Dec 10th, 12:00 AM

The Generativity of Remixing: Understanding Knowledge Reuse Process for Innovation in Online Communities

Yue Han
Le Moyne College, hanyue0610@gmail.com

Jeffrey V. Nickerson
Stevens Institute of Technology, jnickerson@stevens.edu

The Generativity of Remixing: Understanding Knowledge Reuse Process for Innovation in Online Communities

Completed Research Paper

Yue Han
Le Moyne College
1419 Salt Springs Rd, Syracuse, NY 13214
hany@lemoyne.edu

Jeffrey V. Nickerson
Stevens Institute of Technology
1 Castle Point Terrace, Hoboken, NJ 07030
jnickerson@stevens.edu

Abstract

Remixing, a method to harness collective intelligence, is used in many online innovation communities. It is also an important form of online engagement. What actions lead to a remix that is generative? This paper addresses this question by using a knowledge reuse process model previously applied in offline settings as the basis for a series of hypotheses about online communities. An empirical study is performed to examine the relationship between three major actions in the knowledge reuse process model and the generativity of the remix created. An analysis of the reuse of proposals in an online innovation community, Climate CoLab, shows that those including prevalent topics and metaknowledge about integration are more generative. These findings provide insights to strategies and tools that can support knowledge reuse for innovation in online communities.

Keywords: Innovation, knowledge reuse, online communities, metaknowledge, remixing

Introduction

Researchers have created various frameworks to study knowledge reuse (Grant 1996; Markus 2001; Szulanski 2000). While much knowledge is reused just to accomplish short term goals, in design situations, knowledge is reused for the purpose of discovering new knowledge; this is knowledge reuse for innovation (Armbrecht et al. 2001; Majchrzak et al. 2004). Majchrzak et al. (2004) built a six-stage process model to better understand knowledge reuse for innovation; they also suggested that three major actions were involved in this process—reconceptualizing the problem, searching and evaluating ideas to reuse, and developing the selected idea. Studies that followed discussed how the performances of these actions affect the quality of the innovation (Boh 2008; Cheung et al. 2008). Another outcome variable particularly important in online communities is the amount of reuse of the newly created innovation (Kyriakou et al. 2017); this measure of reuse is sometimes referred to as the generativity of the innovation (Hill and Monroy-Hernández 2013). While knowledge is sometimes reused for convenience, as when one acquires a technology product, in this paper we focus on situations in which knowledge is reused in order to build deeper knowledge, a process called knowledge reuse for innovation. We are interested in knowing if an artifact that is the result of an innovation process is itself reused. We are also interested in understanding which processes might drive this reuse.

In recent years, along with the discussion of open innovation and crowdsourcing, researchers have started to explore knowledge reuse in online communities (Hill and Monroy-Hernández 2013; Nickerson 2015; Sojer and Henkel 2010). Knowledge reuse for innovation has been incorporated into the design of these online communities in the form of remixing. These sites permit users to search for and reuse user-generated content in order to create their own innovations, which are in turn shared for others to reuse (e.g., Github, ccMixter, Scratch, Thingiverse, Climate CoLab). They are also 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 these digital traces, it is more straightforward to study the reusability of an innovation in online communities versus offline communities.

In this paper, we aim to address the following research question: How do the three major actions in knowledge reuse process for innovation affect the generativity of an innovation in an online community? More simply, what processes may help a user create a remix that can in turn generate more remixes?

To answer this question, we conducted an empirical study and analyzed data from an online innovation community, Climate CoLab. In the Climate CoLab website, community members are encouraged to participate in different contests by creating novel proposals that address global climate change (Malone et al. 2017). In the empirical study, we analyzed proposals entered in contests that were designed to encourage knowledge reuse for innovation. In these contests, proposal creators search for and integrate pre-existing proposals when creating their novel entries. For each proposal, we identified three features of the proposal to measure the performance of the three major actions in the knowledge reuse process: proposal topic prevalence, number of high-quality proposals reused, and encoded metaknowledge about the rationale for the integration.

This paper aims to help both knowledge workers and online innovation community designers understand what factors within their control increase the reuse of innovations. Our empirical study suggests that incorporating prevalent topics when reconceptualizing the problem increases the generativity of the final creation; encoding metadata about the integration of the ideas reused also increases its generativity. These findings contribute to the knowledge reuse literature by exploring the relationship between the knowledge reuse process and the generativity of the outcome. They also shed light on the design of online innovation communities: Knowledge workers can adopt certain strategies to increase the generativity of their creations; online innovation community designers can also incorporate tools to better support knowledge reuse for innovation in an online community.

This paper starts with a brief review of related work, followed by our hypotheses. Then we describe our empirical study and present the results. Finally, we discuss the results and consider the implications for theory and practice.

Theoretical Development

Knowledge Reuse for Innovation

Knowledge reuse is commonly interpreted as the process of locating and using shared knowledge (Alavi and Leidner 2001). Researchers believe that knowledge reuse is important to study because it contributes to combinative capabilities (Grant 1996; Kogut and 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 1999; Grant 1996; Lenoard and 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 et al. (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 will be referred to as the *knowledge reuse process model* in this paper.

The knowledge reuse process model has been used as the foundation of later studies. A few researchers extended the discussion and suggested enhancements (Chewar and McCrickard 2005; Jing et al. 2015; Petter and Vaishnavi 2007). One paper, following Majchrzak et al.'s (2004) research design and case selection process, constructed a four-stage software development process for knowledge reuse and integration (Chewar and McCrickard 2005). Another recent work developed an idea creation framework based on the model (Jing et al. 2015). 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 and Fadel 2016; Faniel and Majchrzak 2007; Kankanhalli et al. 2011; Khedhaouria and Jamal 2015; Majchrzak et al. 2013). For example, one study was interested in finding ways to optimize the knowledge management systems and technologies to help knowledge reuse for innovation (Faniel and Majchrzak 2007). Another study discussed how to better use knowledge electronic repositories at the search stage of knowledge reuse (Durcikova and Fadel 2016). In addition, a few papers have explored how adapters, metaknowledge, and other factors influence knowledge reuse (McGrath and Parkes 2007; Tung et al. 2014).

Hill and Monroy-Hernández (2013) examined an online remixing community and found that there is a trade-off between generativity and originality, a dimension of quality. 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. To better understand generativity in a remixing community, Kyriakou et al. (2017) studied a 3D printing design community and discussed the relationship between reuse and metamodels—a kind of reuse for innovation. But the study didn't focus on the sequence of steps taken to create an innovation. We do not know the relationship between the knowledge reuse process and the reuse of the resulting innovation. This review suggests that there is a need to revisit the process of knowledge reuse for innovation to understand how the process affects generativity.

Three Major Actions in Knowledge Reuse for Innovation

In their 2004 paper, Majchrzak, Cooper, and Neece (2004) identified a six-stage knowledge reuse process for innovation (Figure 1). The six stages are: reconceptualize the problem and 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.

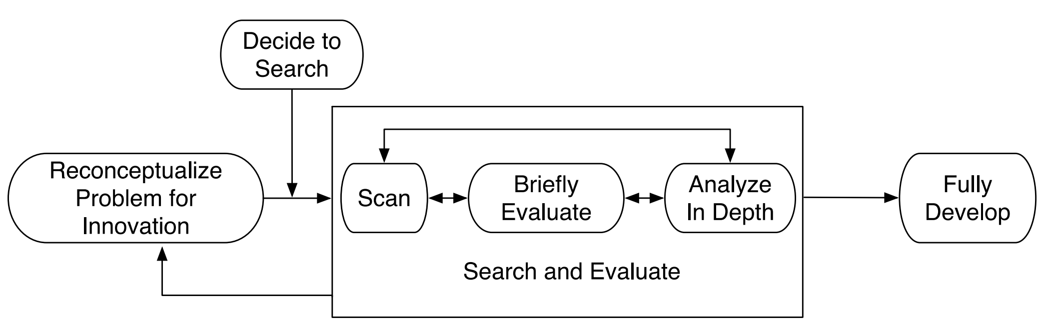

Figure 1. Knowledge Reuse Process for Innovation

(Redrawn from Majchrzak et al., 2004)

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 which 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 and Albert 1999), or because of familiarity (Argote 1999; Brown and Duguid 1991; Nonaka 1994; Hutcheon 2006). 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.*

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 indications 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 and Monroy-Hernández 2013). However, another study examined a music remixing community and suggested that a music remix that has reused more previous music work is less likely to be reused by others because users find it easier to reuse a single work than to recombine multiple sources (Cheliotis and 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. Therefore, 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.*

Major Action 3: Develop the Selected Idea

The third action in the knowledge reuse process for innovation is the 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. Direct evaluation on the integration level of the outcome is both challenging and subjective, as the evaluators need to fully understand both the final idea and all ideas that have been reused. An alternative way to measure the performance of this action is to check if there is metaknowledge expressed about the integration: whether the creators have explicitly explained how they integrate 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 increases the generativity of this creation. 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 metaknowledge about integration is more likely to be reused.*

Research Design

Site Selection

To answer our research question and test the above hypotheses, we conducted an empirical study using data from an online innovation community—Climate CoLab (Introne et al., 2011; Malone, 2010; Nagar, 2013). In Climate CoLab, members collaborate with each other to enter contests by creating proposals addressing global climate change. So far, the website has nearly 75,000 registered members and over 500,000 visitors.

We choose Climate CoLab for the following reasons: First, 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. Second, 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 exploitation condition for knowledge reuse for innovation (Armbrecht et al. 2001). Third, 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). Last but not the least, 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.

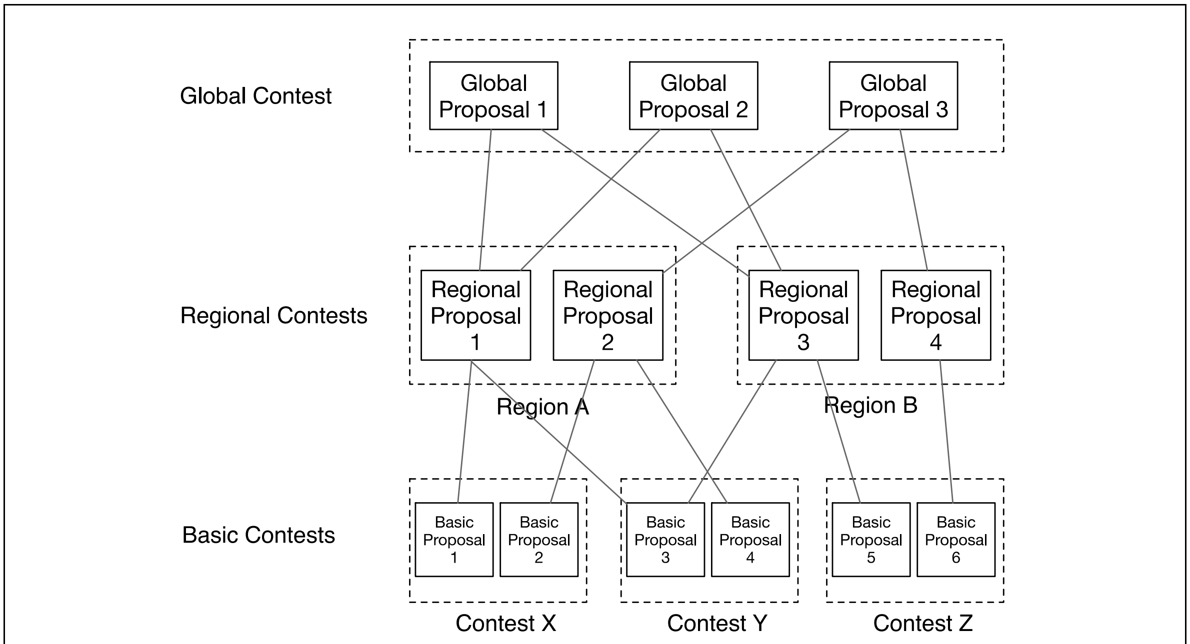

Figure 2. Climate CoLab Proposal Reuse Structure

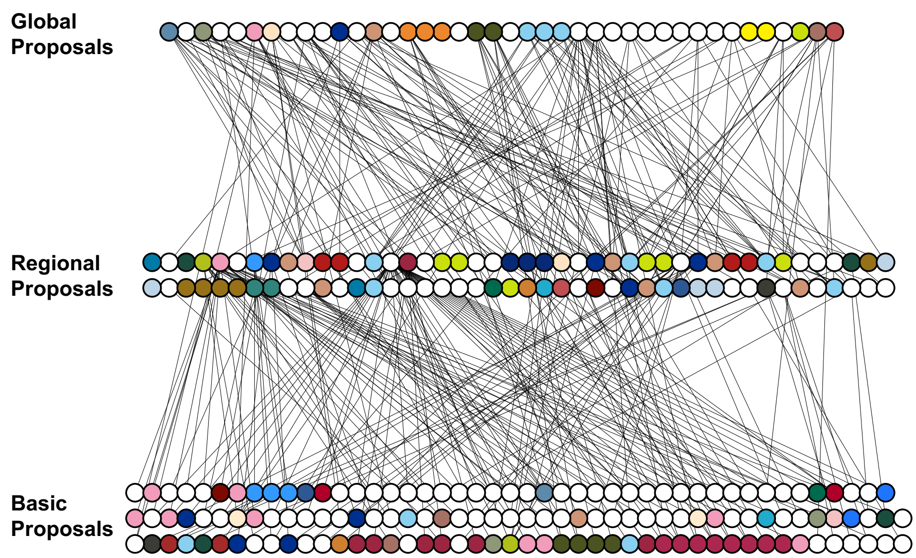

Figure 3. Proposal Reuse Network in the Climate CoLab Website

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.

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.

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 is 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, in this section, 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.

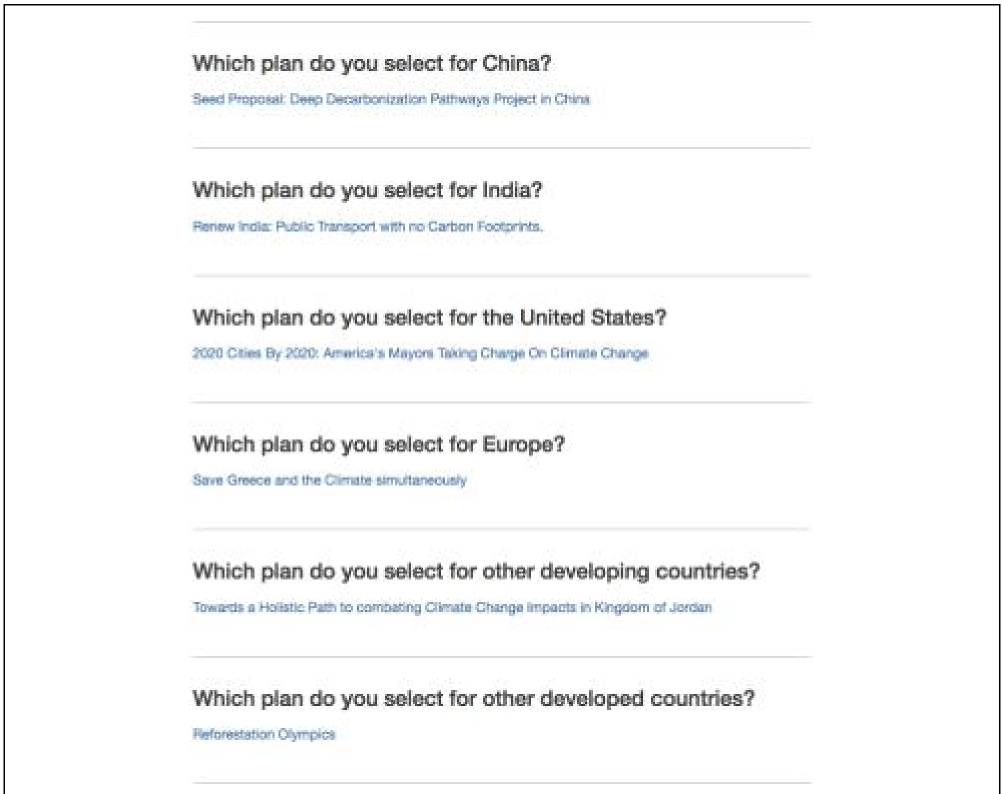

Figure 4. Links to Regional Proposals in a Section of Global Proposals

Independent Variables

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. 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. We followed the procedure described in previous research for this calculation (Ozturk et al. 2016). We first conducted a textual analysis to identify the topics of each proposal. We used the Wikipedia category structure as our ontology 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 et al. (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). We traversed four levels of the category graph, as at that point the high-level topics were sufficiently general. For each proposal, we recorded all identified categories as the topics it covers.

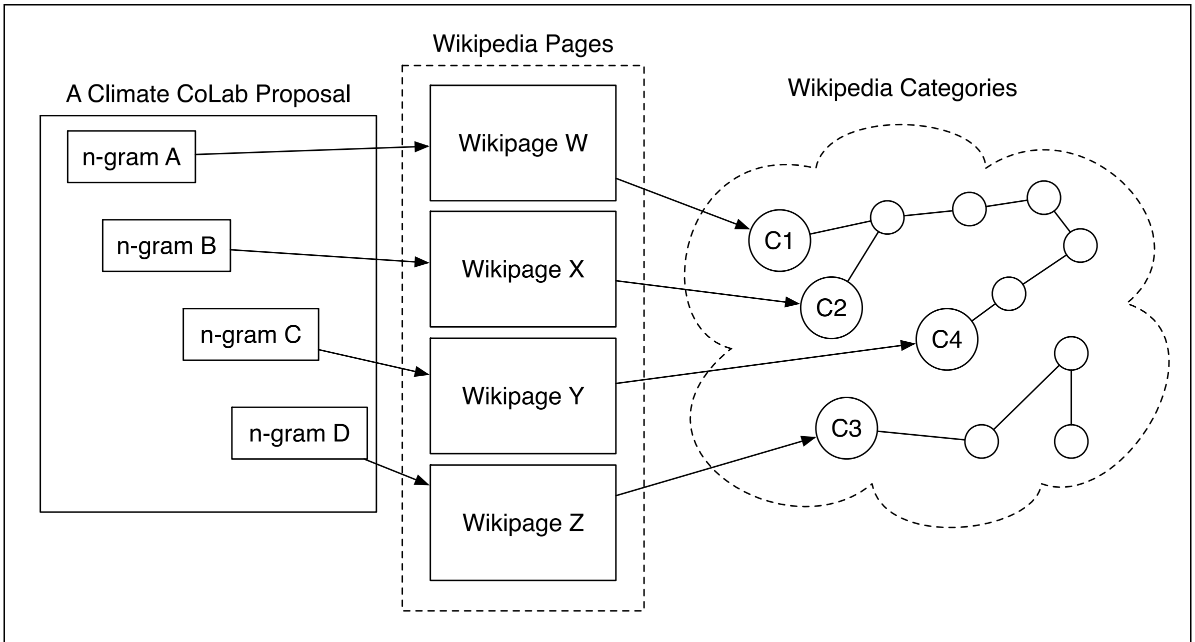

Figure 5. Identifying Topics in a Regional Proposal

Then we calculated the topic prevalence score for each topic within a contest (Figure 6). The topic prevalence score is the degree of topic node divided by the max possible topic degree. For example, in contest X, there are four proposals and topic 2 is covered in one proposal. The topic prevalence score 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.

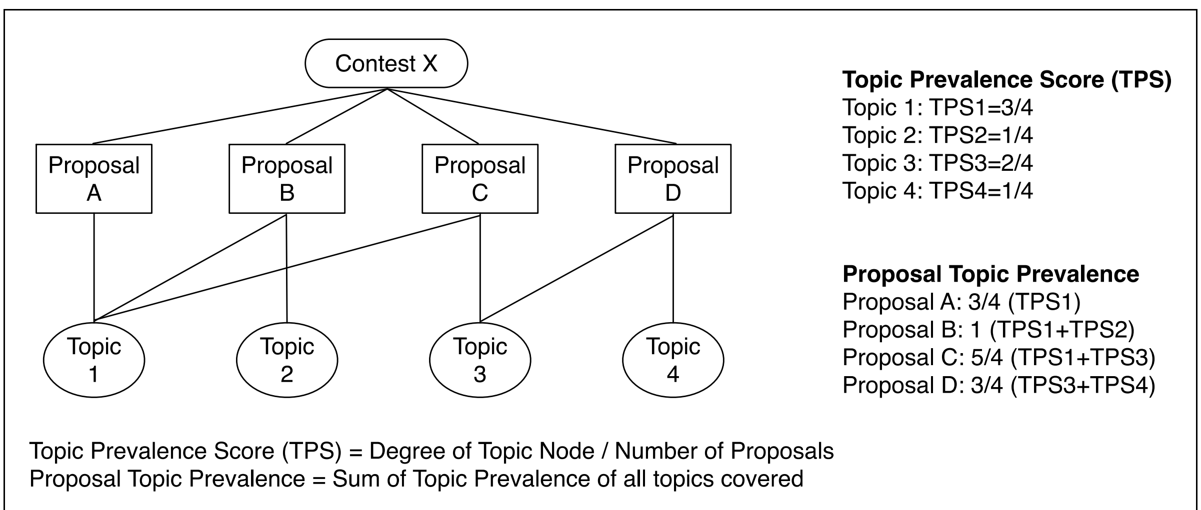

Figure 6. Calculating Proposal Topic Prevalence

Number of High-quality Proposals Reused is our independent variable for H2. There is a special section in each regional proposal where proposal creators can create links 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.

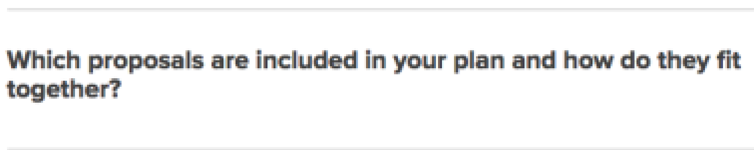

Figure 7. A Section in Regional Proposals that Provides Reuse Links and Metaknowledge about Integration

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.

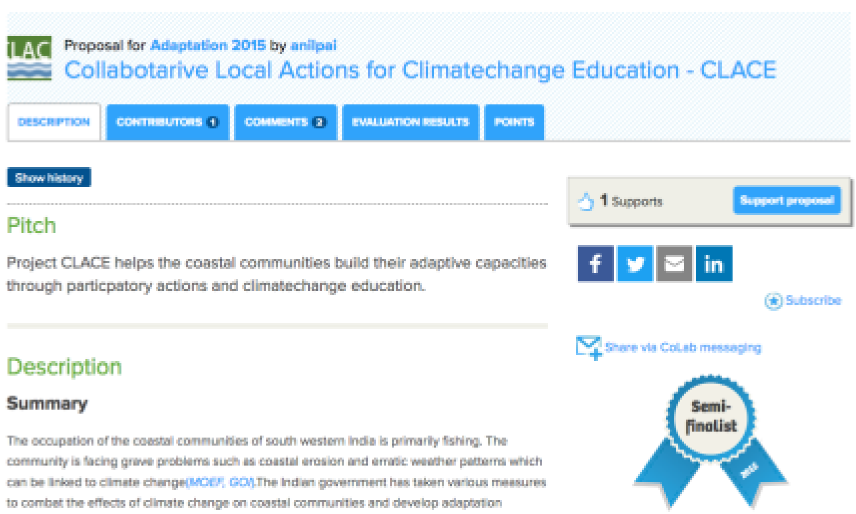

Figure 8. An Example of a High-quality Basic Proposal

Metaknowledge about Integration is our independent variable for H3. We extracted information from the section shown in Figure 7 for each regional proposal and coded all regional proposals. If proposal creators talked about how they integrated the basic proposals reused and provided links to these basic proposals, we considered that metaknowledge about integration was expressed and coded the proposal as True for this binary variable and recorded it as 1 for our regression models (e.g. Figure 9). Otherwise, we coded the proposal as False and recorded it as 0 (e.g., Figure 10).

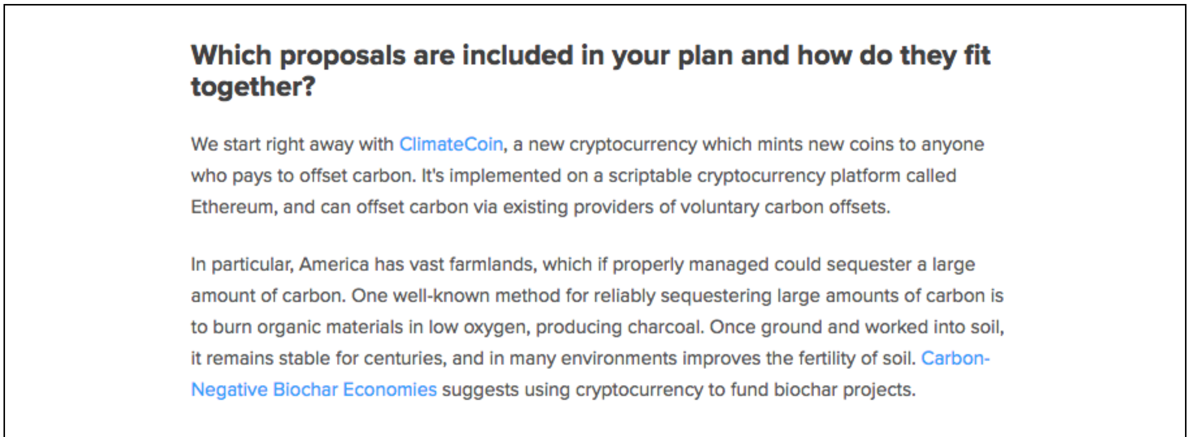

Figure 9. A Proposal with Metaknowledge about Integration

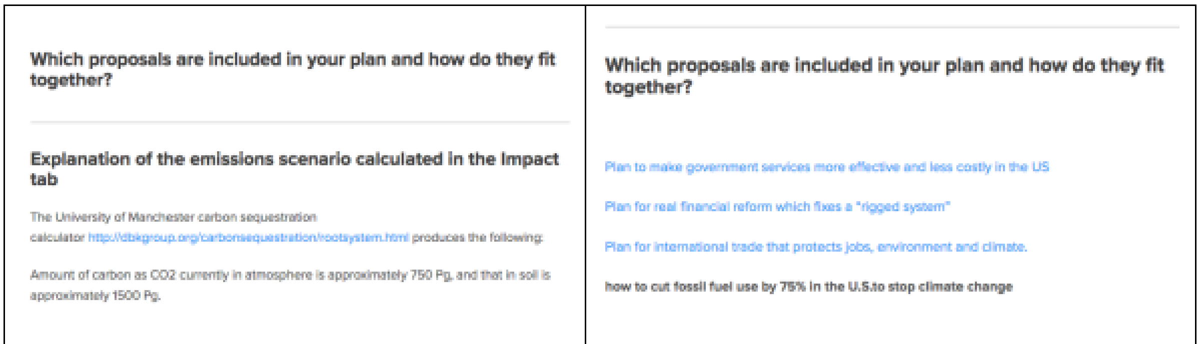

Figure 10. Example Proposals without Metaknowledge about Integration

Control Variables

In this study, the control variables are the *Number of Contributors*, the *Sequence of Proposal Creation*, the *Proposal Owner's Tenure*, and the fixed effect of the *Regional Contest*. 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 and Albert 1999). 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. 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: 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 et al. 2011; Faraj et al. 2015; Goh et al. 2016; Kyriakou et al. 2017). 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.

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 statistics of all variables are listed in Table 1. We have standardized all the independent variables and control variables in all regression models. The correlation table, and multicollinearity check can be found in the Appendix. The results are presented in Table 2.

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
Metaknowledge about Integration	81	0.296	0.459	0	0	1
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

Table 1. Statistics of All Variables

Four Poisson regression models are presented in Table 2, model 1 is a basic regression model with all control variables. The number of contributors 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.

In model 2 we mainly tested the relationship between the proposal topic prevalence with the generativity of a regional proposal. The result suggests that proposal topic prevalence has a positive influence on the generativity of a remix, which means that including more prevalent topics are more likely to be reused in the future. Therefore, hypothesis 1 is supported.

In model 3 we studied 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 metaknowledge about integration. We examined if including metaknowelge about integration would have an influence on the generativity of a regional proposal. The result suggests that including metaknoweldge about integration has a positive influence on the generativity of a remix. In addition, we conducted a Welch's two sample t-test with this variable and found the same result: When a proposal includes metaknowledge about integration, this proposal is more likely to be reused in the future (Table 3). Therefore, hypothesis 3 is also supported. Based on our analysis and results, we summarized our findings in table 4.

		Model 1	Model 2	Model 3	Model 4
	Constant	-0.816*	-0.833*	-0.678	-0.891*
Control	Number of Contributors	0.103	0.104	0.146	0.161
	Sequence of Proposal Creation	-0.094	0.005	0.019	-0.126
	Proposal Owner's Tenure	0.444***	0.459***	0.477***	0.436***
Fixed Effect (Contest)	1303007	1.255**	0.839	0.639	0.677
	1302013	1.333**	0.321	0.005	0.184
	1302019	0.541	0.754	0.595	0.905
	1302025	0.864*	0.880*	0.664	0.574
	1302031	1.177**	0.959*	0.764	0.951*
H1	Proposal Topic Prevalence		0.465***	0.523***	0.512***
H2	Number of High-quality Proposals Reused			0.122	-0.476
	Number of High-quality Proposals Reused (squared)			-0.247	0.172
H3	Metaknowledge about Integration				0.388**
Number of Observations		81	81	81	81
R-Square		0.21	0.27	0.28	0.31
Adjusted R-square		0.13	0.18	0.18	0.21
Sig of model (p)		0.00***	0.00***	0.00***	0.00***

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

Table 2. Poisson Regression Model for Generativity

	True (with metaknowledge about integration)	False (without metaknowledge about integration)
N	24	57
Mean of Generativity	2.08	0.75
Std.Dev	1.8	1.1
t	3.3343	
df	30.449	
P-value	0.002258**	

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

Table 3. Welch's Two Sample T-test of Generativity

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 metaknowledge about integration is more likely to be reused.	Supported

Table 4. Summary of Findings

Discussion and Implications

The Effects of Three Major Actions on Generativity

This empirical study explored the relationship between the performances of three major actions in the knowledge reuse process for innovation and the generativity of the innovative outcome created. As shown in Table 4, 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 work reused in a remix and the generativity of this remix follows a U-shape relationship. However, in our study H2 is not supported by the result. This surprising result 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 including metaknowledge about integration increases the generativity of a remix (H3). Including metaknowledge about integration signals 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, metaknowledge about integration 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 more limited in the time they can spend on a creation, this information creates a quick access to knowledge and improves the efficiency in knowledge reuse.

Our findings suggest that creators can adopt certain strategies to increase the reusability of their creations when they build off previous artifacts. They can incorporate prevalent topics when reconceptualizing the problem; they can also provide metaknowledge about the integration.

Supporting Knowledge Reuse for Innovation in Online Communities

Increasing the generativity of remixes is also beneficial in 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 to help creators perform better in each step of the knowledge reuse process for innovation.

One essential factor that influences creators' performance 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. Because of the number of artifacts in online communities is constantly growing, it can be very difficult to browse all submissions. Therefore, it might be helpful to automatically detect and summarize 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 innovation communities. Online 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 amount of available artifacts but streamlining search. Therefore, we conjecture that developing tools like recommender systems (e.g. Siangliulue et al. 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 metaknowledge might be encouraged 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.

Future Work

Are our findings generalizable? 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 studied 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 creation? We can examine this via network analysis in the future, in particular checking for network autocorrelation. In addition, we can examine the co-occurrence of proposals that are being reused: 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.

Conclusion

Reusing knowledge for innovation is a complex task. This study looked at 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 the remix in online communities. Knowledge workers can adopt certain 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 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. 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.

Acknowledgements

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

Appendix

	1	2	3	4	5	6	7
1. Generativity	1						
2. Metaknowledge about Integration	0.41*	1					
3. Number of High-quality Proposals Reused	0.19	0.58*	1				
4. Proposal Topic Prevalence	0.66*	0.37*	0.40*	1			
5. Number of Contributors	-0.06	0.04	0.23	-0.03	1		
6. Sequence of Proposal Creation	-0.27	-0.01	-0.15	-0.35*	-0.01	1	
7. Proposal Owner's Tenure	0.53*	0.31*	0.25	0.28	-0.17	-0.16	1

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

Table 5. Correlation Table of All Variables

Variable	**VIF**	**1/VIF**
Number of Contributors	1.30	0.766
Sequence of Proposal Creation	1.68	0.594
Proposal Owner's Tenure	1.28	0.779
Proposal Topic Prevalence	2.19	0.457
Number of High-quality Proposals Reused	15.79	0.063
Number of High-quality Proposals Reused (Squared)	13.93	0.072
Metaknowledge about Integration	1.99	0.503
Contest		
1302007	1.67	0.598
1302013	1.86	0.537
1302019	2.27	0.440
1302025	1.48	0.676
1302031	1.83	0.546
Mean VIF	3.94	

Table 6. Multicollinearity Check

References

Alavi, M., and Leidner, D. E. 2001. "Review: Knowledge Management and Knowledge Management Systems: Conceptual Foundations and Research Issues," *MIS quarterly* (25:1), pp. 107-136.

Argote, L. 1999. Organizational Learning: Creating, Retaining and Transferring Knowledge, Boston, MA: Kluwer Academic Publishers.

Armbrecht, F. M., Chapas, R. B., Chappelow, C. C., Farris, G. F., Friga, P. N., Hartz, C. A., Mcllvaine, E., Postle, S. R., and Whitwell, G. E. 2001. "Knowledge Management in Research and Development," *Research-Technology Management* (44:4), pp. 28-48.

Barabási, A. L., and Albert, R. 1999. "Emergence of Scaling in Random Networks," *Science* (286:5439), pp. 509-512.

Bateman, P. J., Gray, P. H., and Butler, B. S. 2011. "Research note—the impact of community commitment on participation in online communities," *Information Systems Research* (22:4), pp. 841-854.
Boh, W. F. 2008. "Reuse of Knowledge Assets from Repositories: A Mixed Methods Study," Information & Management (45:6), pp. 365-375.
Brown, J. S., and Duguid, P. 1991. "Organizational Learning and Communities-of-practice: Toward a Unified View of Working, Learning, and Innovation," *Organization Science* (2:1), pp. 40-57.
Cheliotis, G., and Yew, J. 2009. "An Analysis of the Social Structure of Remix Culture," in *Proceedings of the Fourth International Conference on Communities and Technologies*, University Park, PA, pp. 165-174.
Cheliotis, G., Hu, N., Yew, J., and Huang, J. 2014. "The Antecedents of Remix," in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing*, Baltimore, MD, pp. 1011-1022.
Cheung, P. K., Chau, P. Y., and Au, A. K. 2008. "Does Knowledge Reuse Make a Creative Person More Creative?" *Decision Support Systems* (45:2), pp. 219-227.
Chewar, C. M., and McCrickard, D. S. 2005. "Links for a Human-centered Science of Design: Integrated Design Knowledge Environments for a Software Development Process," in *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, Big Island, HI, pp. 256c-256c.
Dabbish, L., Stuart, C., Tsay, J., and Herbsleb, J. 2012. "Social Coding in GitHub: Transparency and Collaboration in an Open Software Repository," in Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work, Seattle, WA, pp. 1277-1286.
Durcikova, A., and Fadel, K. J. 2016. "Knowledge Sourcing from Repositories: The Role of System Characteristics and Psychological Climate," *Information & Management* (53:1), pp. 64-78.
Faniel, I. M., and Majchrzak, A. 2007. "Innovating by Accessing Knowledge across Departments," *Decision Support Systems* (43:4), pp. 1684-1691.
Faraj, S., Kudaravalli, S., and Wasko, M. 2015. "Leading Collaboration in Online Communities," *MIS Quarterly* (39:2), pp. 393-412.
Genc, Y., Mason, W. A., and Nickerson, J. V. 2013. "Classifying Short Messages using Collaborative Knowledge Bases: Reading Wikipedia to Understand Twitter", in *#MSM2013 Concept Extracting Challenge Proceedings,* Rio de Janeiro, Brazil, pp. 50-53.
Grant, R. M. 1996. "Prospering in Dynamically-competitive Environments: Organizational Capability as Knowledge Integration," *Organization Science* (7:4), pp, 375-387.
Green, S. B. 1991. "How many subjects does it take to do a regression analysis", *Multivariate behavioral research* (26:3), pp. 499-510.
Goh, J. M., Guodong (Gordon) Gao, and Agarwal, R. 2016. "The Creation of Social Value: Can an Online Health Community Reduce Rural-Urban Health Disparities?" *MIS Quarterly* (40:1), pp. 247-263.
Hill, B. M., and Monroy-Hernández, A. 2013. "The Remixing Dilemma: The Trade-off between Generativity and Originality," *American Behavioral Scientist* (57:5), pp. 643-663.
Hutcheon, L. 2012. *A Theory of Adaptation*, Abingdon, OX: Routledge.
Introne, J., Laubacher, R., Olson, G., and Malone, T. 2011. "The Climate CoLab: Large Scale Model-based Collaborative Planning," in *2011 International Conference on Collaboration Technologies and Systems, Philadelphia*, PA, pp. 40-47.
Jing, D., Yang, H., Shi, M., and Zhu, W. 2015. "Developing a Research Ideas Creation System through Reusing Knowledge Bases for Ontology Construction," in *2015 IEEE 39th Annual Computer Software and Applications Conference*, Taichung, Taiwan, pp. 175-180.
Kankanhalli, A., Lee, O. K. D., and Lim, K. H. 2011. "Knowledge Reuse through Electronic Repositories: A Study in the Context of Customer Service Support," *Information & Management*, (48:2), pp. 106-113.
Khedhaouria, A., and Jamal, A. 2015. "Sourcing Knowledge for Innovation: Knowledge Reuse and Creation in Project Teams," *Journal of Knowledge Management* (19:5), pp. 932-948.
Kogut, B., and Zander, U. 1992. "Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology," *Organization Science* (3:3), pp. 383-397.
Kyriakou, H., Nickerson, J. V., and Sabnis, G. 2017. "Knowledge Reuse for Customization: Metamodels in an Open Design Community for 3D Printing," *MIS Quarterly* (41:1), pp. 315-322.
Leonard, D., and Sensiper, S. 1998. "The Role of Tacit Knowledge in Group Innovation," *California Management Review* (40:3), pp. 112-132.
Lim, W. C. 1994. "Effects of reuse on quality, productivity, and economics," *IEEE Software* (11:5), pp. 23-30.

Majchrzak, A., Cooper, L. P., and Neece, O. E. 2004. “Knowledge Reuse for Innovation,” *Management Science* (50:2), pp. 174-188.

Majchrzak, A., and Malhotra, A. 2013. “Towards an Information Systems Perspective and Research Agenda on Crowdsourcing for Innovation,” *The Journal of Strategic Information Systems* (22:4), pp. 257-268.

Malone, T. W., Laubacher, R., and Dellarocas, C. 2010. “The Collective Intelligence Genome,” *MIT Sloan Management Review* (51:3), pp. 21-31.

Malone, T. W., Nickerson, J. V., Laubacher, R. J., Fisher, L. H., De Boer, P., Han, Y., and Towne, B. 2017. “Putting the Pieces Back Together Again: Contest Webs for Large-Scale Problem Solving,” in *20th ACM Conference on Computer-Supported Cooperative Work and Social Computing*, Portland, OR.

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) pp. 57-93.

McGrath, F., and Parkes, C. 2007. “Cognitive Style Fit and Effective Knowledge Reuse,” *International Journal of Knowledge Management Studies* (1:3-4), pp. 484-496.

Nagar, Y. 2013. “Designing a Collective-intelligence System for Evaluating Complex, Crowd-generated intellectual artifacts,” in Proceedings of the 2013 Conference on Computer Supported Cooperative Work Companion, San Antonio, TX, pp. 73-76.

Nickerson, J. V. 2015. “Collective Design: Remixing and Visibility,” in *Design Computing and Cognition'14*, Gero J., Hanna S. (eds.), Switzerland: Springer International Publishing, pp. 263-276.

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

Ozturk, P., Han, Y., Towne, W. B., and Nickerson, J. V. 2016. “Topic Prevalence and Reuse in an Open Innovation Community,” in Collective Intelligence Conference, New York, NY.

Petter, S., and Vaishnavi, V. 2007. “A Generic Process for Individual Knowledge Reuse: Potential Problems and Solutions,” in *AMCIS 2007 Proceedings*, Keystone, CO, pp. 78.

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

Siangliulue, P., Chan, J., Huber, B., Dow, S. P., and Gajos, K. Z. 2016. “IdeaHound: Self-sustainable Idea Generation in Creative Online Communities,” in *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion,* New York, NY, pp. 98-101.

Sojer, M., and Henkel, J. 2010. “Code Reuse in Open Source Software Development: Quantitative Evidence, Drivers, and Impediments,” *Journal of the Association for Information Systems* (11: 12), pp. 868-901.

Szulanski, G. 2000. “The Process of Knowledge Transfer: A Diachronic Analysis of Stickiness,” *Organizational Behavior and Human Decision Processes* (82:1), pp. 9-27.

Tung, Y. H., Chuang, C. J., and Shan, H. L. 2014. “A Framework of Code Reuse in Open Source Software,” in *16th Asia-Pacific Network Operations and Management Symposium*, Hsinchu, Taiwan, pp. 1-6.