. W. (2018). *Superminds: the surprising power of people and computers thinking together.* Boston: Little, Brown and Company.

Margolis, J. D., & Walsh, J. P. (2003). Misery loves companies: Rethinking social initiatives by business. *Administrative Science Quarterly, 48*(2), 268-305.

Mark, G., Lyytinen, K., & Bergman, M. (2007). Boundary objects in design: An ecological view of design artifacts. *Journal of the Association for Information Systems, 8*(11), 34.

Markus, L. M. (2001). Toward a theory of knowledge reuse: Types of knowledge reuse situations and factors in reuse success. *Journal of Management Information Systems, 18*(1), 57-93.

Marom, D., Robb, A., & Sade, O. (2016). Gender dynamics in crowdfunding (Kickstarter): Evidence on entrepreneurs, investors, deals and taste-based discrimination. *SSRN Electronic Journal*: http://ssrn.com/abstract= 2442954.

Mein Goh, J., Gao, G., & Agarwal, R. (2016). The creation of social value: Can an online health community reduce rural–urban health disparities? *MIS Quarterly, 40*(1), 247-263.

Merchant, R. M., Asch, D. A., Hershey, J. C., Griffis, H. M., Hill, S., Saynisch, O., Leung, A. C., Asch, J. M, Lozada, K., Nadkarni, L.D., Kilaru, A.S., Branas, C, Stone, E.M., Starr, L., Shofer, F, Nichol, G., & Becker, L.B. (2013). A crowdsourcing innovation challenge to locate and map automated external defibrillators. *Circulation: Cardiovascular Quality and Outcomes, 6*(2), 229-236.

Mergel, I., & Desouza, K. C. (2013). Implementing open innovation in the public sector: The case of Challenge. gov. *Public Administration Review, 73*(6), 882-890.

McGrath, F., & Parkes, C. (2007). Cognitive style fit and effective knowledge reuse. *International Journal of Knowledge Management Studies, 1*(3-4), 484-496.

Muller, C. L., Chapman, L., Johnston, S., Kidd, C., Illingworth, S., Foody, G., Overeem, A., & Leigh, R. R. (2015). Crowdsourcing for climate and atmospheric sciences: current status and future potential. *International Journal of Climatology, 35*(11), 3185-3203.

Navas, E. (2012). Remix theory: The aesthetics of sampling. New York: Springer.

Nickerson J.V. (2015) Collective Design: Remixing and Visibility. In: Gero J., & Hanna S. (Eds), *Design Computing and Cognition '14* (pp. 263-276). Springer, Cham.

Nicolini, D., Mengis, J., & Swan, J. (2012). Understanding the role of objects in cross-disciplinary collaboration. *Organization Science, 23*(3), 612-629.

Nonaka, I. (1994). A Dynamic Theory of Organizational Knowledge Creation. *Organization Science, 5*(1), 14-37.

Paulini, M., Maher, M. L., & Murty, P. (2014). Motivating participation in online innovation communities. *International Journal of Web Based Communities, 10*(1), 94-114.

Phillips, J. (2010). Open innovation typology*. International Journal of Innovation Science, 2*(4), 175-183.

Resnick, M., Maloney, J., Monroy-Hernández, A., Rusk, N., Eastmond, E., Brennan, K., Silver, J., Silverman, B., & Kafai, Y. (2009). Scratch: programming for all. *Communications of the ACM, 52*(11), 60-67.

Rico, R., Sánchez-Manzanares, M., Gil, F., & Gibson, C. (2008). Team implicit coordination processes: A team knowledge–based approach. *Academy of Management Review, 33*(1), 163-184.

Rittel, H. W., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences, 4*(2), 155-169.

Saxton, G. D., Oh, O., & Kishore, R. (2013). Rules of crowdsourcing: Models, issues, and systems of control. *Information Systems Management, 30*(1), 2-20.

Schenk, E., & Guittard, C. (2011). Towards a characterization of crowdsourcing practices. *Journal of Innovation Economics & Management*, (1), 93-107.

Scherer, A. G., & Palazzo, G. (2011). The new political role of business in a globalized world: A review of a new perspective on CSR and its implications for the firm, governance, and democracy. *Journal of Management Studies, 48*(4), 899-931.

Schlagwein, D., & Bjørn-Andersen, N. (2014). Organizational learning with crowdsourcing: The revelatory case of LEGO. *Journal of the Association for Information Systems, 15*(11), 754-778.

Siangliulue, P., Chan, J., Huber, B., Dow, S. P., & Gajos, K. Z. (2016). IdeaHound: Self-sustainable Idea Generation in Creative Online Communities. *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work & Social Computing* (pp. 98-101). San Francisco, CA.

Sojer, M., & Henkel, J. (2010). Code reuse in open source software development: Quantitative evidence, drivers, and impediments. *Journal of the Association for Information Systems, 11*(12), 868-901.

Srivastava, S. C., Teo, T. S., & Devaraj, S. (2016). You can't bribe a computer: Dealing with the societal challenge of corruption through ICT. *MIS Quarterly, 40*(2), 511-526.

Stanko, M. A. (2016). Toward a theory of remixing in online innovation communities. *Information Systems Research*, 27(4), 773-791.

Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science, 19*(3), 387-420.

Szulanski, G. (2000). The process of knowledge transfer: A diachronic analysis of stickiness. *Organizational Behavior and Human Decision Processes, 82*(1), 9-27.

Vashistha, A., Vaish, R., Cutrell, E., & Thies, W. (2015). The whodunit challenge: Mobilizing the crowd in india. In J. Abascal, S. Barbosa, M. Fetter, T. Gross, P. Palanque, & M. Winckler (Eds.), *Proceedings of Human-Computer Interaction* (pp. 505-521). Bamberg, Germany.

Venkatesh, V., Rai, A., Sykes, T. A., & Aljafari, R. (2016). Combating infant mortality in rural India: Evidence from a field study of eHealth kiosk imlementations. *MIS Quarterly, 40*(2), 353-380.

Vilarinho, T., Pappas, I. O., Mora, S., Dinant, I., Floch, J., Oliveira, M., & Jaccheri, L. (2018). Experimenting a digital collaborative platform for supporting social innovation in multiple settings. In M. Hodoň, G. Eichler, C. Erfurth, & G. Fahrnberger (Eds.), *Proceedings of International Conference on Innovations for Community Services* (pp. 142-157). Žilina, Slovakia.

Von Hippel, E., & von Krogh, G. (2003). Open source software and the “private-collective” innovation model: Issues for organization science. *Organization Science, 14*(2), 209-223.

Wang, J., Farooq, U., & Carroll, J. M. (2013). Does design rationale enhance creativity? In J.M. Carrol (Eds.), *Creativity and Rationale* (pp. 197-222). London: Springer.

Ward, T. B. (2001). Creative cognition, conceptual combination, and the creative writing of Stephen R. Donaldson. *American Psychologist, 56*(4), 350.

White House. (2009). Transparency and Open Government. Retrieved from https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government.

Wikipedia: Size comparisons. (2019). In Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Wikipedia:Size_comparisons.

Wisdom, T. N., & Goldstone, R. L. (2011). Innovation, imitation, and problem solving in a networked group. *Nonlinear Dynamics-Psychology and Life Sciences, 15*(2), 229.

Yu, J., Thom, J. A., & Tam, A. (2007). Ontology evaluation using wikipedia categories for browsing. *Proceedings of the Sixteenth ACM Conference on Information and Knowledge Management* (pp. 223-232). Lisbon, Portugal.

Zhang, C., Hahn, J., & De, P. (2013). Research note—Continued participation in online innovation communities: does community response matter equally for everyone? *Information Systems Research,* 24(4), 1112-1130.

Zhang, W., & Watts, S. A. (2008). Capitalizing on content: Information adoption in two online communities. *Journal of the Association for Information Systems, 9*(2), 73.

Zouaq, A., Gasevic, D., & Hatala, M. (2011). Towards open ontology learning and filtering. *Information Systems, 36*(7), 1064-1081.

Appendix

Table A1. Existing Usage of Crowdsourcing to Address Societal Challenges

Methods	Pros	Cons	Example Paper	Societal Challenge and Context
Crowdsourcing in the form of web-enabled open call	Enable deeper levels of public participation; harness collective intelligence and creative solutions with non-expert knowledge; low cost.	Challenges in sustaining of the online community; harder to motivate non-local participants; risk of low-quality individual entries.	Brabham, 2009	Urban planning and sustainability
			Jarmolowicz, Bickel, Carter, Franck, & Mueller, 2012	Health – smoking cessation
Crowdsourcing in the form of web-enabled contests and competitions	Active citizen participation; contribute to the advancement of democracy and the validity of public institution.	Unclear goal definition leads to failure in generating desired solutions; challenge in collaboration; redundancy in ideas.	Mergel & Desouza, 2013	Governmental challenges (e.g. science and technology, health, international relations etc.)
			Vilarinho et al., 2018	Social innovation in multiple settings
Crowdsourcing in the form of mobile-enabled tournament	Low monetary and time cost; high mobility; high diversity.	Labor intensive quality control; high solution validation time; low scalability.	Merchant et. al, 2013	Health – mapping of automated external defibrillators
			Vashistha, Vaish, Cutrell, & Thies, 2015	Social mobilization in developing countries
Crowdsourcing in the form of Crowdfunding	Leverages the Internet and social network to reach out to an undefined large number of potential investors.	Limited to monetary contribution.	Marom, Robb & Sade, 2016	Gender equality and female entrepreneurship
			Gossel, Brüntje, & Will, 2016	Financial Crisis
Crowdsourcing in the form of remixing	Promotes crowd collaboration; allows task division and integration; deepens domain knowledge and generates comprehensive solutions.	Tech difficulty in platform construction; limited support for encouraging reusability	Malone et al. 2017	Global Climate Change

Table A2. Correlation Table of All Variables

	1	**2**	**3**	**4**	**5**	**6**	**7**	8
1. Generativity	1							
2. Integration metaknowledge	0.40*	1						
3. Number of High-quality Proposals Reused	0.19	0.60*	1					
4. Proposal Topic Prevalence	0.66*	0.37*	0.40*	1				
5. Number of Contributors	-0.06	0.03	0.23	-0.03	1			
6. Sequence of Proposal Creation	-0.27	-0.05	-0.15	-0.35*	-0.01	1		
7. Proposal Owner's Tenure	0.53*	0.32*	0.25	0.28	-0.17	-0.16	1	
8. Owner with Multiple Regional Proposals	0.24	-0.01	-0.01	0.21	-0.25	-0.22	0.37*	1

***p<0.001; **p<0.01; *p<0.05

Table A3. Multicollinearity Check

Variable	**VIF**	**1/VIF**
Number of Contributors	1.34	0.749
Sequence of Proposal Creation	1.70	0.587
Proposal Owner's Tenure	1.45	0.691
Owner with Multiple Regional Proposals	1.48	0.674
Proposal Topic Prevalence	2.19	0.456
Number of High-quality Proposals Reused	17.12	0.058
Number of High-quality Proposals Reused (Squared)	15.05	0.066
Integration metaknowledge	2.14	0.468
Contest		
1302007	1.67	0.598
1302013	1.88	0.532
1302019	2.28	0.440
1302025	1.49	0.672
1302031	1.87	0.534
Mean VIF	**3.97**	

Author Bios

Yue Han is an Assistant Professor of Information Systems in the Madden School of Business at Le Moyne College. She holds a Ph.D. degree in Business Administration with the concentration in Information Systems and Analytics, and a master's degree in Project Management from Stevens Institute of Technology. Her research primarily focuses on collective intelligence and crowd innovation, online collaborations, and information diffusion on social media platforms.

Pinar Ozturk is an Assistant Professor of Information Systems Management in the Palumbo-Donahue School of Business at Duquesne University. She holds a Ph.D. in Business Administration from Stevens Institute of Technology, an MBA and an M.Eng. in Electrical Engineering from University of Hartford, CT. Her research interests include collective computer-supported cooperative work, collective intelligence, crowdsourcing and online collaborations. Specifically, she focuses on understanding origins and dynamics of collective group behavior on mass collaboration systems and social media platforms.

Jeffrey V. Nickerson is Professor and Associate Dean of Research in the School of Business at Stevens Institute of Technology. His research is focused on collective creativity: how teams, online communities and crowds produce new knowledge. He is currently investigating how systems composed of humans and machines explore the space of potential designs for 3D objects, encyclopedia articles, skills, occupations, and policies.