Who Contributes Knowledge? Core-Periphery Tension in Online Innovation Communities

Hani Safadi
Terry College of Business
University of Georgia
Athens, GA, USA
hanisaf@uga.edu

Steven L. Johnson
McIntire School of Commerce
University of Virginia
Charlottesville, VA, USA
sljohnson@virginia.edu

Samer Faraj
Desautels Faculty of Management
McGill University
Montreal, QC Canada
samer.faraj@mcgill.ca

Abstract: Where do valuable contributions originate from in online innovation communities? Prior research provides conflicting answers. One view, consistent with a community of practice perspective, is that valued knowledge contributions are primarily provided by central participants at the core of a community. In contrast, other research—including work adopting an open innovation perspective—predicts that valuable ideas primarily emerge from peripheral participants, those at the margins of a field of knowledge who provide novel ideas and viewpoints. We integrate these contrasting perspectives by considering two distinct forms of position: social embeddedness (a core social position within the social network of participants interacting within a community), and epistemic marginality (a peripheral epistemic position based on the network of topics discussed by a community). Analyzing contributions by 697,412 participants of 52 Stack Exchange online innovation communities, we find that both participants who are socially embedded and participants who are epistemically marginal provide knowledge contributions that are highly valued by fellow community participants. Importantly, among epistemically marginal participants, those with high social embeddedness are more likely to provide contributions valued by the community; by virtue of their epistemic marginality these participants may offer novel ideas while by virtue of their social embeddedness they may be able to more effectively communicate their ideas to the community. Thus, the production of knowledge in an online innovation community involves a complex interaction between the novelty emanating from the epistemic periphery and the social embeddedness required to make ideas congruent with existing social and epistemic norms.

Keywords: Online Communities, Social Networks, Organizing for Innovation in the Digitized World, Quantitative Text Analysis, Archival research: extant data, Digital Innovation, Knowledge Sharing

Acknowledgments: The authors are grateful to Gino Cattani and the two anonymous reviewers for their guiding feedback. The authors also recognize Ahmed Abbasi, Michael Barrett, Nicholas Berente, Steve Borgatti, Marie-Claude Boudreau, Patrick Cohendet, Kevin Crowston, Peter Gray, Ola Henfridsson, James Howison, Elena Karahanna, Brent Kitchens, Paul Leonardi, Ann Majchrzak, Carolina Salge, Suprateek Sarker, Amrit Tiwana, and Richard Watson for helpful comments on earlier drafts of this paper. The authors benefited from feedback by participants at the Academy of Management 2018 OCIS division, the 2017 Terry College of Business MIS Seminar, and the 2018 Virginia Research Seminar Series. Hani Safadi acknowledges support from the 2018 Terry-Sanford Summer Research Award. Steven L. Johnson acknowledges support from the McIntire School of Commerce Foundation. Samer Faraj acknowledges support from the Canada Research Chairs program.

Who Contributes Knowledge? Core-Periphery Tension in Online Innovation Communities

1. Introduction

The tension between the core and the periphery as the primary source of valuable ideas has been observed across various fields of knowledge. In many contexts, knowledge generation occurs when existing ideas and concepts are further developed by a core community of knowledge seekers dwelling in a shared epistemic community (e.g., Boland and Tenkasi 1995, Lave and Wenger 1991, Polanyi 1958). As a result, central participants in a knowledge field are likely to possess the expertise and status to have the contextualized understanding and social integration necessary for engaging with new ideas. In other settings, however, participants who are peripheral to a knowledge field provide creative approaches to solve problems. Peripheral participants have often been found to be the source of novel insight, contributing ideas from different institutional fields (Cattani et al. 2017), fresh perspectives from users uninvolved in product design (von Hippel 2017), and relevant knowledge from adjacent areas of expertise (Jeppesen and Lakhani 2010).

In this paper, we explore how core-periphery tension manifests in online innovation communities. An online innovation community brings together participants with shared interests in a generative collaboration through the open sharing of ideas, experiences, perspectives, problems, and solutions in order to sustain learning and develop collective knowing. Such communities form around wikis, question and answer (Q&A) sites, open-source software development, collaborative digital design platforms, and other forms of open innovation (e.g., Faraj et al. 2016). By supporting knowledge exchange among thousands of dispersed participants, online innovation communities have unique characteristics that may impact knowledge contributions. These communities typically form around a shared field of inquiry and offer participants a chance to learn both by observing knowledge flows and by participating in community interactions (Faraj et al. 2011, Preece 2000). Participation is usually open to anyone interested in joining together in the pursuit of common interests. Participants serve both as knowledge seekers and knowledge providers, and they also comprise an interactive, collaborative audience that clarifies, elaborates upon, and identifies the contributions that are the most valuable. Finally, although the majority of participants rarely meet or interact with each one beyond the online community, participation is recognizably social and leads to strong connections among group members (Kraut and Resnick 2011, Rheingold 1993).

Where do knowledge contributions originate from in online innovation communities? Online community research—consistent with a community of practice perspective (Lave and Wenger 1991)—emphasizes the importance of a central core of active, expert participants for community sustenance and knowledge production (Dahlander and O'Mahony 2011, Faraj et al. 2015, Levina and Arriaga 2014, O'Mahony and Ferraro 2007) and suggests that valuable contributions primarily arise from this core. Indeed, a small percentage of community participants contribute the majority of content and, accordingly, play a dominant role in ongoing community maintenance and the production of knowledge (Dahlander and Frederiksen 2012, Johnson et al. 2015). In contrast, open innovation researchers have emphasized the role of novel idea sources in the generation of valuable knowledge (Bogers et al. 2017, von Hippel 2005, Jeppesen and Lakhani 2010). In this view, creativity and innovation originate from marginal participants, from beyond the primary field of inquiry.

To integrate these conflicting perspectives, we investigate the relationship between a participant's position in a knowledge field and how the audience of fellow participants values their contributions. To inform our understanding of where valued contributions originate, we analyzed data from online innovation communities hosted by Stack Exchange, an online knowledge platform. Through our analysis, we demonstrate the importance of considering two distinct forms of position: a social position in the network of social ties (who interacts with whom) and an epistemic position in the content-derived network of topics (who discusses what). Using a multilevel mixed-effects regression analysis of 52 Stack Exchange online innovation communities, encompassing 1,906,434 quarterly contributions by 697,412 participants over 7 years, we find that both core participants who are socially embedded and peripheral participants who are epistemically marginal provide knowledge contributions that are highly valued by fellow community participants.[1] Furthermore, social embeddedness complements epistemic marginality: epistemically marginal participants who are high on social embeddedness are those most likely to provide highly valued contributions. By showing how social and epistemic positions interact in the production of knowledge, we provide a more nuanced understanding of core-periphery tension in knowledge fields.

2. Where do valuable knowledge contributions originate?

2.1 Core-periphery tension in knowledge fields

Researchers have long extolled the advantages of peripheral participants as sources of novelty and alternative perspectives (e.g., Merton 1973). As Gieryn and Hirsh (1983) note, "the 'marginal' scientist who unexpectedly contributes major theoretical or technical innovations is part of the lore of science" (p. 87). Yet, despite the widely recognized potential benefits of marginality as a source of valuable knowledge, these advantages appear to be infrequently realized. One explanation for the lack of contribution from outsiders is that the creation and development of knowledge within communities is a highly social process (Carlile 2002, Nonaka and von Krogh 2009, Tsoukas 2009). The value of new knowledge contributions can only be understood within the context of a field since knowledge itself is a situated and social accomplishment (Bourdieu 1993). Due to the sticky, localized nature of knowledge, developing and sharing it requires common ground and contextualized shared mental models that arise among participants through field-specific interactions (Argote et al. 2003, von Hippel 1994, Szulanski 1996). As a result, knowledge creation requires deep social interactions, a shared understanding of accepted practices for evaluating new contributions, and an embrace of the generative dance both between tacit and explicit knowledge as well as between knowledge and knowing (Boland and Tenkasi 1995, Orlikowski 2002, Polanyi 1958).

[1] In describing network position in a field, we adopt the terms *embeddedness* and *marginality* to denote opposite ends of a core-periphery continuum: embeddedness is equivalent to a more central, core position; marginality to a less central, more peripheral position.

Still, there is no consensus regarding where valued contributions are likely to emerge within a knowledge field. Phillips and Zuckerman (2001) identified both the core and the periphery as sources of novel ideas but argue that peripheral participants face obstacles in gaining recognition, legitimacy, and acceptance for their new ideas because they lack the social standing and resource advantages of insiders. In communities of knowing, core participants often possess high levels of expertise, social capital, and resources (Collins 1998, Nahapiet and Ghoshal 1998), which enable them to effectively express and develop ideas that otherwise may not resonate with others in the field. Meanwhile, a competing perspective emphasizes that peripheral participants provide unique experiences, expertise, and problem-solving approaches. By leveraging diverse and unique experiences, outsiders with expertise in distant fields can contribute novel ideas that would not be apparent to insiders and help resolve previously unsolvable problems (Jeppesen and Lakhani 2010, Majchrzak et al. 2018). By utilizing alternative ways of thinking about problems and strategies for identifying solutions, marginal participants can serve as catalysts in the problem-solving process (Afuah and Tucci 2012, Gieryn and Hirsh 1983).

Finally, a third perspective suggests that participants in an intermediate position between the core and the periphery may enjoy unique benefits that enhance their ability to make valued contributions. These participants can garner social legitimacy and support by virtue of their closeness to the core while benefiting from the new ideas and inputs that emerge at the periphery of the network. For example, in their study of the motion picture industry from 1992 to 2003, Cattani and Ferriani (2008) found that individuals with an intermediate position garnered the most awards. Likewise, a study of an online innovation community supporting digital audio producers found that the most innovative participants resided at a mid-point between the core and the periphery (Dahlander and Frederiksen 2012).

Where valuable ideas arise within a knowledge field may be affected by facilitating conditions such as community needs, external jolts, and audience expectations. For example, in a study investigating the success of the Cubism movement in gaining acceptance within the art field, Sgourev (2013) found that core participants, concerned by an increasingly fragmented art market, became more receptive to what they originally viewed as an unwelcome approach to painting. Likewise, when analyzing the legitimation of outsider John Harrison's innovative approach to measuring longitude at sea, Cattani et al. (2017) found that external jolts (i.e., response to a major maritime disaster) can raise the urgency of finding a solution and thus propel an outsider's innovative solution toward acceptance. Additionally, the characteristics of the idea itself may be packaged in ways that facilitate acceptance. For example, an analysis of Thomas Edison's system of electric illumination showed that its integration of design elements familiar to gaslight users—such as lampshades, burners, gas statutes, and metered billing—eased its integration into existing usage patterns and user lifestyles (Hargadon and Douglas 2001). Finally, different audiences may evaluate the same contribution differently. In the film industry, for instance, peer audiences were unlikely to value contributions from the periphery, while critic audiences not only "fail to privilege core incumbents, but they actually favor peripheral producers when awarding accolades" (Cattani et al. 2014, p. 276).

2.2 Valued knowledge in online innovation communities

Many online settings—including online social media like Facebook, Twitter, and Reddit—are reminiscent of "great good places" such as pubs, coffee shops, bookstores, and other community centers that promote public association, debate, and social engagement (Oldenburg 1989). Online innovation communities have similar sociality, but as places where knowledge is shared, they also align with the ethos of scientific fields. Built on the assumption that knowledge is cumulative, they are guided by a logic of discovery. They bring together participants who share an interest in the concepts and practices of a complex area of inquiry. Much like Lakatos's (1978) concept of research programs, online innovation communities are concerned with deepening knowledge within a specific domain of interest. In contrast to fields of science, however, online innovation communities do not benefit from established structures that allow science to proceed in a more or less orderly fashion. They lack the institutional structures and processes through which science operates, such as journals, editors, a formal review process, and conferences. Thus, online innovation communities may appear more closely aligned with Feyerabend's (1975) idea of epistemological anarchy than an academic discipline.

The open nature of online innovation communities—with fluid membership, public visibility of all interactions, and no formal positions of authority—favors a highly social process. As observed by Knorr-Cetina's (1999) study of scientific fields, social relations play a major role in what types of contributions are valued and what passes for knowledge. Science-like norms structure the interactions that take place, the equality of participation, and the regulating process (Longino 2002). Similarly, an online innovation community coalesces around an arena of knowledge, the problems the community deems worth addressing, methods for justifying solutions, and strategies for solving problems encountered in practice. Further, given the widespread use of pseudonyms by participants, which typically obscure institutional affiliation, social cues and authority status play a diminished role in the process of contributing knowledge. At the same time, this process is necessarily social, prioritizes the demonstration of expertise, and involves a convergent process of moving toward—if not a singular truth—a collective vetting of useful knowledge.

Because online innovation communities have few barriers to entry and attract a flow of new participants who bring in novel insights and perspectives, they must balance open participation with the need to establish and maintain sustainable behavior norms (Faraj et al. 2015, Kraut and Resnick 2011, Ma and Agarwal 2007). Participants interact by sharing ideas and experiences, debating practices, and helping each other solve problems. This collaborative spirit contrasts with innovation contests and other crowd-sourcing platforms where knowledge seekers pose bespoke problems for their own benefit and offer monetary rewards to the individuals (or teams) that independently generate the best solution (Lakhani 2016). It also differs from other online settings where the focus is on matching knowledge seekers with knowledge providers (e.g., Haas et al. 2015, Jeppesen and Lakhani 2010). Instead, in online innovation communities, the knowledge process is emergent and dialogic in nature; participants clarify and elaborate for each other, building on partial answers, and work together to validate new knowledge (Kudaravalli and Faraj 2008, Tsoukas 2009). Thus, the knowledge process is relationally complex and contingent on social and epistemic norms.

2.3 Social embeddedness and epistemic marginality

Like other social fields, participants in online innovation communities are embedded in a network of social interactions: they occupy a range of positions from core insiders to marginal outsiders. Building on a network perspective to examine participants' interactions (Borgatti and Everett 1999), we use the terms *embeddedness* and *marginality* to denote opposite ends of a core-periphery continuum: embeddedness is equivalent to a central position; marginality to a peripheral position. Historically, the content of flows in a network, although viewed as important, was assumed to be immeasurable in practice (Borgatti and Lopez-Kidwell 2011). Thus, previous conceptualizations of position in a network have singularly focused on measuring ties without considering the knowledge flowing through those ties. Today, the visibility of online interactions affords the observation of human behavior with greater precision and, in turn, the development of a multidimensional conceptualization and understanding of social phenomena (Latour 2010, Lazer et al. 2009). At the same time, the online availability of a content corpus affords differentiation between a social network (who knows or interacts with whom) and a topic network (who is conversing about what). This distinction enables separate analyses of the two network to provide a rich characterization of the community and the social and epistemic positions of its participants.

Participants in online innovation communities possess diverse social and positional capital that can influence how systems of meaning are negotiated within the community. Socially embedded participants are best positioned to understand what constitutes a novel contribution compared to existing content, to recognize topics of interest to the community, and to communicate most effectively in the manner consistent with community norms (Johnson et al. 2015, Levina and Arriaga 2014, Qureshi and Fang 2011, Rullani and Haefliger 2013). Fulfilling the espoused goal of building knowledge—in the sense of developing community-validated useful knowledge—involves a process of interactive dialogue, the setting of contextual conditions, and the deepening of shades of difference (Hadjimichael and Tsoukas 2019, Nonaka and von Krogh 2009). This deepening of perspective making involves the construction and alignment of specific narratives and the raising of communal boundaries regarding topics of relevance

(Boland and Tenkasi 1995, Brown and Duguid 2001). Thus, through repeated interactions, socially embedded participants develop differentiated expertise and mastery of community language and expectations, gains that allow them to offer valuable contributions.

The importance of an authoritative core of community experts is consistent with findings that a relatively small number of online participants contribute the majority of content (Dahlander and O'Mahony 2011, Johnson et al. 2014, Kudaravalli and Faraj 2008, Rullani and Haefliger 2013). Prior research on virtual collaboration environments found that more central participants were viewed as more influential by other participants (Johnson et al. 2014, Sutanto et al. 2011). Studying the Linux Debian community, O'Mahony and Ferraro (2007) found that the community operated as a meritocracy and that sustained interaction with other participants increased the propensity of a community participant becoming a leader. Further, in a survey of participants in an online immigration lawyer community, Wasko and Faraj (2005) found a positive relationship between a participant's self-reported expertise and their centrality in online community discussions. Given that online innovation communities are open, weakly-structured social fields, that social presence is enacted primarily through written communication, and that participants focus on deepening highly contextualized knowledge, we argue that socially embedded participants are more likely to provide knowledge contributions that are valued by the community audience. Therefore, we propose:

> *Proposition 1: In online innovation communities social embeddedness is positively associated with providing valued knowledge contributions.*

Online innovation community participants are united by what Knorr-Cetina (1999) called an epistemic culture: a set of norms, arrangements, and mechanisms that make up what and how they know. The availability of digital traces offers a promising avenue to more comprehensively evaluate interactions, expressions, and social action in the making of knowledge compared to pre-digital approaches to studying knowledge production, which lacked access to traces of such activities (Hampton 2017, Lazer and Radford 2017). Considering the content of exchanges opens up the possibility of identifying who engages on which topics. Each participant's contributions reflect a unique epistemic position based on their perspective, experience, and expertise. Thus, in online innovation communities, participants not only occupy a social position based on whom they interact with, but also a distinct epistemic position based on the content of their contributions. As a result and for example, we can conceptualize a participant as epistemically marginal to the extent that the topics they discuss are distant from the dominant discourse of the community.

This conceptualization aligns with a recent examination of the role of intellectual dissimilarity or distance in online knowledge production. In a study of 2,130 evaluator–proposal pairs for academic research seed grants, Boudreau et al. (2016) identify intellectual distance based on differences between evaluator expertise and proposal topic areas, as measured by the overlap of keywords reflecting evaluator expertise and proposal content. Likewise, Jeppesen and Lakhani (2010) identify technical marginality based on the dissimilarity of a participant's expertise to a focal problem, which they measure as a self-assessed estimate of the distance between a solver's field of expertise and the problem they addressed. Similarly, in a study of evaluations of 105,127 crowdsourced suggestions provided to 922 organizations, Piezunka and Dahlander (2015) identify content distance based on the dissimilarity of word-usage frequency in new suggestions compared to previously considered suggestions. Thus, a participant's epistemic position in a field is reflected in the content of their contributions, which is separate and distinct from their social position based on interpersonal interactions.

Several characteristics of online innovation communities may reduce barriers to the acceptance of novel, innovative solutions generated by peripheral participants. Online communities frequently embrace egalitarian expectations that contributions are judged based on the value of their content rather than on the identity of the author (O'Mahony and Ferraro 2007). Indeed, the pragmatic nature of valuing knowledge—is a contribution useful?—places greater emphasis on the content of contributions than on who provides it. In addition, open access reduces barriers to participation: few restrictions limit who can join a community, view existing content, or contribute new content. Participants with relevant expertise can immediately begin making contributions. Further, online innovation communities offer opportunities for both newcomers and experienced participants to adopt situationally specific roles based on emergent needs (Faraj et al. 2011). Finally, by allowing participants to hide their real-world identities (e.g., name or organizational affiliation), online communities can encourage uninhibited participation (Kraut and Resnick 2011, Sproull and Kiesler 1991). Therefore, epistemically marginal participants may face reduced barriers to active engagement in online innovation communities.

Empirical studies of open innovation contests point to the importance of outsiders in generating creative solutions. For example, in a study of over 12,000 scientists participating in 166 science problem-solving challenges, Jeppesen and Lakhani (2010) found that technical marginality, as measured by the conceptual distance between a participant's technical expertise and the source field of the problem, was associated with providing successful solutions. Likewise, in a study of 14 open innovation challenges sponsored by NASA, the solutions provided by outsiders considerably outperformed those of internal R&D engineers (Lifshitz-Assaf 2018). Similarly, in a study analyzing 8,801 ideas generated by 4,285 participants in the first two years of Dell's Ideastorm community, Bayus (2012) found that peripheral participants provided more diverse, higher quality solutions than core participants. Thus, we suggest that epistemically marginal participants may be particularly valuable because they offer more creative solutions or fresh insights to complex problems that cannot be solved by marshaling the community's established knowledge. Therefore, we propose:

Proposition 2: In online innovation communities epistemic marginality is positively associated with providing valued knowledge contributions.

The arguments above offer an explanation of why valued knowledge contributions may arise from either socially embedded (Proposition 1) or epistemically marginal (Proposition 2) participants. The essence of core-periphery tension is the observation that participants who are best equipped to make a novel contribution are quite frequently also in a disadvantaged social position in a knowledge field. Extant research has largely considered the core and the periphery as endpoints on a single scale where a participant can only be in a single position—the core or the periphery. Specifically, past research has often viewed the social and epistemic position as concomitant and mutually constitutive, such that a core or peripheral position in one dimension strongly suggests an equivalent position in the other. Leveraging the visibility of online interactions facilitates disentangling social and epistemic positions and, accordingly, opens up the ability to consider participants who are *both* epistemically marginal and socially embedded.

In online innovation communities, a participant benefitting from high epistemic marginality and high social embeddedness could leverage the advantages of both. By virtue of their epistemic marginality, such participants could offer novel ideas, while by virtue of their social embeddedness they may be able to more effectively communicate those ideas to the community. As has long been noted, because participants in the social core feel secure about their social position, they may be more willing to actively seek out and promote novel ideas (Phillips and Zuckerman 2001). For participants, getting a novel solution accepted required dexterous navigation of the social field. More specifically, participants' persistent social engagement enables their ideas to overcome initial skepticism and opposition (Cattani et al. 2017). As an example, epistemically marginal health professionals were able to achieve the adoption of their expert recommendations by building and emphasizing their social connection to participants embedded in the social core (DiBenigno 2019). Finally, in online settings, the most influential participants were those able to customize their language to reflect the accepted norms and values of the community (Johnson et al. 2015). Thus, epistemically marginal participants may benefit from high social embeddedness—in terms of a tacit understanding of norms, dominant frames, and community language—in order to more effectively establish their novel contributions. Therefore, we propose:

> *Proposition 3: In online innovation communities social embeddedness strengthens the positive association between epistemic marginality and valued knowledge contributions.*

Figure 1 provides a visual summary of the three propositions and the relationships among social embeddedness, epistemic marginality, and valued knowledge contributions in online innovation communities explored in this paper.

3. Research setting

To test our propositions we perform an empirical analysis of contributions to Stack Exchange, a platform supporting over one hundred communities "that are created and run by experts and enthusiasts ... who are passionate about a specific topic" and provide "answers to practical, detailed questions" (Stack Exchange Tour 2019). Stack Exchange communities provide participants with a supportive community in which they can gain valuable information, be exposed to creative ideas and ways of thinking about problems, and have access to motivated experts jointly engaged in problem-solving. As a place where participants engage with challenging problems, Stack Exchange facilitates participants' learning and supports innovation within organizations. Stack Exchange started in 2006 as a single community for computer programmers (Stack Overflow) and expanded in 2010 into a network of stacks related primarily to technology but also additional domains, including business, arts, and science. With 5.1 billion visits in 2018 and 9.5 million registered users (as of January 2019), Stack Exchange is one of the world's most popular web destinations (Stack Exchange About 2019).

The example shown in Figure 2 illustrates the social processes that participants engage in to expand community understanding of complex knowledge topics. The question titled "Create a unterminable process in Windows" was posted by "user20825" to the Information Security stack (Create a ... 2019). (This question is indicative of typical questions across a variety of stacks; Appendix Table A lists titles of questions with similar complexity in other stacks.) The initial comments on the question (shown immediately below the question) provide skeptical responses. The first comment (posted by "tylerl") uses humor to question the question's premise. The second comment (posted by "Bob Watson") helpfully provides three links to external information sources with a summary of their content. Such a response, succinct and authoritative, implies that the question has already been addressed elsewhere. At the same time, two contemporaneous responses (not shown in the figure) express skepticism that the question can be satisfactorily answered. Then, "Polynomial" provides a long, detailed answer that refutes the emerging consensus, summarizing ten different potential methods to achieve the original question's objectives. This answer was marked as an accepted answer by the question author, denoting that it provided a solution to their problem.[2] This answer also received an enthusiastic community response with 93 upvotes, surpassing the 40 upvotes for the question itself. In the comments on this answer, the original question author ("user20825") asked for (and received) additional clarification about the answer while the answer author ("Polynomial") also responded to a question by "SteveS" and received an additional 7 upvotes for the additional information provided.

This exchange illustrates multiple social processes that help shape valued knowledge contributions. On Stack Exchange, participants are encouraged to upvote content they find useful and, alternatively, to downvote questions lacking evidence of prior research, unproductive comments, and incorrect answers. Thus, the community as a whole serves as an interactive audience that, through votes and comments, encourages (or discourages) lines of thinking and focuses attention on both topics of interest as well as the most promising lines of inquiry. The community is itself the arbiter of what is a useful contribution. Further, there is no clear separation between novices, peers, and experts. All participants (regardless of their tenure or expertise) can ask questions and contribute answers. This egalitarian approach is expressed in the website interface, which minimizes participants' identity markers. Participants can optionally disclose information in a personalized profile (including geographic location and links to social media accounts), but quite limited profile information is shown in Q&A threads, and even then, only for the question and answer authors, not individual commenters. In Figure 2, where the answer author's name is shown ("Polynomial"), a small user profile picture appears along with counts and icons signifying a reputation score of over 104K, 36 gold badges, 253 silver badges, and 346 bronze badges. Notably, even when other prolific participants provide valued comments (such as "tylerl" in this example), no additional identification is noted as such on their comments. Finally, while participants remain "on topic" they are also sociable. That is, responses reflect a collective striving for common understanding; meanwhile,

[2] End-user documentation of Stack Exchange features relevant to this example can be found at: https://stackoverflow.com/help/accepted-answer, https://stackoverflow.com/help/privileges/vote-up, https://stackoverflow.com/help/privileges/vote-down, and https://stackoverflow.com/help/badges.

respondents often signal that they have read each other’s responses and frequently refer to each other by name.

4. Data and variables

Our data is based on an archive (Stack Exchange Data Dump 2016) provided by Stack Exchange in December 2016 that contains a complete event history of questions, answers, comments, and votes by all participants in all Stack Exchange communities (also referred to as stacks). Consistent with our interest in mature communities that attract expert participants and enable knowledge contributions, we focused on communities that entered beta status before January 2014 and were promoted to production status by December 2015. We also excluded the idiosyncratic communities of Stack Overflow and the meta, foreign language, and math stacks. Stack Overflow, the first community in the Stack Exchange network, is far larger than other stacks and encompasses a broader range of topics than the more focused communities launched thereafter. Each topic-based stack has an associated meta stack for discussion of stack administration and moderation, which we excluded. In addition, because our analysis of epistemic position is based on linguistic characteristics of contributions, we excluded communities that discuss foreign languages (e.g., French and Russian language stacks) and math, which relies heavily on mathematical formulas and equations. Ultimately, we focused our analysis on the remaining 52 communities that met these criteria.

The data archive contains fine-grained trace data about participants’ activities including the questions, answers, and comments contributed to each community. As illustrated in Figure 3, we organize this data into three levels of analysis: community, community member,[3] and time-stamped contributions (summarized by quarter starting from the inception of each community through the fourth quarter of 2016). Our choice of a quarterly time scale balances the need for sufficient data for analysis with minimizing potentially confounding factors (Zaheer et al. 1999). A shorter time frame (e.g., days) may

[3] In the discussion of the empirical model we use the term *member* interchangeably with *participant* as all participants in Stack Exchange discussions must register as community members in order to contribute content.

only partially capture the interaction around a topic while a much longer time frame (e.g., years) would not accurately capture the network position at the time of contribution. The online appendix provides detailed descriptive statistics for the summarized archival data, including stack characteristics (Appendix Table D) and aggregate growth over time (Appendix Figure B). The same individual can participate in more than one community; however, because each contribution they make can only appear in a single community, we treat their contributions in each community independently. Thus, the unit of analysis for our empirical model is an individual community member's contributions per quarter per community.

4.1 Valued knowledge contribution

We rely on two measures to assess a member's valued knowledge contributions. First, *upvoted posts* measures the number of upvotes received on a member's contributions (questions, answers, and comments). Upvotes reflect the extent to which other community participants found a contribution to be useful. Second, *accepted answers* measures the number of accepted answers for a participant. Accepted answers, detailed by the question author, signal that the answer provided a solution to their problem. Together, these measures show how a member's contributions are recognized and valued within a community. Importantly, both measures reflect the *quality* of contributions, not just the *quantity* of contributions. As noted below, our model includes the quantity of contributions as a control variable.

4.2 Measuring member position in a field

In order to analyze member contributions, we consider each Stack Exchange community to represent a distinct field of knowledge within which each member's position can be established. As illustrated in Figure 4, we identify a member's social position and epistemic position for each quarter for each community they contribute to. We compile a social network and a topic network per quarter per community, both based on members' contributions. Social embeddedness is based on the position of members in the social network formed according to who interacts with whom. Members are considered to have a social tie when they contribute content (question, answer, or comment) associated with the same question (identified as *thread* in Figures 4 and 5). Epistemic marginality is measured based on position in the topic network formed by the content of all member contributions to a community in a quarter. Two members will have a similar epistemic position when they contribute content about similar topics. This epistemic similarity would be entirely distinct from a social tie if they contributed similar content to different threads. Whereas members are nodes in both networks, the social network connects members via co-appearance in threads and the topic network connects members via topics discussed in their contributions. Thus each stack member has a distinct social and epistemic position per quarter per stack based on their position in the member-thread network (social) and member-topic network (epistemic) for the community.

Next, we consider the appropriate measure to assess a member's position in each network. Existing research has taken a variety of approaches. Studies of creative performance have favored "coreness" measures to evaluate the position of actors in a social network (Aadland et al. 2019, Cattani et al. 2014, Cattani and Ferriani 2008). Coreness measures partition the overall network into a number of exclusive groups based on their connectivity. For example, a core-periphery distinction splits the networks into two groups: a core of high density and a periphery of low density (Borgatti and Everett 1999), and a *k*-core decomposition partitions the network into *k* groups where the *k*-core is the group in which every node has at least a degree of *k* (Seidman 1983). Early studies of online innovation communities favored *degree centrality* (e.g., Ahuja et al. 2003, Wasko and Faraj 2005), defined as the number of connections a node in a network has with other nodes. Later studies (e.g., Sutanto et al. 2011) incorporated *closeness centrality*, which is the average length of the shortest paths from a particular node to all other nodes in the network; and, *betweenness centrality*, which is the percentage of shortest paths in the network that a particular node arbitrates.

For a number of conceptual and empirical reasons, we select closeness centrality as an appropriate measure of network position in both the social and the topic networks. First, centrality measures can effectively incorporate weighted network ties. Given that online communities follow a power-law distribution (von Hippel and von Krogh 2003, Johnson et al. 2014) by weighing the edges, closeness centrality takes into account the uneven frequency of participation. Alternative measures, such as coreness, are more appropriate for unweighted networks where all edges are equally important (Borgatti and Everett 1999). Second, centrality measures avoid the assumption that a network is characterized by a single core-periphery structure. Though it may serve as a useful representation, an idealized core-periphery structure is rarely found in social networks (Cattani et al. 2014). Third, among the family of centrality measures, closeness centrality (rather than betweenness or degree) best reflects coreness. As noted by Borgatti and Everett (1999), not all centrality measures reflect coreness. Specifically, betweenness centrality is not an adequate measure to investigate core-periphery tension because it assigns "high values to actors who are not strongly connected to a core group of people, but who link two otherwise unconnected regions of a network" (Borgatti and Everett 1999, p. 393). Likewise, degree centrality captures the first-degree of separation, namely the number of contacts a node has. Although this is a useful manner to identify individual-level social capital (e.g., Wasko and Faraj 2005), degree centrality reflects local connectivity but does not take into account a node's position within the global network. By considering not only direct ties to nodes but also the connectedness of those nodes, closeness centrality better reflects network position as conceptualized according to a core-periphery perspective. Fourth, as detailed in the online appendix, we empirically tested our networks for significant core-periphery structures (Kojaku and Masuda 2018). We found support for the social networks exhibiting a core-periphery structure but not for the topic networks. We analyzed the network structure using a continuous coreness measure (Rombach et al. 2014) and compared it to the closeness centrality approach. The two measures were strongly correlated, which further supports our use of closeness centrality.

4.2.1 Social embeddedness

We create a single network per community per quarter to assess each member's position in the social network. As described further below, we followed a two-step process: (i) for each community, for each quarter, we prepare an affiliation network reflecting all community contributions and (ii) computed closeness centrality for each member. First, we create an affiliation network (e.g., Borgatti and Halgin 2011) based on members (authors) and threads (a question, related comments, related answers, and their related comments). A conceptual illustration of a highly simplified affiliation network is shown in Figure

5 (left) with 3 threads (questions and related content) and 5 members (authors of that content). In this example, in Thread 1, member *A* poses a question and members *B* and *C* provide related content. In the affiliation network representing these contributions, nodes *A*, *B*, and *C* are connected to the node *Thread 1*. When all threads and members are considered, a single social network emerges. In our actual networks, the single social network per community per quarter would typically encompass hundreds or thousands of threads and members. The embeddedness of members in this social network is reflective of their positions in the network. For example, *A*, *B*, and *D* are more highly embedded than *C* and *E* (who are marginal nodes). To reflect the uneven quantity of participation within a thread, the thread-member edges in the affiliation network are weighted based on the number of contributions to that thread. Further, questions that received no answers and no comments are excluded from the affiliation network.

Second, we compute the closeness centrality of a member *u* as the reciprocal of the sum of the shortest path distances from *u* to all *n-1* other nodes. Since the sum of distances depends on the number of nodes in the graph, closeness is normalized by the sum of the minimum possible distances *n-1* (Hagberg et al. 2008). Thus, we define the distance between a member and a thread as the reciprocal of the number of contributions. To illustrate, a thread participant contributing nine comments and one answer to a question has a distance of 0.1 (1 / (1 + 9)). Another participant who makes only two contributions, a comment and an answer, has a distance of 0.5. The first member is closer to the thread than the second member because of the former's higher number of contributions to the thread. This distance function allows the centrality measurement to capture multiple contributions to the same thread, a common occurrence in dialogic knowledge creation communities such as Stack Exchange.

Finally, although the closeness centrality has a theoretical range from 0 to 1, it is unlikely that all networks will have similar ranges in practice. Therefore, to ensure comparable interpretation of the measure across networks of different communities and over time, we normalize the closeness centrality by dividing it over its largest value in the network. In summary, we measure social embeddedness for all (*n*) members (*u*) as the closeness centrality of a member in the thread-member (v*,* *u*) affiliation network:

$$Closeness\ centrality_{network}(u) = \frac{n-1}{\sum_{v=1}^{n-1} distance(v,u)}$$

$$Social\ embeddedness(u) = \frac{Closeness\ centrality_{social\ network}(u)}{Max_{v=1}^{n-1}(Closeness\ centrality_{social\ network}(v))}$$

4.2.2 Epistemic marginality

To calculate epistemic marginality, we create a single topic network per community per quarter. We use this network to establish each member's position in the network of topics discussed by all community members. A community member is high in epistemic marginality if they provide content about topics that are less common and more distinctive compared to other community members' contributions in that quarter. In contrast, a member is low in epistemic marginality if they contribute content about frequently discussed topics that are central to the community's discussion that quarter.

Epistemic marginality reflects a distinct position from social embeddedness both because (a) members can discuss multiple topics in a single thread and (b) the same topic can appear in more than one thread. In Figure 5 (right), we demonstrate a simplified example where two topics are discussed in three threads.[4] Assuming threads 1 and 2 are about Topic 1, nodes *A*, *B*, *C*, and *D* are connected to Topic 1 in the topic network. However, node *D* also connects to Topic 2 because of the involvement of *D* in Topic 2. In this example, member *D* has both high social and epistemic embeddedness—as reflected in their central location in both the member-thread network and also in the member-topic network. Alternatively, members *A* and *B* are socially embedded (i.e., they have high centrality in the member-thread network) yet are epistemically marginal (i.e., they have low centrality in the member-topic network). Finally, members *C* and *E* are both socially and epistemically marginal (i.e., they have low centrality in both networks). As shown in this example, the social and epistemic position has a simple, intuitive

[4] A simplifying assumption of this illustration is that all of the content in each of three example threads is about a single topic; this assumption is relaxed in the actual analysis.

conceptualization; yet, particularly given the challenge of mapping content to topics in a robust manner, the epistemic position is complex to operationalize.

A simplistic approach to mapping content to topics is to use member-selected tags associated with Stack Exchange content. For example, ideally the two topics in Figure 5 (right) may be reflected by tags used by members discussing these topics in the three threads. Unfortunately, this approach is prone to multiple shortcomings. First, Stack Exchange associates tags with questions only. In dialogically intense discussions where multiple ideas are discussed and debated, it is erroneous to consider all responses to a question as comprising the same epistemic position. Indeed, if all the answers to a single question were considered epistemically equivalent, it would greatly restrict the ability to explore the association between epistemic position and the accepted answer outcome variable. Second, tags are assigned by the author of the question and are subject to author idiosyncrasies (Small 2011). For example, question authors may choose tags based on the perceived likelihood of attracting responses, rather than accurate anticipation of the content of associated answers. Third, due to community-specific norms, the nature and quality of tags vary considerably both within the community and across the community. Finally, Stack Exchange offers automatic tag suggestions, making some tag designations an artifact of algorithmic processes rather than well-informed, deliberate choices by stack participants.

As an alternative approach to relying on tags, we use contribution content to identify the network of topics discussed by a community. We can then identify the relative epistemic position of each participant via the topics associated with their content. We use topic modeling, an automated technique for coding the content of text into a set of substantively important coding categories (in this case, topics). Topic modeling has been applied in a variety of organizational and management scholarship (DiMaggio 2015, Grimmer and Stewart 2013, Hannigan et al. 2019, Mützel 2015). As an algorithmic approach that requires minimal human intervention, the method is more inductive than traditional approaches to text analysis in the social and human sciences (Mohr and Bogdanov 2013). Furthermore, the major limitation of topic modeling—the difficulty of retroactively applying intuitive labels to algorithmically defined content collections—is not relevant to our goal of identifying the relative position of participants.

Specifically, we employ a Latent Dirichlet allocation (LDA) topic model (Blei et al. 2003, Blei and Lafferty 2007). In this model, the content of questions, answers, and comments form a set of documents composed of a set of words. Each word is associated probabilistically with a set of topics. These probabilities are estimated from the data by the LDA algorithm. One issue in training the topic model is that the number of topics is one meta parameter that needs to be specified (Mohr and Bogdanov 2013). Because we have no reason to believe that one stack may have a certain number of topics or that different stacks share the same number of topics, we use the number of tags as a rough indicator of the number of topics in each stack. More specifically, we train one topic model per stack per quarter and set the number of topics equal to the number of tags appearing more than once. Although reliance on tags may introduce some of the same data quality issues noted above, we posit that the number of distinct tags used by a community provides a reasonable "order-of-magnitude" estimate for the number of community topics and, importantly, has no discernible upward or downward bias.

Because each document is one single post (question, answer, or comment), we can aggregate the distribution of topics in documents by members. This approach is a particular case of the more general author-topic model where documents are collaboratively created by multiple authors (Rosen-Zvi et al. 2010). We use this aggregate distribution to connect members to topics with an affiliation network that we refer to as the topic network. An example topic network is shown on the right side of Figure 6. Similar to the social network, the topic network is an affiliation network in which members connect to topics. Likewise, it is possible to use the same closeness centrality measurement to compute the epistemic embeddedness of members in their communities. To apply the centrality measure, we use the reciprocal of P(member | topic) as a distance function representing the distance between a member and associated topics. For example, if two members are associated with the same topic with different probabilities, the member with the higher probability is closer to the topic than the member with the lower probability.

$$P(member|topic) = \frac{\sum_{post \in member\ posts} P(post|topic)}{|member\ posts|}$$

To operationalize epistemic *marginality* (rather than embeddedness), we subtract the closeness centrality value from the largest value of centrality in the network. The member with the highest centrality becomes the one with the lowest marginality and vice-versa. Finally, we normalize the measure by dividing it over the largest value of closeness centrality. This measure yields value ranges from 0 to 1 in all networks and thus is useful for making comparisons across communities over time. We note here that the measures, social embeddedness and epistemic marginality, have comparable interpretation given the similar structure of the social and topic networks and the underlying use of closeness centrality to evaluate position in the two networks. This similarity is useful for comparing the outcomes of the two theoretical constructs.

$$Epistemic\ marginality(u) = \frac{Max_{v=1}^{n-1}\left(Closeness\ centrality_{topic\ network}(v)\right) - Closeness\ centrality_{topic\ network}(u)}{Max_{v=1}^{n-1}\left(Closeness\ centrality_{topic\ network}(v)\right)}$$

To illustrate the difference between member social and epistemic position, Figure 6 plots the social network and the topic network in the Stack Exchange GIS community (GIS SE 2019) in 2013 Quarter 3 side-by-side. The layout is a force-directed graph where the relative distance between nodes is proportional to their closeness to the same threads or topics. For ease of interpretation, thread nodes, topic nodes, and edges are not shown in the figure. Darker colors represent contributors with a higher number of valued knowledge contributions (as measured by accepted answers). The social network (a) shown on the left exhibits a visible core-periphery structure in which members of high centrality are densely clustered. Members of decreasing centrality form sparser layers around the core. In the topic network (b) shown on the right, however, a core-periphery structure is less evident. Beyond this distinction, Figure 6 shows the relationship between the position of the same nodes in the social network and their corresponding position in the topic network. Nodes are consistently colored in both networks, with darker nodes providing higher-quality contributions than lighter nodes. In the social network (left) higher-quality contributions (darker nodes) are more likely to appear in the central core. Alternatively, in the topic network (right) higher-quality contributions (darker nodes) are more likely to appear toward the peripheral edges.

4.3 Control variables

We control for several factors that may provide alternative explanations for members' valued knowledge contribution. We have two sets of controls: community-level control variables and member-level control variables. As with the dependent and independent variables, these variables are evaluated for each community for each quarter observed.

4.3.1 Community-level control variables

To account for the large heterogeneity of the 52 communities (Appendix Table C), we control for multiple community-level variables: *community age* (the number of quarters since the inception of the community), *community membership* (which is the number of members in the community), and *community maturity* (the average reported age of community members). We also control for *community crowding*, which is the existence of multiple new questions in the community competing for the attention of members available to answer. Crowding is shown to play an important role in problem-solving communities (Haas et al. 2015, Piezunka and Dahlander 2015). From the knowledge-provider perspective, members who answer questions have limited attention, and although moderate levels of crowding may stimulate attention, high levels ultimately dilute attention and lead to decreased contribution (Haas et al. 2015). We operationalize community crowding as the number of new questions posted in a quarter divided over the number of active members in the community that quarter.

4.3.2 Member-level control variables

Because members can be affiliated with multiple communities, spillover effects may exist across communities. For example, generalized reciprocity suggests that if the member was helped in one community, they may be more willing to help in another community (Faraj and Johnson 2011). Therefore, we control for the number of *concurrent communities* a member is affiliated with. Likewise, members choose the extent to which they disclose their real identities in their communities. Stack

Exchange enables members to curate a personal profile in each of their communities. This profile has five components: name, age, location, personal photo, and web address. We operationalize *pseudonymity* as the number of missing components in the member profile. Also, members' patterns of contribution can affect how such contribution is valued. The order of answers and comments in question threads can play a role in having these posts accepted and upvoted by others. In particular, a first-mover advantage suggests that the earlier responder will be more likely to receive acceptance and upvotes. We included the variables *comment rank* and *answer rank* to measure the median rank of a member's comments and answers to questions, respectively. Higher rank means later responses overall while early responses overall correspond with a lower rank. We also account for members' *total contributions* measured as the total number of questions, answers, and comments made by a member within a community in a quarter. Members who post more have more opportunities to receive upvotes and accepted answers. We also control for members' *tenure* in their communities by considering the (fractional) number of years since their first post.

As the sole medium of communication in online communities, language plays a significant role in driving interactions among members. That is, how members express knowledge in their contributions can influence the value of their contributions. As a result, we control for the average number of *links per post*. In online innovation communities such as Stack Exchange, members sometimes refer to external resources or other threads within the community. In this context, providing links substitutes for writing original content. We also control for the average number of *mentions per post* because addressing members directly may affect how others react to the content (Johnson et al. 2015). For example, more mentions may increase the likelihood of getting an upvote from the mentioned member. We control for the *readability* of posts using the SMOG index (McLaughlin 1969). This index estimates the years of education needed to understand a piece of text. We reverse code the index in order to measure readability on a scale of simplicity rather than complexity. Further, we control for the *language prototypicality* of each member, defined as using a language similar to that used by other members, or language that is prototypical in the community. Using a corpus of question, answer, and comment texts in a stack, we train a trigram statistical language model (SLM) with Lidstone's additive smoothing (Chen and Goodman 1996). This model estimates the probabilities of short sequences of words (with lengths of 1, 2, and 3 words) based on their frequency in the text. These probabilities can then be used to evaluate the prototypicality of new text (Johnson et al. 2015). In particular, we computed the cross-entropy of the trigram model for a given new text. Because we are interested in prototypicality, we reverse code the cross-entropy score. Next, the linguistic prototypicality of each member is the average of the prototypicality of their contributions.

$$Prototypicality(post, SLM) = -H(post, model) = \frac{\sum_{t \in trigrams\ in\ post} P(t|SLM)}{|trigrams\ in\ post|}$$

$$Prototypicality(member) = \frac{\sum_{post \in member's\ post} Prototypicality(post, SLM)}{|member's\ post|}$$

Finally, following Johnson et al. (2015), we control for members' *positive sentiment* using AFINN, a word-based classifier of sentiments based on a dictionary of emotionally rated English words tailored to the Internet language of blogs, discussion forums, and tweets (Nielsen 2011).

Table 1 summarizes the research variables and measurements.

5. Model

We use multilevel analysis to model the nested structure in the data. Specifically, we use a three-level mixed-effects model to represent the quarterly contributions of members within communities. This type of model supports the careful identification required to account for the significant heterogeneity of the communities and members. Outcomes can be impacted by community-level factors, member-level factors, and time-specific factors; thus, grouping and collapsing data drawn from heterogeneous populations could lead to biased results and erroneous conclusions. Further, observations from the same groups (e.g., members and communities) are likely to be more similar to each other than observations from different groups. This similarity violates the assumption of ordinary least squares regression that observations are independent (Hox et al. 2017, p. 16).

Multilevel mixed-effects modeling enables us to capture the heterogeneity across communities and members. In particular, we model valued knowledge contribution (upvoted posts and accepted answers) in a quarter (q) for a member (m) within a community (c) as a function of our research variables as fixed-effects (X). In addition, as random-effects, we include a random intercept that differs per community (v_{0c}) and member (u_{0mc}). The random member-intercept captures the unobserved heterogeneity across members in the same community, whereas the random community-intercept captures the heterogeneity across communities. The inclusion of community age and member tenure within the community controls for time effects on contributions.

$$\text{Valued Knowledge contribution}_{qmc} = \pi_{0mc} + \sum_{\substack{i=1 \\ X \in vars}}^{|vars|} \beta_{i00} X_{qmc} + e_{qmc}$$

$$\text{where} \quad \pi_{0mc} = \beta_{000} + v_{0c} + u_{0mc}$$

This equation was estimated using three steps and appears below (without the subscript notation). In step (1) we include all community-level and member-level control variables. In step (2) we introduce the social embeddedness and epistemic marginality variables, and in step (3) we include their interaction effect. We estimate each step for each of the two dependent variables measuring valuable knowledge contribution: upvoted contributions and accepted answers. Both dependent variables are counts, and they are overdispersed (Table 2). Multilevel count models such as Poisson and negative binomial are not feasible because these specifications rely on computational optimization in which convergence is slow and sometimes fails with large data sets like ours (Rabe-Hesketh and Skrondal 2008, West et al. 2014). Instead of using a count model, we control for the dispersion in the dependent variables by taking their natural logarithms (Wooldridge 2010, p. 723). Finally, we only include members with at least one comment or one answer in models with the upvoted posts dependent variable (1, 2, and 3), and we only include members with at least one answer in models for the accepted answers dependent variable (4, 5, and 6). This constraint is important to reduce the number of irrelevant observations. Accordingly, contributing an answer is necessary to get an answer accepted. Including members without an answer in the sample would zero-inflate the dependent variable. This exclusion yields different sample sizes for the two dependent variables: 1,574,560 observations (i.e., quarterly member contributions) for upvotes and 734,911 observations for accepted answers.

$$\begin{aligned}\log(\text{Upvoted posts}) = {} & \beta_0 + \beta_1\text{Community age} + \beta_2\text{Community membership} + \beta_3\text{Community crowding} \\ & + \beta_4\text{Community maturity} + \beta_5\text{Concurrent communities} + \beta_6\text{Pseudonymity} + \beta_7\text{Comment rank} \\ & + \beta_8\text{Answer rank} + \beta_9\text{Total contributions} + \beta_{10}\text{Tenure} + \beta_{11}\text{Links per post} + \beta_{12}\text{Mentions per post} \\ & + \beta_{13}\text{Readability of posts} + \beta_{14}\text{Language prototypicality} + \beta_{15}\text{Positive sentiment}\end{aligned}\tag{1}$$

$$+ \beta_{16}\text{Social embeddedness} + \beta_{17}\text{Epistemic marginality}\tag{2}$$

$$+ \beta_{18}\text{Social embeddedness} \times \text{Epistemic marginality}\tag{3}$$

$$\begin{aligned}\log(\text{Accepted answers}) = {} & \gamma_0 + \gamma_1\text{Community age} + \gamma_2\text{Community membership} + \gamma_3\text{Community crowding} \\ & + \gamma_4\text{Community maturity} + \gamma_5\text{Concurrent communities} + \gamma_6\text{Pseudonymity} + \gamma_7\text{Comment rank} \\ & + \gamma_8\text{Answer rank} + \gamma_9\text{Total contributions} + \gamma_{10}\text{Tenure} + \gamma_{11}\text{Links per post} + \gamma_{12}\text{Mentions per post} \\ & + \gamma_{13}\text{Readability of posts} + \gamma_{14}\text{Language prototypicality} + \gamma_{15}\text{Positive sentiment}\end{aligned}\tag{4}$$

$$+ \gamma_{16}\text{Social embeddedness} + \gamma_{17}\text{Epistemic marginality}\tag{5}$$

$$+ \gamma_{18}\text{Social embeddedness} \times \text{Epistemic marginality}\tag{6}$$

6. Findings

The results of the regression analyses are illustrated in Table 3. First, all analyses yield significant goodness of fit (χ^2 and log-likelihood). The model-difference tests between (1) & (2), (2) & (3), (4) & (5), and (5) & (6) are all significant. Next, we note that the coefficients of the community-level and member-level control variables are consistent with prior findings in the literature on online communities. Community age has negative associations with both dependent variables ($\beta_1 = -.0178^{***}$& $\gamma_1 = -.0055^{***}$) which aligns with prior findings of the degradation of contributions in online communities over time (Butler 2001). Similar to Haas et al. (2015), community crowding has a positive effect ($\beta_3 = .4613^{***}$ & $\gamma_3 = .0951^{***}$), which suggests that communities with more questions attract more contributions from members. Member pseudonymity is negatively associated with both dependent variables ($\beta_6 = -.0208^{***}$& $\gamma_6 = -.0287^{***}$), which supports a social view of knowledge contribution such that members who self-disclose their identities are more likely to be recognized for their valued knowledge. Both comment rank and answer rank have negative coefficients, suggesting that early participants have a higher chance of acquiring recognition for their contributions. Similarly, high volume participants have a better chance of earning recognition for their contributions as evidenced by the positive coefficient of the total number of contributions ($\beta_9 = .0122^{***}$& $\gamma_9 = .0070^{***}$). The qualitative aspects of contributions are also important in determining their value. Links per post have a positive coefficient (β_{11} = .2067***& γ_{11} = .0332***), suggesting that providing external references is valuable. Mentions per post are also positive (β_{12} = .0126***& γ_{12} = .0057***), which again suggests that knowledge contribution is a social process where members who interact with others have a better chance of earning recognition for their contributions. Similar to Johnson et al. (2015), we find a positive coefficient of the readability of posts (β_{13} = .0196***& γ_{13} = .0130***) and language prototypicality (β_{14} = .0055***& γ_{14} = .0006**), findings that suggest that members using the linguistic norms of the community have a better chance of gaining recognition (Danescu-Niculescu-Mizil et al. 2013).

Of note, between the two dependent variables, tenure and positive sentiment have opposing effects. Member tenure has a negative effect on upvoted posts but a positive effect on accepted answers (β_{10} =-.0409***& γ_{10} = .0040*). This contrast suggests that the experience that accompanies tenure is recognized by the knowledge seeker but not necessarily by the audience. Similarly, whereas positive sentiment is positively associated with upvotes (β_{15} = .0764***& γ_{15} = -.0066*), it is negatively associated with accepted answers, a contrast suggesting that the audience is sensitive to the sentiment of the post while the knowledge seeker is not.

In support of P1 and P2, in models (2) & (5) we find consistent, large, significant effects for both measures of knowledge contribution. Social embeddedness is positively associated with knowledge contribution (β_{16} = 1.1183***& γ_{16} =.2170***), and epistemic marginality is positively associated with knowledge contribution (β_{17} = 1.4763***& γ_{17} =.5965***). Furthermore, these effects are practically significant. A unit increase in social embeddedness results in a 205% increase in upvoted posts (exp(1.1183) - 1 = 2.05) and a 24% increase in accepted answers. Because social embeddedness has a range from zero to one, we can relate a 1% increase in social embeddedness to a 2% increase in upvoted posts and a 0.24% increase in accepted answers. Similarly, a unit increase in epistemic marginality results in a 337% increase in upvoted posts and an 81% increase in accepted answers. Because epistemic marginality also ranges from zero to one, we can connect a 1% increase in epistemic marginality to a 3.37% increase in upvoted posts and a 0.81% increase in accepted answers.

In support of P3, the interaction effect between social embeddedness and epistemic marginality in models (3) and (6) is positively associated with upvoted posts and accepted answers (β_{18} = 2.7511***& γ_{18} =1.5713***). Thus, we find that members who are socially embedded and members who are epistemically marginal are more likely to make valuable knowledge contributions. In addition, social embeddedness moderates the relationship between epistemic marginality and valued knowledge contributions. Figure 7 plots this interaction. Consider an epistemically core member with epistemic marginality at its minimum value of 0. Holding all other variables constant, a unit increase of social embeddedness from (almost) zero to one is estimated to result in increasing log(Upvoted posts) from .28 to .77. This .48 unit increase translates to 63% more upvotes (1-exp(.48)). However, for a member with epistemic marginality at its maximum value of one, holding all other values constant, a unit increase in social embeddedness results in increasing log(Upvoted posts) from 0.08 to 3.3. This 3.2 unit increase translates to a 240% increase in upvoted posts. This four-fold difference is represented by the difference in the slopes of the two linear relationships in Figure 7, on the left. The right side of Figure 7 illustrates this relationship for accepted answers. Here, we see a more pronounced interaction. Among epistemically marginal participants, those with high social embeddedness are more likely to provide contributions valued by the community.

We performed extensive robustness checks to ensure the validity of the results under different model specifications and sampling approaches. The following robustness checks are detailed in the appendix: sensitivity analyses (excluding members with certain characteristics and restricting to different subsamples), alternative model specifications (cross-sectional analysis of the last quarter with the inclusion of moderator role, and Heckman specification to address self-selection). We also provide further model diagnostics (multicollinearity and omitted variables). Overall, our primary findings remain supported under the preponderance of these checks, thereby providing increased confidence in the validity of our findings under alternative choices of data sampling and model specification.

7. Discussion and conclusion

The goal of this paper is to identify where valued knowledge contributions originate in online innovation communities. To address this question, we analyzed contributions by 697,412 participants in 52 Stack Exchange Q&A communities. Going beyond traditional conceptualizations that treat position in social and knowledge fields as concomitant, we provide a more nuanced differentiation. We consider two distinct positions: social embeddedness (a core position within the social ties network) and epistemic marginality (a peripheral position within the topics network). We find that participants who are socially embedded and participants who are epistemically marginal both contribute knowledge valued by the community. Additionally, among epistemically marginal participants, those who are more socially embedded are more likely to provide knowledge contributions valued by the community. We conclude that neither explanations offered in favor of socially core participants nor those that emphasize the role of epistemically peripheral participants alone fully explain the contribution of valued knowledge. By distinguishing social and epistemic positions, our findings contribute to understanding how core-periphery tension manifests in online innovation communities.

7.1 Implications for the core-periphery debate in knowledge fields

We make multiple contributions to the core-periphery debate in knowledge fields. First, existing research emphasizing the social dimension of knowledge production has focused on the situated, tacit nature of knowledge, the presence of shared epistemic culture, and the importance of social and dialogic processes among core community participants (e.g., Brown and Duguid 2001, Knorr-Cetina 1999, Tsoukas 2009). Our findings indicate that in the context of online innovation communities, social embeddedness only partially explains knowledge contribution. While social embeddedness is positively associated with providing valued knowledge contributions, so, too, is epistemic marginality. This additional consideration may be rooted in the relative lack of formal social structures in online innovation communities. In contrast to more formalized fields of knowledge, online communities have more fluid boundaries, participation is less controlled, and social status cues are less present. As a result, the overlap between epistemic contributions and the social center of the field may be weaker. Thus, outsiders with good ideas may find lower social barriers to gaining recognition for their contributions.

Second, existing research emphasizing the epistemic dimension of knowledge production has focused on the importance of novel contributions by peripheral participants. Peripheral participants have access to new ideas and new ways of thinking about problems, and they are unconstrained by dominant thinking in a field. Thus, they are better able to offer solutions that appear novel to central participants (e.g., Boudreau et al. 2016, Jeppesen and Lakhani 2010). Our findings indicate that being epistemically marginal only partially explains knowledge contribution. In contrast to open innovation contests where there is minimal interaction among participants or, even, between knowledge seekers and knowledge providers, online innovation communities are representative of settings with more extensive epistemic and social interaction. In such settings, the positive merits of a novel idea alone may be insufficient to successfully navigate the complex social processes involving the values, assumptions, and ways of knowing of the local community.

Third, our findings expand upon prior innovation research that highlights barriers social outsiders face in their attempts to gain support from others for their novel ideas. The process of acquiring such support requires the persistent travail of convincing potential allies and contextualizing innovative ideas in the field (e.g., Cattani et al. 2014, 2017). Our findings align with and expand upon these insights. Previous literature has focused on the innovator as an outsider trying to gain acceptance in a reluctant field (e.g., John Harrison's invention of the marine chronometer). Our finding of the importance of social embeddedness as a complement to epistemic marginality can shed some light on the process of novelty adoption. For example, Harrison's journey with the Royal Society could be partially seen as a process of socialization; accordingly, acceptance of his idea was only achieved with significant effort expended on gaining allies and earning the goodwill of central figures in the Royal Society. Thus, a challenge for outside innovators is not only to establish the technical merit of their ideas but also to engage with the accompanying social processes necessary to have their ideas accepted by a field.

Fourth, peripheral participants face a number of knowledge-related barriers to acceptance of their ideas. These include knowledge differences and pragmatic boundaries (Carlile 2002), as well as gatekeeping by influential insiders (Cattani et al. 2017, Lifshitz-Assaf 2018). By differentiating social and epistemic positions, we are able to analyze how these positions, both separately and together, influence the production of valued knowledge. When epistemically marginal contributions are made by a socially embedded participant, they may realize not only the benefits of outsider thinking commonly attributed to the periphery but also the benefits of insider social position typically attributed to the core. Thus, we find a crucial role for participants in online innovation communities who are socially embedded and able to offer novel ideas. Additional research is needed to establish the extent to which these same processes apply to other innovation contexts.

Finally, our study addresses core knowing processes in communities formed to resolve problems of practice and doing. Previous work has focused on the knowledge conversion processes (e.g. tacit to explicit) taking place in such online communities (Faraj et al. 2016, Majchrzak and Malhotra 2016, Nonaka and von Krogh 2009). Our study points to the pragmatic nature of knowledge that develops in these online innovation communities. Contrary to the dominant literature that takes a "justified true belief" perspective on knowing (with an emphasis on objective reality and science-like processes of knowledge validation), what passes for knowledge in online innovation communities has a more pragmatist flavor with the emphasis on what works in practice and what is useful. Because of the community validation process, the knowledge that emerges in these communities is practice-oriented, actionable, and useful in daily life. Thus, future studies of innovation communities may benefit from a reorientation toward a pragmatist view of knowledge.

7.2 Implications for online innovation communities

Our study has implications for understanding what it means to be socially embedded online. Much of the previous literature about online behaviors has emphasized the motivation, identity, and utilitarian and reputational goals that drive online participation (e.g., Kraut and Resnick 2011, Ma and Agarwal 2007).

Our study explores, more specifically, social interactions and epistemic positions of participants within an online community over 7 years and across 52 groups. Rather than taking an individual motivation or an individual behavior perspective, we focus on the production of knowledge resulting from complex social processes. Our findings point to the importance of norms and values as essential aspects of social dynamics online. Interestingly, given the open, pseudonymous participation, limited formal roles, and lean personal profiles of online innovation communities, the social status hierarchies present in many social settings are largely missing here. Instead, social position is achieved through participation—specifically, who interacts with whom. At the same time, interactions form on the basis of shared interests—what conversation a participant decides to join and which fellow community participants are attracted to the conversations a participant initiates. Thus, just like in primarily face-to-face settings, social embeddedness in online innovation communities relates to important behaviors, processes, and outcomes. Yet, how sociality and embeddedness operate online remains under-explored.

Second, our findings about the role of epistemic marginality and social embeddedness have implications for the emergent understanding of epistemic processes online. Current research on crowdsourcing and innovation contests stresses the value of online platforms and technologies as a way to attract outsiders, facilitate their participation, and maximize the provision of new ideas and contributions (e.g., Bayus 2012, Jeppesen and Lakhani 2010, Lifshitz-Assaf 2018). Our findings suggest that this focus may understate key social processes when participants interact extensively. Like Lakatos's concept of programs of research or Knorr-Cetina's epistemic communities, online innovation communities seem to engage in a highly dialogic process focused on deepening the understanding of a complex issue and exploring how to apply such knowledge situationally. Nonetheless, online epistemic communities are limited by the same openness to participation that makes them so effective. They struggle with how to attract and retain participants with diverse perspectives while still maintaining norms of engagement and sustaining social harmony. Thus, in contrast to framing online innovation communities as repositories of straightforward questions with simply stated answers, we find that they represent a collective ongoing accomplishment of articulating, understanding, and advancing knowledge important to the community.

Finally, online innovation communities foreground the importance of technology for both enabling knowledge exchange and supporting social interactions. That is, the technology platform developed by Stack Exchange plays a crucial role in structuring interactions. Platform designers shape participation through selected reward structures (e.g., reputation scores), by prioritizing content visibility (e.g., top question lists), and by channeling community feedback (e.g., displaying votes). Further, this research serves as an example of how to leverage the availability of platform data to engage in big data research that builds and tests theory. Our study encompasses 697,412 participants of 52 distinct knowledge communities making 11,697,355 total contributions over the course of 7 years. Avoiding the limitations of empirical research that merely expose hidden empirical relationships, we started with a theoretical puzzle and then proceeded to collect and analyze large scale trace data. Making inferences from the analysis of big data requires deeply understanding the context, ensuring that trace data reflects accurately what people actually do in that context, and deploying appropriate methods for organizing and analyzing large data sets (Berente et al. 2018, Johnson et al. 2019, Lazer et al. 2009). Thus, careful deployment of computational social science approaches, coupled with tight guidance from existing theory, can provide an effective new way to extend our understanding of online phenomena and social science theorizing in general.

7.3 Conclusion

The results of this study suggest several opportunities for future research. Although our setting encompasses 52 different communities covering diverse knowledge topics, all of the Stack Exchange communities we analyzed share the same technology platform with the same participation channels and community-based governance. Although our results are consistent across different types of stacks and across an extended period of time, our research design privileges the meso-level in its focus on collective knowledge production. An opportunity exists for research focused on identifying the trajectory of contributions of participants within a community as well as the conditions under which influential participants are more likely to emerge. Other boundary conditions for our theorizing include the following: participation is open to anyone; participants are largely (though not exclusively) pseudonymous; there is no formal role differentiation between knowledge seekers and knowledge providers; and, the audience for contributions consists of fellow community participants. As a result, further work is needed to establish whether our model holds for other settings, including different knowledge domains, platforms, and governance structures.

The challenge of balancing the novelty that outsiders provide with the ability to integrate new and existing knowledge is a universal challenge for established fields of knowledge. As online innovation communities become increasingly prevalent across a variety of settings, it is important to understand how core-periphery tension manifests in online settings. The visibility of online interactions enables the differentiation between social embeddedness and epistemic marginality as two distinct positions in a field. We found that both socially embedded and epistemically marginal participants provide valued knowledge contributions to online innovation communities. Importantly, among epistemically marginal participants, those with more social embeddedness are even more likely to provide contributions valued by the community. Thus, this study suggests that valued knowledge contributions arise from participants who not only are able to offer novel information, viewpoints, or perspectives but also are able to navigate complex social and epistemic expectations by engaging effectively in conversations that are central to the community.

References

Aadland E, Cattani G, Ferriani S (2019) Friends, Cliques and Gifts: Social Proximity and Recognition in Peer-Based Tournament Rituals. *Academy of Management Journal* 62(3):883–917.

Afuah A, Tucci CL (2012) Crowdsourcing as a solution to distant search. *Academy of Management Review* 37(3):355–375.

Ahuja MK, Galletta DF, Carley KM (2003) Individual centrality and performance in virtual R&D groups: An empirical study. *Management Science* 49(1):21–38.

Anon (2016) Stack Exchange Data Dump. *Internet Archive*. Retrieved (February 11, 2017), https://archive.org/details/stackexchange.

Anon (2019) GIS SE. *Geographic Information Systems Stack Exchange*. Retrieved (August 1, 2019), https://gis.stackexchange.com/.

Anon (2019) Stack Exchange About. Retrieved (August 1, 2019), https://stackexchange.com/about.

Anon (2019) Stack Exchange Tour. Retrieved (August 1, 2019), https://stackexchange.com/tour.

Anon (2019) Create a ... *Information Security Stack Exchange*. Retrieved (August 5, 2019), https://security.stackexchange.com/questions/30985/create-a-unterminable-process-in-windows.

Argote L, McEvily B, Reagans R (2003) Managing knowledge in organizations: An integrative framework and review of emerging themes. *Management science* 49(4):571–582.

Bayus BL (2012) Crowdsourcing New Product Ideas over Time: An Analysis of the Dell IdeaStorm Community. *Management Science* 59(1):226–244.

Berente N, Seidel S, Safadi H (2018) Research Commentary—Data-Driven Computationally Intensive Theory Development. *Information Systems Research* 30(1):50–64.

Blei DM, Lafferty JD (2007) A correlated topic model of Science. *The Annals of Applied Statistics* 1(1):17–35.

Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. *the Journal of machine Learning research* 3:993–1022.

Bogers M, Zobel AK, Afuah A, Almirall E, Brunswicker S, Dahlander L, Frederiksen L, et al. (2017) The open innovation research landscape: established perspectives and emerging themes across different levels of analysis. *Industry and Innovation* 24(1):8–40.

Boland RJ, Tenkasi RV (1995) Perspective making and perspective taking in communities of knowing. *Organization Science* 6(4):350–372.

Borgatti SP, Everett MG (1999) Models of core/periphery structures. *Social Networks* 21(4):375–395.

Borgatti SP, Halgin DS (2011) Analyzing affiliation networks. *The Sage handbook of social network analysis*:417–433.

Borgatti SP, Lopez-Kidwell V (2011) Network Theory. Scott J, Carrington PJ, eds. *The SAGE handbook of social network analysis*. (SAGE publications, London; Thousand Oaks, Calif), 40–54.

Boudreau KJ, Guinan EC, Lakhani KR, Riedl C (2016) Looking across and looking beyond the knowledge frontier: Intellectual distance, novelty, and resource allocation in science. *Management Science* 62(10):2765–2783.

Bourdieu P (1993) *The field of cultural production: Essays on art and literature* (Columbia University Press).

Brown JS, Duguid P (2001) Knowledge and organization: A social-practice perspective. *Organization Science*:198–213.

Butler BS (2001) Membership size, communication activity, and sustainability: A resource-based model of online social structures. *Information systems research* 12(4):346–362.

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

Cattani G, Ferriani S (2008) A Core/Periphery Perspective on Individual Creative Performance: Social Networks and Cinematic Achievements in the Hollywood Film Industry. *Organization Science* 19(6):824–844.

Cattani G, Ferriani S, Allison PD (2014) Insiders, Outsiders, and the Struggle for Consecration in Cultural Fields: A Core-Periphery Perspective. *American Sociological Review*.

Cattani G, Ferriani S, Lanza A (2017) Deconstructing the Outsider Puzzle: The Legitimation Journey of Novelty. *Organization Science* 28(6):965–992.

Chen SF, Goodman J (1996) An empirical study of smoothing techniques for language modeling. *Proceedings of the 34th annual meeting on Association for Computational Linguistics*. (Association for Computational Linguistics), 310–318.

Collins R (1998) *The Sociology of Philosophies: A Global Theory of Intellectual Change* (Harvard University Press, Cambridge, MA).

Dahlander L, Frederiksen L (2012) The Core and Cosmopolitans: A Relational View of Innovation in User Communities. *Organization Science* 23(4):988–1007.

Dahlander L, O'Mahony S (2011) Progressing to the Center: Coordinating Project Work. *Organization Science* 22(4):961–979.

Danescu-Niculescu-Mizil C, West R, Jurafsky D, Leskovec J, Potts C (2013) No country for old members: User lifecycle and linguistic change in online communities. *Proceedings of the 22nd International World Wide Web Conference*.

DiBenigno J (2019) Rapid Relationality: How Peripheral Experts Build a Foundation for Influence with

Line Managers. *Administrative Science Quarterly*:0001839219827006.

DiMaggio P (2015) Adapting computational text analysis to social science (and vice versa). *Big Data & Society* 2(2):205395171560290.

Faraj S, Jarvenpaa SL, Majchrzak A (2011) Knowledge collaboration in online communities. *Organization Science* 22(5):1224–1239.

Faraj S, Johnson SL (2011) Network exchange patterns in online communities. *Organization Science* 22(6):1464–1480.

Faraj S, von Krogh G, Monteiro E, Lakhani KR (2016) Online Community as Space for Knowledge Flows. *Information Systems Research* 7047:1–17.

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

Feyerabend P (1975) *Against method* (NLB, London).

Gieryn TF, Hirsh RF (1983) Marginality and innovation in science. *Social studies of science* 13(1):87–106.

Grimmer J, Stewart BM (2013) Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis* 21(03):267–297.

Haas MR, Criscuolo P, George G (2015) Which problems to solve? Attention allocation and online knowledge sharing in organizations. *Academy of Management Journal* 58(3):680–711.

Hadjimichael D, Tsoukas H (2019) Towards a better understanding of tacit knowledge in organizations: Taking stock and moving forward. *ANNALS*.

Hagberg A, Swart P, S Chult D (2008) Exploring network structure, dynamics, and function using NetworkX.

Hampton KN (2017) Studying the digital: Directions and challenges for digital methods. *Annual Review of Sociology* 43:167–188.

Hannigan T, Haans RFJ, Vakili K, Tchalian H, Glaser V, Wang M, Kaplan S, Jennings PD (2019) Topic modeling in management research: Rendering new theory from textual data. *Academy of Management Annals* (ja).

Hargadon AB, Douglas Y (2001) When innovations meet institutions: Edison and the design of the electric light. *Administrative science quarterly* 46(3):476–501.

von Hippel E (1994) “Sticky information” and the locus of problem solving: implications for innovation. *Management Science* 40(4):429–439.

von Hippel E (2005) *Democratizing innovation* (the MIT Press).

von Hippel E (2017) Free Innovation by Consumers—How Producers Can Benefit: Consumers’ free innovations represent a potentially valuable resource for industrial innovators. *Research-Technology Management* 60(1):39–42.

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.

Hox JJ, Moerbeek M, Van de Schoot R (2017) *Multilevel analysis: Techniques and applications* (Routledge).

Jeppesen LB, Lakhani K (2010) Marginality and Problem-Solving Effectiveness in Broadcast Search. *Organization Science* 21(5):1016–1033.

Johnson SL, Faraj S, Kudaravalli S (2014) Emergence of Power Laws in Online Communities: The Role of Social Mechanisms and Preferential Attachment. *MIS Quarterly* 38(3).

Johnson SL, Gray P, Sarker S (2019) Revisiting IS research practice in the era of big data. *Information and Organization* 29(1):41–56.

Johnson SL, Safadi H, Faraj S (2015) The emergence of online community Leadership. *Information Systems Research* 26(July):35–68.

Knorr-Cetina K (1999) *Epistemic cultures: How the sciences make knowledge* (Harvard University Press).

Kojaku S, Masuda N (2018) A generalised significance test for individual communities in networks. *Scientific reports* 8(1):7351.

Kraut RE, Resnick P (2011) Encouraging contribution to online communities. *Building successful online communities: Evidence-based social design*:21–76.

Kudaravalli S, Faraj S (2008) The structure of collaboration in electronic networks. *Journal of the Association for Information Systems* 9(10/11):706–726.

Lakatos I (1978) *The Methodology of Scientific Research Programmes*

Lakhani KR (2016) Managing communities and contests to innovate with crowds. *Revolutionizing Innovation: Users, Communities, and Open Innovation*. (MIT Press Cambridge, MA), 109.

Latour B (2010) Tarde's idea of quantification. *The Social After Gabriel Tarde: Debates and Assessments*. 145–162.

Lave J, Wenger E (1991) *Situated learning: Legitimate peripheral participation* (Cambridge university press).

Lazer D, Pentland AS, Adamic L, Aral S, Barabasi AL, Brewer D, Christakis N, et al. (2009) Life in the network: the coming age of computational social science. *Science* 323(5915):721.

Lazer D, Radford J (2017) Data ex machina: introduction to big data. *Annual Review of Sociology* 43:19–39.

Levina N, Arriaga M (2014) Distinction and Status Production on User-Generated Content Platforms: Using Bourdieu's Theory of Cultural Production to Understand Social Dynamics in Online Fields. *Information Systems Research* 25(3):468–488.

Lifshitz-Assaf H (2018) Dismantling Knowledge Boundaries at NASA: The Critical Role of Professional Identity in Open Innovation. *Administrative Science Quarterly* 63(4):746–782.

Longino HE (2002) *The fate of knowledge* (Princeton University Press).

Ma M, Agarwal R (2007) Through a glass darkly: Information technology design, identity verification, and knowledge contribution in online communities. *Information Systems Research* 18(1):42.

Majchrzak A, Griffith TL, Reetz DK, Alexy O (2018) Catalyst organizations as a new organization design for innovation: The case of hyperloop transportation technologies. *Academy of Management Discoveries* 4(4):472–496.

Majchrzak A, Malhotra A (2016) Effect of knowledge-sharing trajectories on innovative outcomes in temporary online crowds. *Information Systems Research* 27(4):685–703.

McLaughlin GH (1969) SMOG grading: A new readability formula. *Journal of reading* 12(8):639–646.

Merton RK (1973) *The sociology of science: Theoretical and empirical investigations* (University of Chicago press).

Mohr JW, Bogdanov P (2013) Introduction—Topic models: What they are and why they matter. *Poetics* 41(6):545–569.

Mützel S (2015) Facing Big Data: Making sociology relevant. *Big Data & Society* 2(2):205395171559917.

Nahapiet J, Ghoshal S (1998) Social capital, intellectual capital, and the organizational advantage. *Academy of Management Review* 23(2):242–266.

Nielsen F (2011) A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *Proceedings of the ESWC2011 Workshop on "Making Sense of Microposts": Big things come in small packages*:93–98.

Nonaka I, von Krogh G (2009) Tacit knowledge and knowledge conversion: Controversy and advancement in organizational knowledge creation theory. *Organization Science* 20(3):635–652.

Oldenburg R (1989) *The great good place: Café, coffee shops, community centers, beauty parlors, general stores, bars, hangouts, and how they get you through the day.* (Paragon House Publishers).

O'Mahony S, Ferraro F (2007) The emergence of governance in an open source community. *Academy of Management Journal* 50(5):1079–1106.

Orlikowski WJ (2002) Knowing in practice: Enacting a collective capability in distributed organizing. *Organization science* 13(3):249–273.

Phillips DJ, Zuckerman EW (2001) Middle-Status Conformity: Theoretical Restatement and Empirical Demonstration in Two Markets. *American Journal of Sociology* 107(2):379–429.

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

Polanyi M (1958) *Personal knowledge: Towards a post-critical philosophy* (University of Chicago Press).

Preece J (2000) *Online communities: Designing usability and supporting socialbilty* (John Wiley & Sons, Inc.).

Qureshi I, Fang Y (2011) Socialization in open source software projects: A growth mixture modeling approach. *Organizational Research Methods* 14(1):208–238.

Rabe-Hesketh S, Skrondal A (2008) *Multilevel and longitudinal modeling using Stata* (STATA press).

Rheingold H (1993) *The virtual community: Finding connection in a computerized world* (Addison-Wesley Longman Publishing Co., Inc.).

Rombach MP, Porter MA, Fowler JH, Mucha PJ (2014) Core-periphery structure in networks. *SIAM Journal on Applied mathematics* 74(1):167–190.

Rosen-Zvi M, Chemudugunta C, Griffiths T, Smyth P, Steyvers M (2010) Learning author-topic models from text corpora. *ACM Transactions on Information Systems (TOIS)* 28(1):4.

Rullani F, Haefliger S (2013) The periphery on stage: The intra-organizational dynamics in online communities of creation. *Research Policy* 42(4):941–953.

Seidman SB (1983) Network structure and minimum degree. *Social networks* 5(3):269–287.

Sgourev SV (2013) How Paris gave rise to Cubism (and Picasso): Ambiguity and fragmentation in radical innovation. *Organization Science* 24(6):1601–1617.

Small TA (2011) What the hashtag?: A content analysis of Canadian politics on Twitter. *Information Communication and Society* 14(6):872–895.

Sproull L, Kiesler S (1991) Computers, Networks and Work. *Scientific American* 265(3):116–127.

Sutanto J, Tan CH, Battistini B, Phang CW (2011) Emergent Leadership in Virtual Collaboration Settings: A Social Network Analysis Approach. *Long Range Planning* 44(5–6):421–439.

Szulanski G (1996) Exploring internal stickiness: Impediments to the transfer of best practice within the firm. *Strategic management journal* 17(S2):27–43.

Tsoukas H (2009) A dialogical approach to the creation of new knowledge in organizations. *Organization Science* 20(6):941–957.

Wasko M, Faraj S (2005) Why Should I Share: Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice. *MIS Quarterly* 29(1):35–47.

West BT, Welch KB, Galecki AT (2014) *Linear mixed models: a practical guide using statistical software* (Chapman and Hall/CRC).

Wooldridge JM (2010) *Econometric analysis of cross section and panel data* (MIT press).

Zaheer S, Albert S, Zaheer A (1999) Time scales and organizational theory. *Academy of Management Review* 24(4):725–741.

Figures and Tables

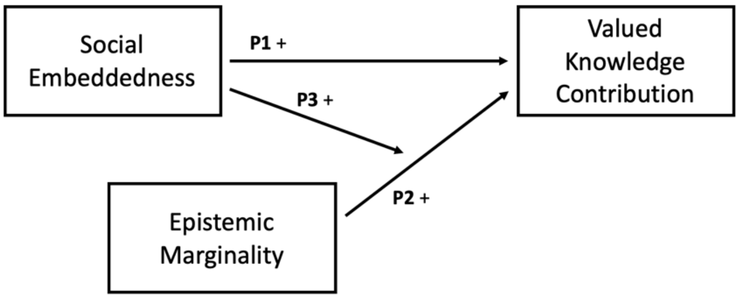
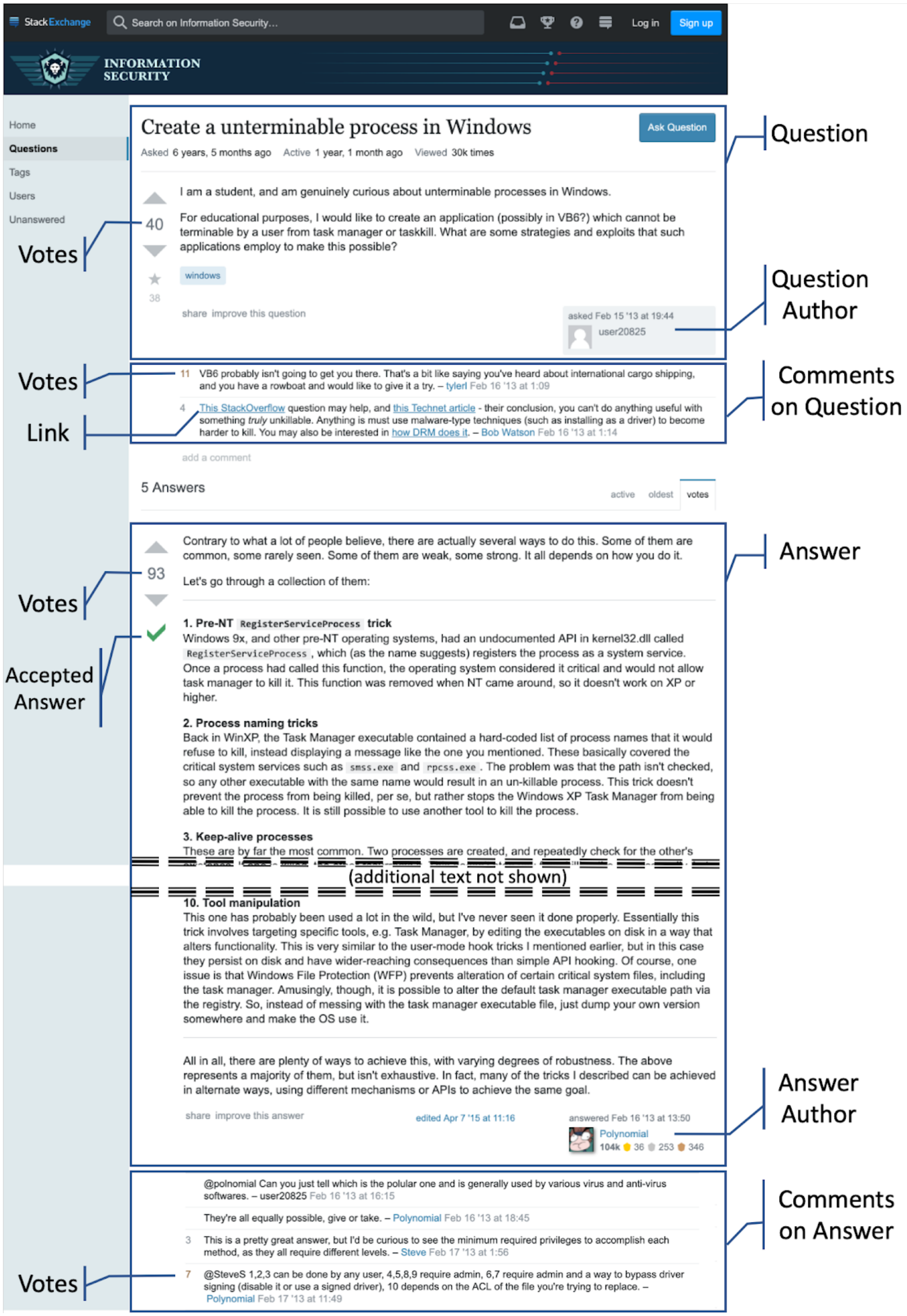


Figure 1: Research model



Create a unterminable process in Windows

Asked 6 years, 5 months ago Active 1 year, 1 month ago Viewed 30k times

40

I am a student, and am genuinely curious about unterminable processes in Windows.

For educational purposes, I would like to create an application (possibly in VB6?) which cannot be terminable by a user from task manager or taskkill. What are some strategies and exploits that such applications employ to make this possible?

windows

38

share improve this question

asked Feb 15 '13 at 19:44
user20825

11 VB6 probably isn't going to get you there. That's a bit like saying you've heard about international cargo shipping, and you have a rowboat and would like to give it a try. – tylerl Feb 16 '13 at 1:09

4 This StackOverflow question may help, and this Technet article - their conclusion, you can't do anything useful with something *truly* unkillable. Anything is must use malware-type techniques (such as installing as a driver) to become harder to kill. You may also be interested in how DRM does it. – Bob Watson Feb 16 '13 at 1:14

add a comment

5 Answers

active oldest votes

93

Contrary to what a lot of people believe, there are actually several ways to do this. Some of them are common, some rarely seen. Some of them are weak, some strong. It all depends on how you do it.

Let's go through a collection of them:

1. Pre-NT `RegisterServiceProcess` trick
Windows 9x, and other pre-NT operating systems, had an undocumented API in kernel32.dll called `RegisterServiceProcess`, which (as the name suggests) registers the process as a system service. Once a process had called this function, the operating system considered it critical and would not allow task manager to kill it. This function was removed when NT came around, so it doesn't work on XP or higher.

2. Process naming tricks
Back in WinXP, the Task Manager executable contained a hard-coded list of process names that it would refuse to kill, instead displaying a message like the one you mentioned. These basically covered the critical system services such as `smss.exe` and `rpcss.exe`. The problem was that the path isn't checked, so any other executable with the same name would result in an un-killable process. This trick doesn't prevent the process from being killed, per se, but rather stops the Windows XP Task Manager from being able to kill the process. It is still possible to use another tool to kill the process.

3. Keep-alive processes
These are by far the most common. Two processes are created, and repeatedly check for the other's

(additional text not shown)

10. Tool manipulation
This one has probably been used a lot in the wild, but I've never seen it done properly. Essentially this trick involves targeting specific tools, e.g. Task Manager, by editing the executables on disk in a way that alters functionality. This is very similar to the user-mode hook tricks I mentioned earlier, but in this case they persist on disk and have wider-reaching consequences than simple API hooking. Of course, one issue is that Windows File Protection (WFP) prevents alteration of certain critical system files, including the task manager. Amusingly, though, it is possible to alter the default task manager executable path via the registry. So, instead of messing with the task manager executable file, just dump your own version somewhere and make the OS use it.

All in all, there are plenty of ways to achieve this, with varying degrees of robustness. The above represents a majority of them, but isn't exhaustive. In fact, many of the tricks I described can be achieved in alternate ways, using different mechanisms or APIs to achieve the same goal.

share improve this answer

edited Apr 7 '15 at 11:16

answered Feb 16 '13 at 13:50
Polynomial
104k ● 36 ● 253 ● 346

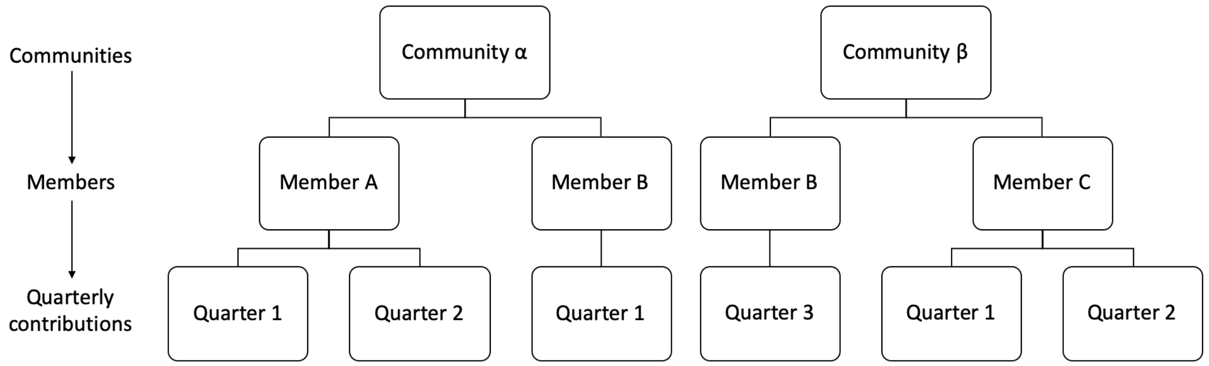
@polnomial Can you just tell which is the polular one and is generally used by various virus and anti-virus softwares. – user20825 Feb 16 '13 at 16:15

They're all equally possible, give or take. – Polynomial Feb 16 '13 at 18:45

3 This is a pretty great answer, but I'd be curious to see the minimum required privileges to accomplish each method, as they all require different levels. – Steve Feb 17 '13 at 1:56

7 @SteveS 1,2,3 can be done by any user, 4,5,8,9 require admin, 6,7 require admin and a way to bypass driver signing (disable it or use a signed driver), 10 depends on the ACL of the file you're trying to replace. – Polynomial Feb 17 '13 at 11:49

Figure 2: Annotated example of Stack Exchange contributions; as of August 5, 2019, from https://security.stackexchange.com/questions/30985/create-a-unterminable-process-in-windows



Note: Some members are affiliated with more than one community (e.g., Member B). Their contributions to each community are treated independently.

Figure 3: Nested structure of Stack Exchange member contribution data

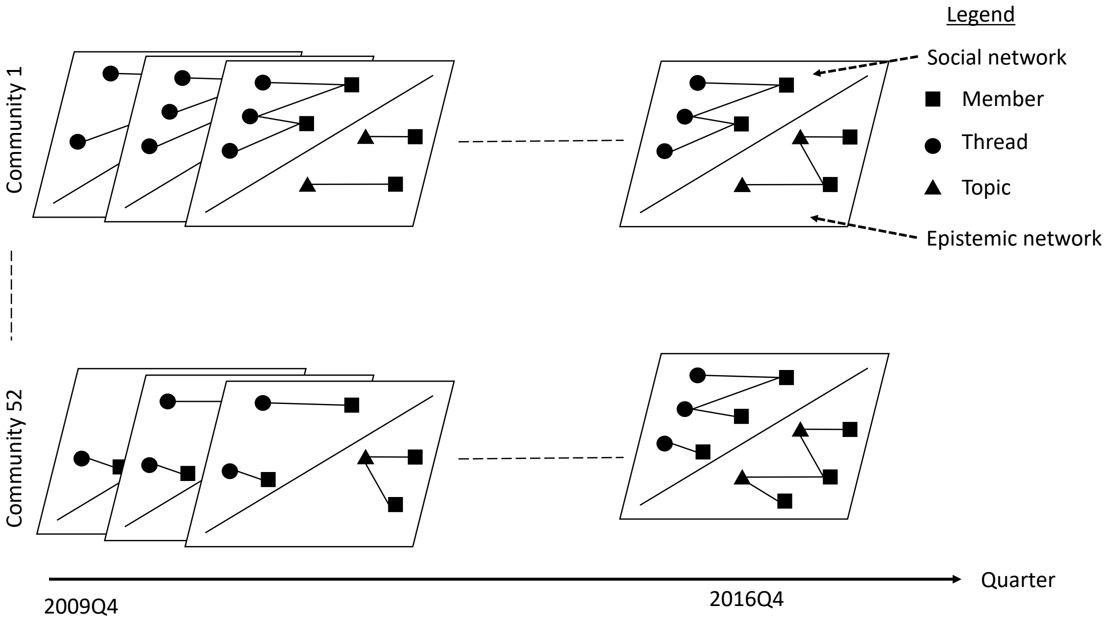


Figure 4: Construction of the networks from the archival data

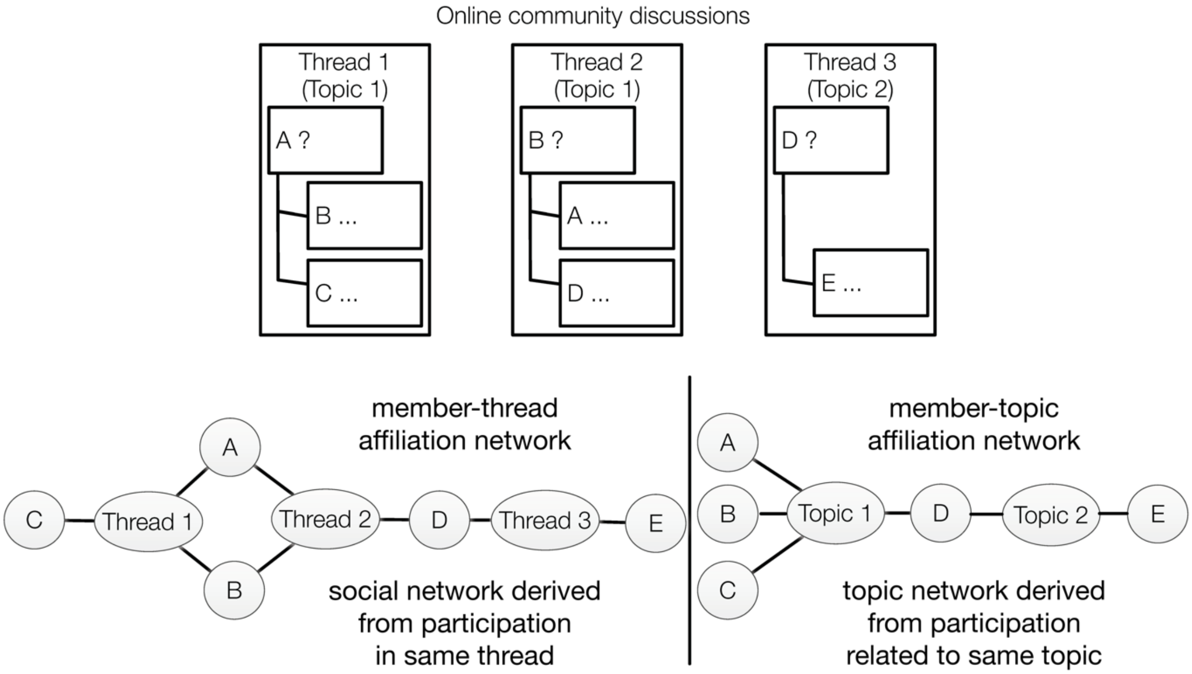
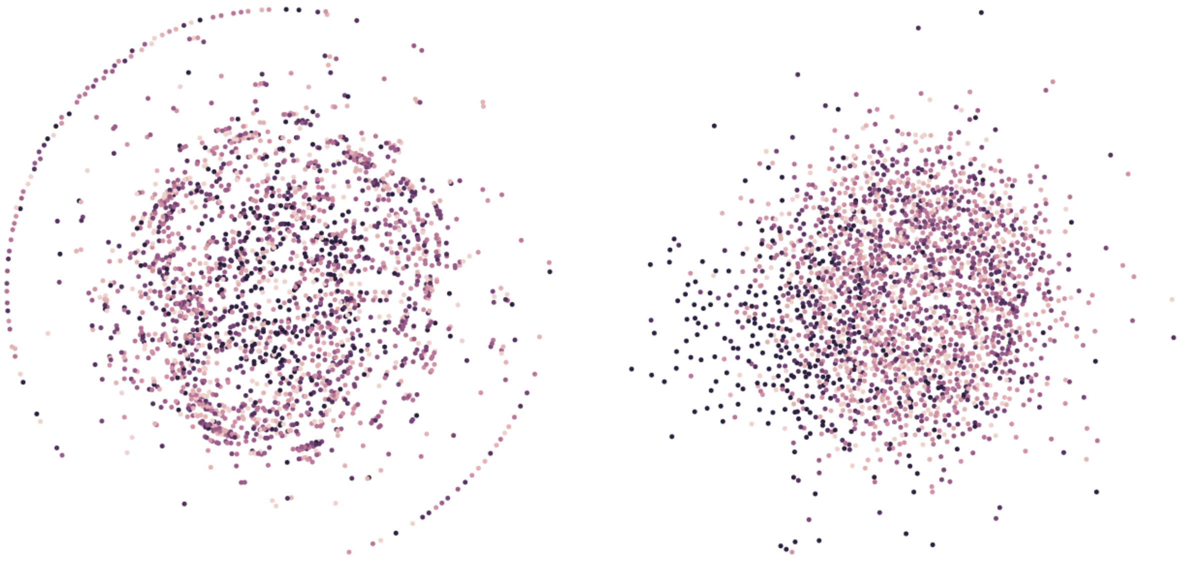


Figure 5: Creation of social and topic affiliation networks based on online community discussions

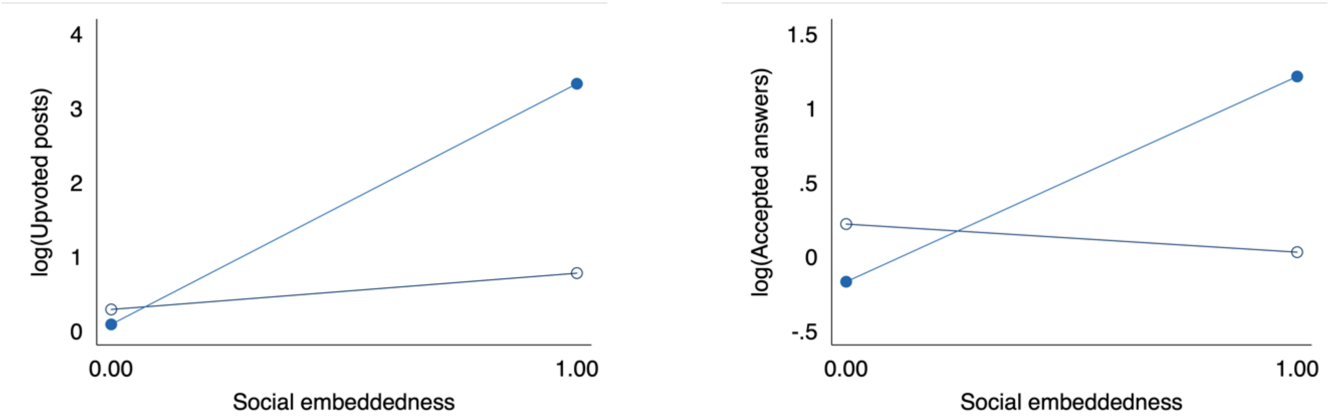


(a) Social position based on the centrality of threads that a member posted to

(b) Epistemic position based on the centrality of topics in a member's posts

The darker dots represent members with a higher number of valued knowledge contributions.
In the social network the preponderance of darkest dots is towards the center; in the topic network, towards the edges.

Figure 6: Member social and epistemic position for Stack Exchange GIS community (3rd quarter of 2013)



Epistemic marginality values at ○ a minimum of zero, and ● a maximum of one
Figure 7: interaction plots for upvoted posts (left) and accepted answers (right)

Table 1: Theoretical constructs and their operationalizations

Construct	**Description**
Dependent variable: valued knowledge contribution	
(1) Upvoted posts	Number of upvoted posts
(2) Accepted answers	Number of accepted answers
Community-level controls	
(3) Community age	The number of quarters since the inception of the community
(4) Community membership	The number of members in the community
(5) Community crowding	The number of questions divided over the number of members
(6) Community maturity	The average age of community members
Member-level controls	
(7) Concurrent communities	The number of Stack Exchange communities that a member is affiliated with
(8) Pseudonymity	The incompleteness of the member profile based on the following information: real name, age, location, picture, and web address
(9) Comment rank	The median order of the member's comments within the discussion threads
(10) Answer rank	The median order of the member's answers within the discussion threads
(11) Total contributions	The sum of the number of questions, answers, and comments
(12) Tenure	The number of years since a member's first post in the stack
(13) Links per post	The average number of links per member post
(14) Mentions per post	The average number of mentions of other members per post
(15) Readability of posts	The average readability per post using the reverse-coded SMOG index
(16) Language prototypicality	The inverse cross-entropy of posting using a statistical trigram language model
(17) Positive sentiment	The average positive sentiment per post using AFINN classifier
Independent variables: member position	
(18) Social embeddedness	Closeness centrality within the member-thread affiliation network
(19) Epistemic marginality	Closeness centrality within the member-topic affiliation network (reversed)

Table 2: Descriptive statistics and correlation matrix

	Mean	**Standard deviation**	**Minimum**	**First quartile**	**Median**	**Third quartile**	**Maximum**
(1) Upvoted posts	6.156	34.63	0	0	1	3	4313
(2) Accepted answers	.457	3.661	0	0	0	0	459
(3) Community age	14.50	6.661	0	9	15	20	28
(4) Community membership	3418.9	3198.8	17	1349	2486	3979	14340
(5) Community crowding	1.012	.365	.352	.765	.934	1.193	8.927
(6) Community maturity	33.24	2.174	24.72	31.75	33.10	34.55	45.58
(7) Concurrent communities	2.467	3.683	1	1	1	2	38
(8) Pseudonymity	3.208	1.377	0	2	3	4	5
(9) Comment rank	7.804	8.962	1	2	3	13	58
(10) Answer rank	8.088	7.718	1	2	7	11	89
(11) Total contributions	7.148	32.26	1	1	2	5	2587
(12) Tenure	.867	1.170	.000000467	.127	.244	1.219	7.244
(13) Links per post	.214	.536	0	0	0	.200	84
(14) Mentions per post	.162	1.489	0	0	0	.0588	1036
(15) Readability of posts	-8.844	5.000	-133.4	-11.66	-9.831	-7.399	0
(16) Language prototypicality	11.69	5.936	0	5.781	13.72	17.26	20.36
(17) Positive sentiment	.200	.390	-18.80	0	.174	.393	10.14
(18) Social embeddedness	.503	.175	.000100	.427	.506	.606	1
(19) Epistemic marginality	.272	.234	0	.128	.198	.313	1

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**	**(15)**	**(16)**	**(17)**	**(18)**	**(19)**
(1)	1.00																		
(2)	.78*	1.00																	
(3)	-.05*	-.03*	1.00																
(4)	-.05*	-.02*	.29*	1.00															
(5)	.02*	.05*	-.16*	-.10*	1.00														
(6)	.06*	.01*	-.40*	-.37*	-.06*	1.00													
(7)	.01*	-.00*	.00	-.06*	-.08*	.05*	1.00												
(8)	-.07*	-.07*	-.01*	.04*	.06*	-.01*	.06*	1.00											
(9)	-.07*	-.07*	.07*	.14*	-.12*	-.04*	-.04*	.15*	1.00										
(10)	-.09*	-.11*	-.02*	.05*	-.15*	.06*	.03*	.03*	-.03*	1.00									
(11)	.74*	.83*	-.02*	-.02*	.06*	.02*	.01*	-.08*	-.11*	-.11*	1.00								
(12)	.07*	.07*	.26*	.04*	-.07*	-.07*	.07*	-.19*	-.16*	-.04*	.09*	1.00							
(13)	.03*	.03*	-.01*	-.01*	.01*	.00*	-.02*	-.04*	.13*	-.14*	.00*	.02*	1.00						
(14)	.01*	.01*	.02*	.01*	.01*	-.00*	.01*	-.01*	-.02*	.01*	.01*	.02*	.00*	1.00					
(15)	.15*	.15*	-.01*	.03*	.04*	.01*	.01*	-.07*	-.27*	-.06*	.21*	.07*	-.08*	-.04*	1.00				
(16)	-.02*	-.01*	.27*	.49*	-.21*	-.21*	-.02*	-.01*	.10*	.11*	-.01*	.07*	-.01*	.01*	-.02*	1.00			
(17)	-.00*	-.01*	-.05*	-.08*	.03*	.05*	-.02*	-.01*	-.06*	.02*	-.01*	-.04*	.04*	-.01*	-.00*	-.05*	1.00		
(18)	.22*	.18*	-.05*	-.09*	.01*	.04*	.05*	-.07*	-.31*	-.02*	.28*	.05*	-.06*	.02*	.25*	-.03*	.02*	1.00	
(19)	.31*	.26*	-.12*	-.19*	.03*	.08*	.04*	-.12*	-.26*	-.20*	.35*	.10*	.03*	.01*	.24*	-.10*	.02*	.44*	1.00

1,906,434 quarterly contributions of 697,412 members in 52 communities over 29 quarters; * p<0.05

Table 3: Multi-level regression results

		(1)	(2)	(3)		(4)	(5)	(6)
			log(Upvoted posts)				log(Accepted answers)	
Community-level								
Community age	β_1	-.0178***	-.0136***	-.0135***	γ_1	-.0055***	-.0044***	-.0043***
		(.0031)	(.0028)	(.0027)		(.0006)	(.0006)	(.0006)
Community membership	β_2	-.0000	-.0000	-.0000	γ_2	-.0000	-.0000	-.0000
		(.0000)	(.0000)	(.0000)		(.0000)	(.0000)	(.0000)
Community crowding	β_3	.4613***	.4551***	.4554***	γ_3	.0951***	.0914***	.0951***
		(.1275)	(.1261)	(.1241)		(.0206)	(.0236)	(.0234)
Community maturity	β_4	.0261*	.0176	.0158	γ_4	-.0018	-.0056*	-.0068*
		(.0122)	(.0117)	(.0115)		(.0023)	(.0028)	(.0027)
Member-level								
Concurrent communities	β_5	.0006	-.0033**	-.0038***	γ_5	-.0001	-.0019***	-.0024***
		(.0010)	(.0011)	(.0011)		(.0004)	(.0004)	(.0004)
Pseudonymity	β_6	-.0208***	-.0110**	-.0099**	γ_6	-.0287***	-.0188***	-.0192***
		(.0046)	(.0034)	(.0033)		(.0022)	(.0015)	(.0015)
Comment rank	β_7	-.0214***	-.0074***	-.0081***	γ_7	-.0065***	-.0004	-.0014***
		(.0027)	(.0016)	(.0015)		(.0007)	(.0003)	(.0004)
Answer rank	β_8	-.0323***	-.0226***	-.0221***	γ_8	-.0285***	-.0276***	-.0263***
		(.0060)	(.0044)	(.0043)		(.0062)	(.0060)	(.0056)
Total contributions	β_9	.0122***	.0079***	.0065***	γ_9	.0070***	.0059***	.0052***
		(.0010)	(.0007)	(.0006)		(.0005)	(.0004)	(.0004)
Tenure	β_{10}	-.0409***	-.0144**	-.0151**	γ_{10}	.0040*	.0060***	.0075***
		(.0055)	(.0050)	(.0049)		(.0019)	(.0018)	(.0019)
Links per post	β_{11}	.2067***	.1864***	.1924***	γ_{11}	.0332***	.0353***	.0375***
		(.0190)	(.0175)	(.0172)		(.0026)	(.0027)	(.0027)
Mentions per post	β_{12}	.0126***	.0040	.0032	γ_{12}	.0057***	.0025***	.0016*
		(.0026)	(.0023)	(.0024)		(.0010)	(.0007)	(.0008)
Readability of posts	β_{13}	.0196***	.0048***	.0027***	γ_{13}	.0130***	.0069***	.0054***
		(.0015)	(.0007)	(.0007)		(.0008)	(.0004)	(.0004)
Language prototypicality	β_{14}	.0055***	.0031*	.0032**	γ_{14}	.0006**	.0003	.0003
		(.0015)	(.0012)	(.0012)		(.0002)	(.0002)	(.0002)
Positive sentiment	β_{15}	.0764***	.0671***	.0710***	γ_{15}	-.0066*	-.0190***	-.0135**
		(.0099)	(.0093)	(.0097)		(.0033)	(.0045)	(.0045)
Social embeddedness	β_{16}		1.1183***	.4896***	γ_{16}		.2170***	-.1893***
			(.1254)	(.1218)			(.0267)	(.0257)
Epistemic marginality	β_{17}		1.4763***	-.2013*	γ_{17}		.5965***	-.3886***
			(.0872)	(.1000)			(.0479)	(.0507)
Social embeddedness	β_{18}			2.7511***	γ_{18}			1.5713***
× Epistemic marginality				(.2096)				(.0946)
Constant	β_0	.2887	-.7567	-.3548	γ_0	.5505***	.2240*	.4886***
		(.4192)	(.4031)	(.3924)		(.0797)	(.0972)	(.0962)
Random effects								
$\sigma^2(v_{0c})$		-.9546***	-1.0779***	-1.0828***		-2.9427***	-2.7485***	-2.8137***
		(.0910)	(.0910)	(.0911)		(.2285)	(.1492)	(.1386)
$\sigma^2(u_{0mc})$		-.9626***	-1.1672***	-1.1926***		-1.5753***	-1.6729***	-1.6865***
		(.0524)	(.0511)	(.0520)		(.0402)	(.0406)	(.0395)
Observations		1574560	1574560	1574560		734911	734911	734911
Log-likelihood		-1981073	-1794707	-1780722		-448782	-409029	-400724
χ^2		7485.69	12032.80	12804.65		5818.30	10719.94	9641.68

Robust standard errors in parentheses, * p<.05, ** p<.01, *** p<.001, Probability > χ^2 = 0 for all models

Online Appendix

A. Measurement Validation

To validate the use of closeness centrality as a measure of embeddedness (the inverse of marginality) within the topic network, we perform further analyses to assess the face validity and the external validity of the measure.

B. Epistemic marginality across online innovation communities

To provide more insight into the role of epistemic marginality in online innovation communities, we performed additional exploratory analyses. In the main analyses reported in the paper, we assessed epistemic marginality within each community. Yet, the same individual can be a central member of one community and a marginal member in another. Because Stack Exchange maintains a persistent identity of members who participate in more than one stack, we can gain insight into the role of boundary spanning by investigating these multi-stack members. We report the analyses below to understand better the roughly 35% of members who are concurrent members of multiple communities and, thus, are inter-community boundary spanners.

First, we consider the affiliation network formed by individual stack membership across all stacks in our sample. From this perspective, a participant in multiple stacks will be in a more core position within the overall Stack Exchange ecosystem than the more peripheral participants in single stacks. Analyzing this ecosystem-level network, we assigned each individual in our sample an ecosystem centrality score that reflects their community spanning. We find that the correlation between community spanning and epistemic marginality is negatively correlated ($r = -0.014$, $p < .001$). The modest negative correlation suggests that community spanners—those who are members in multiple communities—are less likely to be epistemically marginal than members of a single stack. Nonetheless, this evidence is provisional in that the two measurements (community spanning in the ecosystem and epistemic marginality in a stack) operate at different units of analysis.

Second, to strengthen the understanding of inter-community spanning, we created a weighted network that considers not just membership in multiple stacks, but also the level of participation in each. This formulation takes into consideration that if a member *A* is affiliated with two communities and has one contribution in one and nine in another, they are less of community spanner than a participant *B* who has five contributions in each. To measure weighted community spanning, we first calculated the standard deviation of members' frequency of contribution in each community they are affiliated with and then negating it. Members who score high concentrate their contributions in one focal community and participate less in the other communities they are a member of (less spanning). In contrast, a low standard deviation indicates that members contribute almost equally to all communities (more spanning). We then calculated the correlation between the weighted community spanning measure and epistemic marginality (r = -0.13, p < .001). We interpret this result as further evidence that inter-community spanners tend to be epistemically embedded (i.e., less marginal) within their communities as well.

Finally, we investigate the relationship between tenure in a community and epistemic marginality in that community (r = 0.10, p < 0.01). We find that members with a longer duration of membership in a community are more likely to be epistemically marginal. This result is consistent with gaining expertise in a knowledge field over time and, thus, being able to provide more novel contributions than others with less experience. In summary, our additional exploratory analysis provides evidence that epistemic marginality is associated with participating in a single stack over an extended period of time.

C. Closeness centrality versus continuous core-periphery

As noted in the body of the paper, we use closeness centrality to assess members' positions in the social and topic networks for multiple main reasons. We discuss two of these further here. First, given the large size and density of these networks, it is challenging to uncover structurally equivalent positions using classical core-periphery decomposition algorithms. Furthermore, core-periphery assumes discrete cores to which different members belong. On the other hand, we have no reason to expect such distinct cores to exist in a fluid online community. Closeness centrality relaxes this assumption, given that it considers a member position within the overall network. Indeed, closeness centrality has already been suggested to be a better measurement of position when the network does not exhibit a significant core-periphery structure. "From a theoretical point of view, the key difference between a centrality measure and a coreness measure is that coreness carries with it a model of the pattern of ties in the network as a whole. The coreness measure is only interpretable to the extent that the model fits. In contrast, a centrality measure is interpretable no matter what the structure of the network. For example, closeness centrality measures the entire graph-theoretic distance of a node to all others. A node's closeness centrality can be used to predict the time that messages originating at random nodes throughout the network will take to reach that node. The measure holds this interpretation no matter what the structure of the network" (Borgatti and Everett 1999, p. 393).

Notwithstanding this conceptual argument, there are recent methodological advancement in core-periphery measurement that relax the constraint of discrete cores. New *continuous* core-periphery measures extend discrete core-periphery measures to assess the coreness of nodes within the network without an a priori determination of discrete communities (Borgatti and Everett 1999, Rombach et al. 2014). At the same time, there has been advancement (Boyd et al. 2006, Kojaku and Masuda 2018) in testing for the statistical significance of a core-periphery structure.

We perform two validation checks. First, we assess whether the social and the topic network exhibit a significant discrete core-periphery structure that justifies using classical core-periphery rather than closeness centrality as a measure. Then to check the external validity of closeness centrality, we correlate it with the continuous core-periphery measure. We note here that the second test depends on the first. If the networks do not exhibit a significant core-periphery structure, then closeness centrality is not expected to correlate with any core-periphery measure that assumes such a structure in the network.

In these validations, we estimate continuous core-periphery measure and core-periphery structure significance on a subset of the 17 smallest communities in our data. We only tested the smallest communities because of computational limitations of calculating the continuous core-periphery measure (Rombach et al. 2014) and the significance of the core-periphery structure in the networks (Kojaku and Masuda 2018). In particular, the latter compares the core-periphery structure of the network with that of random Erdős–Rényi networks of the same size. As a result, the already computationally expensive calculation has to be carried on extra random networks (in our case we simulated 100 random networks for each social and topic network in the sample).

The statistical significance of the core-periphery structure is presented in Table B and Figure A. Because we have multiple networks per stack-quarter (348 social networks and 348 topic networks), we report the average significance as well as its standard error. The box plot in Figure A illustrates the variability of core-periphery structure significance by the stack category. The results demonstrate that the social networks exhibit a significant core-periphery structure with all 348 social networks having a core-periphery significance of $p < 0.0025$.

On the other hand, topic networks do not exhibit a core-periphery structure as evident by the nonsignificant results of the community structure test. The average statistical significant is $p = .72$, which is highly insignificant, suggesting that the core-periphery partition of these networks using the continuous core-periphery algorithm is not better than a random partitioning. A minority of these topic networks exhibit a core-periphery structure as evident by the whiskers and outlier points in the box plot.

Second, we correlate the continuous core-periphery measure of member nodes with the closeness centrality of member nodes in these networks. The correlation matrix is presented in Table C. In the social networks, the correlation between closeness centrality and continuous core-periphery is 0.41 ($p < 0.05$). This moderate value suggests that closeness centrality reflects the continuous core-periphery measure of the social networks. In the topic networks, the correlation between closeness centrality and continuous core-periphery is much lower (r=0.12, $p < 0.05$). This low correlation, compared to that in the social networks, suggests that closeness centrality captures an aspect of network structure that is not reflected in the continuous-core periphery measure. Given that the topic networks do not exhibit a core-periphery structure to start with, this comparison validates the use of closeness centrality as a measure of embeddedness.

As for the social network, both the closeness centrality and the continuous core-periphery are valid measures for coreness. However, in addition to the advantages descried in the body of the paper, we opted to use closeness centrality because although the statistical test was highly significant, we were only able to run it on a selected sample of the networks, and thus there is no guarantee that much larger social networks would exhibit a single core structure. Further, using the same measure for both networks also provides the advantage of making similar comparisons and interpretations of its effects.

D. Derived topics versus author-assigned tags

Because tags are assigned to questions, rather than to individual content items, to use tags as a proxy for topics would greatly limits the ability to test our research model. Nonetheless, to inform future research efforts, we report on the comparison of tags and topics. We compared the number of tags a member used to topics extracted from their contributions using the LDA algorithm. We find that tags have a greater variability compared to topics. The distribution of tags is overdispersed (Mean= 2.46, SD= 4.95), whereas the distribution of topics is not (Mean= 44.11, SD= 25.99). This evidence suggests that members vary greatly in their tagging behavior. We also observe this relationship between-group which suggests that communities have different tagging norms. The between-group standard deviation of tags (1.05) is higher relative to the mean compared to the between-group standard deviation of topics (4.80). Finally, 37% of members do not use tags at all compared to only 5% of members who have no topics associated with their contributions (mostly because of concise content).

E. Community-Level Descriptive Statistics

Table D presents aggregate statistics of the 52 communities, or stacks, that are studied in this paper since inception until the end of 2016. Each community can be visited via the URL [community name].stackexchange.com (e.g., physics.stackexchange.com). Consistent with their categorization by the Stack Exchange platform, the communities are grouped into five categories (Culture / Recreation, Life / Arts, Professional, Science, and Technology). For each community the table reports the age of the community in quarters, the total number of members, questions, answers, comments, accepted answers and upvoted posts. Figure B presents another view of this information. The sparkline plots show how these total numbers across all communities change over time. Active members are members contributing in each quarter. Finally, Table A presents several example questions from the communities.

F. Model Diagnostics

In order to ensure a ceteris paribus interpretation of the regression coefficient, we perform collinearity diagnostics using variance inflation factors (VIF). The average VIF is 1.22 and all independent variables have a VIF less than 1.5. This value is less than the recommended value of 10 and the most conservative recommendation of 5 suggesting the absence of collinearity between independent variables in our data.

In interpreting the interaction effects, it is important to ensure that the two variables, social embeddedness and epistemic marginality, have observations in all possible configurations of high (above mean) values and low (below mean) values of both variables. Table E presents the counts of observation in each configuration of high/low social embeddedness/epistemic marginality. The observations are fairly distributed across the four configurations. Most observations are in the two categories of high/high and low/low, reflecting therefore on the positive correlation (0.44) between social embeddedness and epistemic marginality. However, there is still a substantial number of observations in the other two categories ensuring that there are cases of high/low and low/high social embeddedness and epistemic marginality that the interaction effects interpretation applies to.

Finally, we test for potential model specification errors using Stata's linktest function (Pregibon 1980, Stata 2015). Linktest deals with two potential causes of miss-specification: omitted variables and an incorrect link function. Linktest uses the predicted value (*hat*) and its value squared (*hatsq*) as predictors to rebuild the model. If the model is specified correctly, the predicted value should be significant but its squared value should not as it means that either there omitted relevant variables or that the link function is not correctly specified (Chen et al. 2003). It is worth noting that linktest is not offered for multi-level data, so we estimated standard regression models controlling for community membership. The coefficients are comparable to those of the main model. The value of *hat* is significant economically and statistically for both dependent variables (1.002 and 1.037 respectively). The value of *hatsq* is economically non-significant for both upvoted posts and accepted answers (-6.50e-06 and -.0006 respectively). It is statistically non-significant for upvoted posts. The results confirm an adequate model specification.

G. Sensitivity Analyses

We performed extensive sensitivity analyses to ensure the validity of the results under different model specifications and sampling approaches.

First, because some members (35% of members) are affiliated with multiple communities, results may be affected by spill-overs across communities. For example, what a member learns or contributes to one community can affect that member's contribution to another community. This effect is partially mitigated by controlling for the number of concurrent communities that a member is affiliated with and also by the multi-level model specification. Furthermore, we offer two extra robustness checks to ensure that the results are not affected by spill-overs across communities.

First, we exclude from the analysis members who have affiliations in multiple communities (Table F). In this analysis, the "Concurrent communities" variable is automatically dropped because it does not vary. This constraint excludes observations of multi-community-affiliated members. However, the results are consistent with those of the main analyses. Second, we run another analysis in which we only examine observations about communities in which multi-affiliated members are most active (Table G). Once again the results are consistent with those of the main analysis.

Second, the main analysis includes both knowledge seekers (i.e., members who post questions) and knowledge providers (i.e., members who provide answers and comments) in the same sample. However, because our theoretical interest is about knowledge contribution, one may argue that knowledge seekers should not be included in the model. A counter-argument suggests that knowledge seekers are still important participants in the discourse because they identify the knowledge gaps that need to be addressed in the community of knowing. In other words, no answers are possible without valid questions. However, as a robustness check, we only consider members who provide answers and comments but no questions (Table H). The results are overall consistent with those of the main analysis.

Third, another concern about the findings is that their statistical significance is purely driven by the large sample size rather than real relationships between variables. To address this issue, we randomly select samples of smaller size and repeat the analyses comparing the coefficients. Specifically, we randomly select 10%, 5%, and 1% of observations using a stratified sampling strategy where the strata are the stack-quarter combinations. This stratification is important to ensure we have a representative sub-sample that includes all stacks in all quarters and hence is comparable to the sample used in the main analysis. The results for main effects are presented in Table I, and the results of the interaction effects are presented in Table J. The results are consistent with those of the main analysis including the results of the 1% sample.

Fourth, to ensure that the results are not driven by a few influential members in the communities, we exclude the top 1% of members from the sample based on their total number of contributions. We note here that prior work in online communities recognizes the highly unequal nature of contribution where a few members generate most contribution. In our data, the top 1% of members contribute 15% to 45% of posts in their communities. Nonetheless, the exclusion of these members does not alter the conclusion of the main analysis as outlined in Table K.

H. Alternative Specifications

The two outcome variables (upvoted posts and accepted answers) represent the recognition and validation of the contribution as valuable by the knowledge seeker and the community, respectively. It is important to rule out the alternative explanation that this recognition is resulting from the formal role of the contributing member rather than the value of the contributed knowledge. Although formal roles are often reflective of prior contributions (O'Mahony and Ferraro 2007), contributions of the member with formal roles may receive more recognition because of their authority (Levina and Arriaga 2014). The extensive control variables including member tenure and pseudonymity partially address this concern. However, as a robustness check, we estimate an alternative model that includes the moderator role as a control variable. In Stack Exchange members can nominate and vote other members as moderators. Moderators are endowed with some privileges to intervene in selecting and promoting others' posting.[5] The moderator

[5] https://stackoverflow.blog/2009/05/18/a-theory-of-moderation/

data was only available for the last quarter of observation.[6] The results are in Table L. Moderators tend to gain more upvotes, but there is no significant effect on the acceptance of their answers. However and importantly the inclusion of the moderator roles does not alter the findings of the main analysis.

Finally, we address the concern that self-selection is driving the results. This concern stems from the tendency of some members to both occupy a unique position in the community and gain recognition (upvotes and accepted answers) because of an unobserved variable such as sociality. This concern is partially addressed by the inclusion of various member-level control variables. As a robustness check, we ran a Heckman selection model (Table M). We estimate a population pooled model and we include community dummies. We cluster the standard errors around members to avoid biasing the standard errors. The selection equations introduce two new variables "posting" and "answering" defined as having posts and answers greater than zero respectively. These variables are explained with member demographics: tenure and pseudonymity. The theoretical rationale for using these variables is that members who are long-tenured have more tendency to post and answer (explaining self-selection). On the other hand, members who use ambiguous identities (pseudonymity) have a lower tendency to post and answer (they probably would want to free-ride and get questions answered). The non-selection hazard (inverse mills ratio) is computed from the two selection equations and included in the main equation. Interestingly, the negative and significant coefficients of the inverse mills ratio suggest a negative selection bias. This is plausible given that upvotes and accepted answers depend on the quality of the contribution which can degrade if a member is aiming for quantity. However, correcting for self-selection results in comparable estimates to those of the main analysis.

[6] https://stackoverflow.com/election

I. References

Borgatti SP, Everett MG (1999) Models of core/periphery structures. *Soc. Networks* 21(4):375–395.

Boyd JP, Fitzgerald WJ, Beck RJ (2006) Computing core/periphery structures and permutation tests for social relations data. *Soc. Networks* 28(2):165–178.

Chen X, Ender P, Mitchell M, Wells C (2003) Regression diagnostics. *Regres. with Stata*. (UCLA).

Kojaku S, Masuda N (2018) A generalised significance test for individual communities in networks. *Sci. Rep.* 8(1):1–10.

Levina N, Arriaga M (2014) Distinction and Status Production on User-Generated Content Platforms: Using Bourdieu's Theory of Cultural Production to Understand Social Dynamics in Online Fields. *Inf. Syst. Res.* 25(3):468–488.

O'Mahony S, Ferraro F (2007) The Emergence of Governance in an Open Source Community. *Acad. Manag. J.* 50(5):1079–1106.

Pregibon D (1980) Goodness of link tests for generalized linear models. *Appl. Stat.*:14–15.

Rombach MP, Porter MA, Fowler JH, Mucha PJ (2014) Core-Periphery Structure in Networks. *SIAM J. Appl. Math.*

Stata (2015) *Stata Reference Manual Release 14* (Stata Press).

Tables and Figures

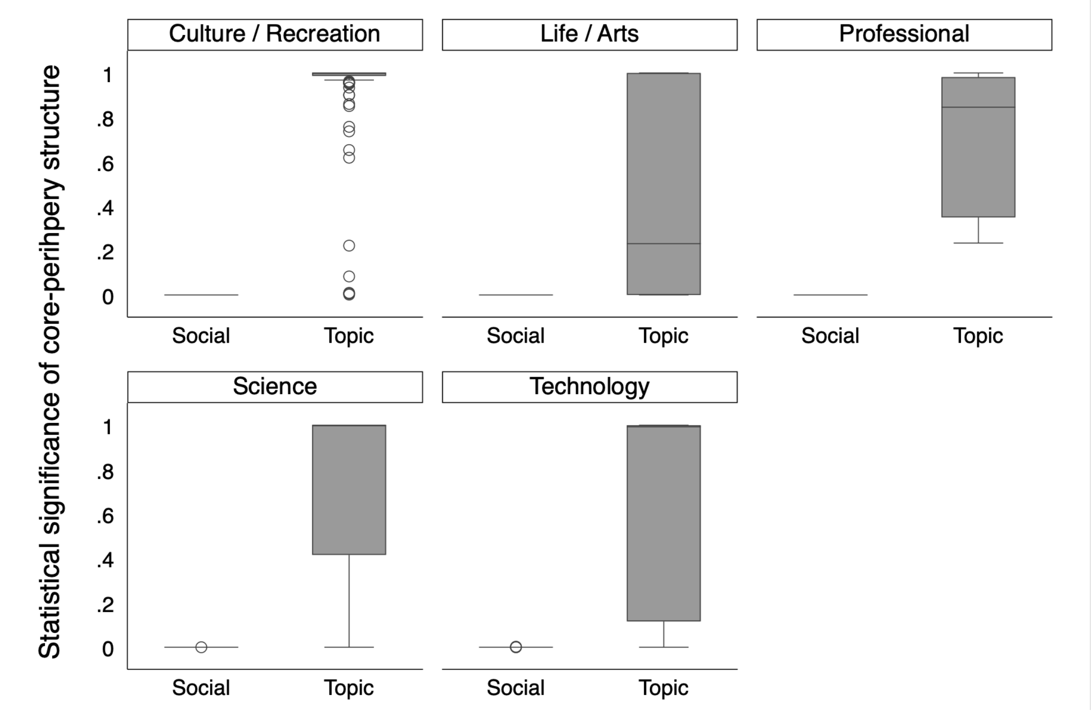


Figure A: Statistical significance of the core-periphery structure of the social and topic networks of communities grouped by their category

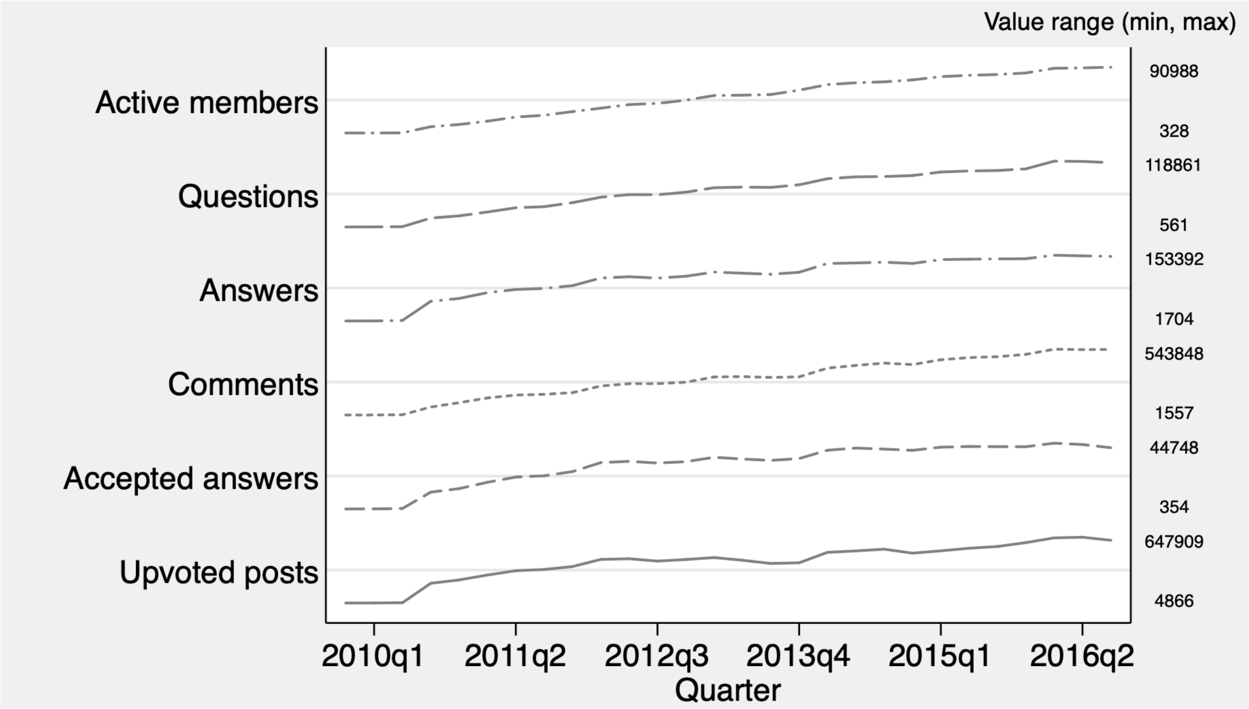


Figure B: Quarterly total number of active members, questions, answers, comments, accepted answers, and upvoted posts in all communities

Table A: Hot questions (snapshot on Monday, Sept. 16, 2019, 3PM)

Stack	Question Title	Asked	Answers
academia	Exam design: give maximum score per question or not?	yesterday	12
askubuntu	Delete empty subfolders, keep parent folder	yesterday	3
aviation	Why don't airports use arresting gears to recover energy from landing passenger planes?	3 days ago	9
codereview	Python web-scraper to download table of transistor counts from Wikipedia	3 days ago	6
cooking	Do household ovens ventilate heat to the outdoors?	yesterday	6
dba	MySQL - How to check for a value in all columns	21 hours ago	5
drupal	How can I add a link on the “Structure” admin page?	10 hours ago	2
electronics	Why do we need to use transistors when building an OR gate?	14 hours ago	5
gaming	How far away from you does grass spread?	11 hours ago	1
graphic design	What exactly is a web font, and what does converting to one involve?	17 hours ago	2
money	Amortized Loans seem to benefit the bank more than the customer	3 days ago	11
movies	Why are there no programmes / playbills for movies?	yesterday	2
music	Is the name of an interval between two notes unique and absolute?	20 hours ago	5
network engineering	Why do IXPs need ASN?	14 hours ago	1
photo	How important is weather sealing for a winter trip?	9 hours ago	1
physics	How is underwater propagation of sound possible?	20 hours ago	3
rpg	Persuading players to be less attached to a pre-session 0 character concept	yesterday	3
scifi	What is the origin of the “being immortal sucks” trope?	yesterday	5
security	How to generate short fixed length cryptographic hashes?	yesterday	2
stats	What's the purpose of autocorrelation?	12 hours ago	2
tex	Plot irregular circle in latex	yesterday	2
travel	Wrong Schengen Visa exit stamp on my passport, who can I complain to?	2 days ago	4
unix	Tips for remembering the order of parameters for ln?	yesterday	9
ux	Should the pagination be reset when changing the order?	9 hours ago	2
workplace	Should I inform my future product owner that there are big chances that a team member will leave the company soon?	yesterday	5

Table B: The mean statistical significance of core-periphery structure in the social and topic network of 17 communities

	Mean of core-periphery statistical significance	Std. Err.	[95% Conf.	Interval]
Social network	7.78e-06	6.75e-06	-5.48e-06	.000021
Topic network	.7211123	.0221631	.6775976	.7646269

Table C: Correlations between closeness centrality and continuous core-periphery measures of member nodes in the social and the topic networks

	(1)	(2)	(3)	(4)
(1) Social network closeness centrality	1.00			
(2) Social network continuous core-periphery	.41*	1.00		
(3) Topic network closeness centrality	-.31*	-.46*	1.00	
(4) Topic network continuous core-periphery	-.09*	-.07*	.12*	1.00

* p<0.0001, n=1,906,434

Table D: Overall characteristics of the 52 Stack Exchange communities

Category	**Community**	**Age**	**Members**	**Questions**	**Answers**	**Comments**	**Accepted answers**	**Upvoted posts***
Culture / Recreation	anime	16	4051	7074	9708	24309	3643	50418
	bicycles	25	7923	9001	22613	67309	4581	82587
	christianity	21	5315	8050	19482	64400	3734	89749
	gaming	25	43713	70308	112484	257850	41118	551926
	judaism	28	4095	20005	32801	152224	6566	144205
	rpg	25	9066	18878	47140	100160	12996	380139
	skeptics	23	5736	6900	8484	68964	2949	130778
	travel	22	16130	19947	34300	127566	8758	279324
Life / Arts	academia	19	14432	16221	39462	135564	7273	403797
	cooking	25	14352	15681	37214	80004	8158	146197
	diy	25	19912	26266	45312	115652	9928	130291
	graphicdesign	23	18234	18830	32911	77490	8333	91235
	money	28	12398	15578	30540	81572	7307	147906
	movies	20	10402	13035	19588	52155	7245	134305
	music	22	8281	8619	22458	48658	4480	87492
	photo	25	13113	17215	42222	103863	8893	166935
	scifi	23	24982	34566	67509	274306	19313	643035
Professional	aviation	12	4764	7493	14891	55841	4505	137814
	workplace	18	14466	12621	41411	132087	6264	407038
Science	biology	20	7547	13424	16235	55004	5864	82818
	chemistry	18	9725	17763	20989	66015	7248	91195
	cs	19	11328	16325	21391	77808	7837	101853
	cstheory	25	4757	7909	12292	51691	3641	112949
	physics	24	37847	83158	125202	439629	34499	444481
	stats	25	43204	83117	86884	351278	27101	307966
Technology	android	25	33686	36495	44727	118992	10513	93228
	apple	25	56430	71475	104097	230113	26133	204104
	askubuntu	25	162696	226293	299498	800197	76854	575097
	blender	14	10718	23302	24942	85658	10346	94778
	codereview	23	29354	37336	64330	174296	22064	326422
	crypto	21	7408	10836	14534	55923	5494	60520
	dba	23	30318	47566	64503	184410	22875	189406
	drupal	23	18554	63677	81014	183471	28230	115137
	dsp	21	6202	10240	13369	45130	4481	35271
	electronics	28	36310	70734	127191	444752	36462	374775
	expressionengine	16	3025	11137	14416	30356	5283	27367
	gamedev	25	22335	34129	55041	155784	18357	190906
	gis	25	31856	72105	88088	254207	27816	255568

Category	Community	Age	Members	Questions	Answers	Comments	Accepted answers	Upvoted posts*
	magento	15	15605	46517	55655	146003	18002	72849
	mathematica	19	11919	36275	54084	229585	19358	343747
	networkengineering	14	6322	7950	11810	32330	3541	35766
	programmers	25	40482	41103	127126	363729	23651	745736
	raspberrypi	18	13620	15035	19813	62030	5305	43046
	salesforce	17	13427	50742	62330	170864	22408	159271
	security	24	28835	34123	66351	181899	16415	367417
	sharepoint	28	19965	63044	81310	167255	24607	106891
	tex	25	45665	122195	158017	654211	70994	965709
	unix	25	56413	98484	150213	422057	48535	481800
	ux	25	19764	21582	58452	124411	10854	254898
	webapps	26	16577	17332	25489	40193	6527	67036
	webmasters	25	17698	23203	38929	78297	10756	95762
	wordpress	25	31569	68505	87412	243539	34020	106429

* posts are answers and comments and a post may receive multiple upvotes from different members

Table E: Covariation of social embeddedness and epistemic marginality

	Social embeddedness below mean	Social embeddedness above mean	Total
Epistemic marginality above mean	143,206	791,066	934,272
Epistemic marginality below mean	501,042	471,120	972,162
Total	644,248	1,262,186	1,906,434

Table F: Results excluding members who have affiliation with multiple communities (concurrent communities > 1)

	(1)	(2)	(3)	(4)	(5)	(6)
	log(Upvoted posts)			log(Accepted answers)		
Community-level						
Community age	-.0172***	-.0133***	-.0132***	-.0055***	-.0044***	-.0043***
	(.0028)	(.0027)	(.0026)	(.0007)	(.0007)	(.0006)
Community membership	-.0000	-.0000	-.0000	.0000	-.0000	-.0000
	(.0000)	(.0000)	(.0000)	(.0000)	(.0000)	(.0000)
Community crowding	.4139**	.4047**	.4031**	.0854***	.0782**	.0804***
	(.1282)	(.1322)	(.1293)	(.0218)	(.0247)	(.0234)
Community maturity	.0279*	.0204	.0183	.0005	-.0035	-.0051
	(.0120)	(.0116)	(.0113)	(.0025)	(.0030)	(.0029)
Member-level						
Concurrent communities	dropped	dropped	dropped	dropped	dropped	dropped
	(.)	(.)	(.)	(.)	(.)	(.)
Pseudonymity	-.0265***	-.0176***	-.0163***	-.0309***	-.0209***	-.0210***
	(.0046)	(.0031)	(.0030)	(.0026)	(.0017)	(.0016)
Comment rank	-.0202***	-.0074***	-.0080***	-.0066***	-.0007*	-.0016***
	(.0029)	(.0016)	(.0016)	(.0008)	(.0003)	(.0003)
Answer rank	-.0270***	-.0189***	-.0184***	-.0266***	-.0258***	-.0244***
	(.0048)	(.0035)	(.0035)	(.0064)	(.0062)	(.0058)
Total contributions	.0126***	.0081***	.0067***	.0074***	.0061***	.0053***
	(.0014)	(.0009)	(.0008)	(.0008)	(.0006)	(.0005)
Tenure	-.0326***	-.0075	-.0085	.0115***	.0127***	.0141***
	(.0058)	(.0049)	(.0049)	(.0024)	(.0022)	(.0024)
Links per post	.1730***	.1561***	.1615***	.0306***	.0314***	.0337***
	(.0168)	(.0154)	(.0152)	(.0029)	(.0029)	(.0030)
Mentions per post	.0124***	.0053**	.0047*	.0063***	.0033***	.0024***
	(.0026)	(.0020)	(.0020)	(.0011)	(.0006)	(.0006)
Readability of posts	.0165***	.0038***	.0018**	.0118***	.0062***	.0046***
	(.0019)	(.0007)	(.0007)	(.0012)	(.0006)	(.0005)
Language prototypicality	.0045**	.0025*	.0026*	.0006*	.0002	.0003
	(.0014)	(.0012)	(.0012)	(.0002)	(.0002)	(.0002)
Positive sentiment	.0614***	.0511***	.0546***	-.0069*	-.0211***	-.0159***
	(.0082)	(.0077)	(.0080)	(.0030)	(.0040)	(.0038)
Social embeddedness		.9843***	.3933***		.1916***	-.2199***
		(.1222)	(.1086)		(.0300)	(.0296)
Epistemic marginality		1.4563***	-.1577		.6435***	-.3988***
		(.0910)	(.0867)		(.0464)	(.0731)
Social embeddedness			2.6938***			1.6970***
× Epistemic marginality			(.2099)			(.1245)
Constant	.2433	-.7190	-.3306	.4700***	.1644	.4382***
	(.4184)	(.3993)	(.3902)	(.0878)	(.1052)	(.1041)
Random effects						
$\sigma^2(v_{0c})$	-.9613***	-1.0770***	-1.0807***	-2.9445***	-2.7974***	-2.8815***
	(.0919)	(.0911)	(.0911)	(.2143)	(.1625)	(.1487)
$\sigma^2(u_{0mc})$	-1.0261***	-1.2284***	-1.2609***	-1.6115***	-1.7155***	-1.7346***
	(.0596)	(.0562)	(.0559)	(.0515)	(.0515)	(.0498)
Observations	992434	992434	992434	479815	479815	479815
Log likelihood	-1194574.47	-1085113.27	-1076535.28	-279759.00	-251733.86	-245586.86
χ^2	4792.66	6499.21	7073.08	3442.88	9081.29	5393.39

Robust standard errors in parentheses, * $p<0.05$, ** $p<0.01$, *** $p<0.001$, Probability > $\chi^2 = 0.0000$ for all models

Table G: Results including observations in communities where members are mostly active (concurrent = max)

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	log(Upvoted posts)			**log(Accepted answers)**		
Community-level						
Community age	-.0189***	-.0143***	-.0142***	-.0063***	-.0049***	-.0048***
	(.0031)	(.0029)	(.0028)	(.0007)	(.0007)	(.0006)
Community membership	-.0000	-.0000	-.0000	-.0000	-.0000	-.0000
	(.0000)	(.0000)	(.0000)	(.0000)	(.0000)	(.0000)
Community crowding	.4574***	.4462**	.4464***	.1021***	.0960***	.1001***
	(.1358)	(.1365)	(.1342)	(.0240)	(.0271)	(.0270)
Community maturity	.0278*	.0194	.0173	-.0018	-.0060	-.0075*
	(.0124)	(.0121)	(.0118)	(.0025)	(.0031)	(.0031)
Member-level						
Concurrent communities	.0922***	.0368***	.0309***	.0205***	.0014	-.0028*
	(.0044)	(.0032)	(.0030)	(.0023)	(.0015)	(.0014)
Pseudonymity	-.0299***	-.0185***	-.0172***	-.0318***	-.0208***	-.0211***
	(.0047)	(.0032)	(.0032)	(.0023)	(.0016)	(.0015)
Comment rank	-.0223***	-.0081***	-.0090***	-.0068***	-.0003	-.0015***
	(.0030)	(.0017)	(.0017)	(.0008)	(.0003)	(.0004)
Answer rank	-.0305***	-.0209***	-.0204***	-.0311***	-.0297***	-.0280***
	(.0055)	(.0040)	(.0039)	(.0069)	(.0066)	(.0062)
Total contributions	.0111***	.0074***	.0060***	.0067***	.0057***	.0050***
	(.0009)	(.0006)	(.0005)	(.0005)	(.0004)	(.0004)
Tenure	-.0375***	-.0097	-.0101*	.0039	.0071***	.0091***
	(.0058)	(.0050)	(.0049)	(.0022)	(.0020)	(.0022)
Links per post	.1916***	.1732***	.1792***	.0373***	.0390***	.0413***
	(.0173)	(.0161)	(.0157)	(.0032)	(.0033)	(.0032)
Mentions per post	.0137***	.0054**	.0046*	.0072***	.0033***	.0021*
	(.0028)	(.0021)	(.0022)	(.0011)	(.0007)	(.0009)
Readability of posts	.0206***	.0055***	.0033***	.0138***	.0072***	.0055***
	(.0016)	(.0007)	(.0006)	(.0008)	(.0005)	(.0004)
Language prototypicality	.0046**	.0024*	.0026*	.0005*	.0002	.0003
	(.0015)	(.0012)	(.0012)	(.0002)	(.0002)	(.0002)
Positive sentiment	.0672***	.0581***	.0621***	-.0080*	-.0226***	-.0162***
	(.0092)	(.0085)	(.0089)	(.0036)	(.0047)	(.0046)
Social embeddedness		1.0720***	.4404***		.2352***	-.2072***
		(.1249)	(.1190)		(.0297)	(.0255)
Epistemic marginality		1.4965***	-.1991*		.6486***	-.4266***
		(.0906)	(.0962)		(.0493)	(.0541)
Social embeddedness			2.7581***			1.7083***
× Epistemic marginality			(.2065)			(.0925)
Constant	.2269	-.7743	-.3496	.5609***	.2234*	.5204***
	(.4335)	(.4171)	(.4038)	(.0896)	(.1079)	(.1072)
Random effects						
$\sigma^2(v_{0c})$	-.8964***	-1.0263***	-1.0297***	-2.8672***	-2.6809***	-2.7467***
	(.0908)	(.0903)	(.0905)	(.2157)	(.1477)	(.1338)
$\sigma^2(u_{0mc})$	-.9494***	-1.1541***	-1.1824***	-1.5182***	-1.6204***	-1.6354***
	(.0549)	(.0536)	(.0544)	(.0419)	(.0427)	(.0419)
Observations	1224275	1224275	1224275	599775	599775	599775
Log likelihood	-1527113.07	-1381997.39	-1370529.55	-386964.09	-351617.63	-343907.98
χ^2	8228.69	13347.84	14930.97	6936.84	11845.30	9615.34

Robust standard errors in parentheses, * p<0.05, ** p<0.01, *** p<0.001, Probability > χ^2 = 0.0000 for all models

Table H: Results excluding members who ask questions (Questions = 0)

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	log(Upvoted posts)			**log(Accepted answers)**		
Community-level						
Community age	-.0156***	-.0110***	-.0110***	-.0044***	-.0034***	-.0034***
	(.0038)	(.0031)	(.0030)	(.0004)	(.0005)	(.0005)
Community membership	.0000	.0000	.0000	-.0000	-.0000	-.0000
	(.0000)	(.0000)	(.0000)	(.0000)	(.0000)	(.0000)
Community crowding	.2451*	.2809**	.2927**	.0663***	.0835**	.0908***
	(.1063)	(.0933)	(.0906)	(.0169)	(.0256)	(.0247)
Community maturity	.0291**	.0223**	.0201*	-.0037*	-.0057*	-.0067**
	(.0109)	(.0085)	(.0082)	(.0018)	(.0026)	(.0023)
Member-level						
Concurrent communities	.0017	-.0036***	-.0043***	-.0010**	-.0027***	-.0033***
	(.0009)	(.0009)	(.0009)	(.0004)	(.0004)	(.0005)
Pseudonymity	-.0372***	-.0216***	-.0214***	-.0227***	-.0125***	-.0129***
	(.0037)	(.0026)	(.0024)	(.0019)	(.0012)	(.0012)
Comment rank	-.0107***	-.0026	-.0020	-.0065***	-.0012**	-.0016***
	(.0018)	(.0014)	(.0014)	(.0009)	(.0004)	(.0004)
Answer rank	-.0572***	-.0443***	-.0436***	-.0232***	-.0240***	-.0229***
	(.0099)	(.0079)	(.0078)	(.0049)	(.0053)	(.0049)
Total contributions	.0134***	.0089***	.0065***	.0081***	.0064***	.0055***
	(.0018)	(.0012)	(.0009)	(.0009)	(.0008)	(.0007)
Tenure	-.0080	-.0003	.0011	.0081**	.0063**	.0073**
	(.0069)	(.0054)	(.0051)	(.0025)	(.0021)	(.0023)
Links per post	.1637***	.1439***	.1496***	.0240***	.0238***	.0262***
	(.0152)	(.0141)	(.0140)	(.0020)	(.0021)	(.0021)
Mentions per post	.0196***	.0087*	.0055	.0063***	.0021	.0008
	(.0040)	(.0043)	(.0049)	(.0014)	(.0015)	(.0016)
Readability of posts	.0140***	.0054***	.0025**	.0114***	.0068***	.0053***
	(.0018)	(.0009)	(.0008)	(.0012)	(.0007)	(.0006)
Language prototypicality	.0060***	.0040***	.0039***	.0008***	.0005**	.0006**
	(.0013)	(.0010)	(.0010)	(.0002)	(.0002)	(.0002)
Positive sentiment	.0116	.0109	.0145*	.0080***	.0005	.0063*
	(.0062)	(.0060)	(.0064)	(.0023)	(.0030)	(.0030)
Social embeddedness		.8462***	.0730		.2885***	-.1045**
		(.1036)	(.0670)		(.0368)	(.0388)
Epistemic marginality		1.3882***	-.7413***		.6514***	-.3833***
		(.0841)	(.0788)		(.0591)	(.0760)
Social embeddedness			3.7282***			1.7229***
× Epistemic marginality			(.1757)			(.1837)
Constant	.0854	-.7153*	-.2685	.5773***	.1778	.4135***
	(.3416)	(.2853)	(.2696)	(.0759)	(.0933)	(.0889)
Random effects						
$\sigma^2(v_{0c})$	-1.2272***	-1.4321***	-1.4320***	-2.9645***	-2.7125***	-2.8115***
	(.0890)	(.0877)	(.0860)	(.1172)	(.1887)	(.1605)
$\sigma^2(u_{0mc})$	-.9919***	-1.2134***	-1.2696***	-1.6680***	-1.8110***	-1.8328***
	(.0528)	(.0492)	(.0474)	(.0553)	(.0522)	(.0484)
Observations	737601	737601	737601	492056	492056	492056
Log likelihood	-853774.24	-778356.82	-762927.98	-252779.58	-217962.33	-211072.46
χ^2	4329.19	8916.00	19811.09	2914.30	2912.66	3744.99

Robust standard errors in parentheses, * p<0.05, ** p<0.01, *** p<0.001, Probability > χ^2 = 0.0000 for all models

Table I: Sample reduction by random sampling

Randomly selecting	10%	5%	1%	10%	5%	1%
	log(Upvoted posts)			**log(Accepted answers)**		
Community-level						
Community age	-.0142***	-.0142***	-.0131***	-.0058***	-.0055***	-.0054***
	(.0031)	(.0032)	(.0032)	(.0009)	(.0011)	(.0014)
Community membership	-.0000	-.0000	-.0000	-.0000	-.0000	.0000
	(.0000)	(.0000)	(.0000)	(.0000)	(.0000)	(.0000)
Community crowding	.4447***	.4476***	.3329**	.0996***	.1218***	.0723*
	(.1284)	(.1275)	(.1101)	(.0225)	(.0285)	(.0329)
Community maturity	.0180	.0194	.0269	-.0087*	-.0077	-.0088
	(.0133)	(.0139)	(.0145)	(.0040)	(.0049)	(.0060)
Member-level						
Multi-community	-.0061***	-.0062***	-.0098***	-.0030***	-.0030***	-.0041*
	(.0016)	(.0016)	(.0022)	(.0007)	(.0009)	(.0019)
Pseudonymity	-.0103**	-.0088*	-.0095	-.0224***	-.0233***	-.0276***
	(.0037)	(.0037)	(.0049)	(.0021)	(.0024)	(.0051)
Comment rank	-.0076***	-.0079***	-.0075***	.0001	.0002	.0006
	(.0016)	(.0017)	(.0020)	(.0005)	(.0005)	(.0008)
Answer rank	-.0244***	-.0238***	-.0239***	-.0336***	-.0318***	-.0356***
	(.0045)	(.0048)	(.0054)	(.0065)	(.0073)	(.0067)
Total contributions	.0071***	.0068***	.0068***	.0055***	.0054***	.0053***
	(.0008)	(.0009)	(.0011)	(.0005)	(.0006)	(.0007)
Tenure	.0027	.0067	.0130	.0234***	.0298***	.0384***
	(.0055)	(.0063)	(.0094)	(.0030)	(.0033)	(.0055)
Links per post	.1754***	.1812***	.1892***	.0344***	.0360***	.0364**
	(.0204)	(.0202)	(.0340)	(.0040)	(.0046)	(.0113)
Mentions per post	.0035	.0005	.0054	.0019	.0026	.0101***
	(.0020)	(.0028)	(.0075)	(.0020)	(.0026)	(.0030)
Readability of posts	.0062***	.0065***	.0072***	.0079***	.0082***	.0073***
	(.0009)	(.0010)	(.0014)	(.0007)	(.0007)	(.0010)
Language prototypicality	.0040**	.0042**	.0045**	.0003	.0001	.0010
	(.0013)	(.0014)	(.0017)	(.0004)	(.0006)	(.0010)
Positive sentiment	.0704***	.0748***	.0728**	-.0243***	-.0193*	-.0234
	(.0118)	(.0141)	(.0269)	(.0066)	(.0076)	(.0130)
Social embeddedness	1.1494***	1.1544***	1.1560***	.2639***	.2700***	.3122***
	(.1287)	(.1374)	(.1428)	(.0354)	(.0391)	(.0414)
Epistemic marginality	1.5964***	1.6318***	1.6578***	.6704***	.6623***	.6491***
	(.0949)	(.0984)	(.1024)	(.0517)	(.0571)	(.0591)
Constant	-.7700	-.8456	-.9846	.3287*	.2626	.3400
	(.4677)	(.4920)	(.5233)	(.1499)	(.1899)	(.2341)
Random effects						
$\sigma^2(v_{0c})$	-1.0671***	-1.0786***	-1.1695***	-2.6265***	-2.6274***	-2.9382***
	(.0954)	(.0967)	(.1058)	(.1459)	(.1521)	(.2767)
$\sigma^2(u_{0mc})$	-1.0309***	-1.0449***	-1.0886***	-1.4999***	-1.5364***	-1.5463***
	(.0466)	(.0530)	(.1699)	(.0467)	(.0514)	(.1739)
Observations	157200	78505	15657	73380	36668	7318
Log likelihood	-183333.97	-91676.60	-18459.91	-44448.31	-22380.33	-4476.77
χ^2	10537.60	9843.98	4678.72	8310.50	6832.35	4965.28

Robust standard errors in parentheses, * $p<0.05$, ** $p<0.01$, *** $p<0.001$, Probability > χ^2 = 0.0000 for all models

Table J: (interaction effects) Sample reduction by random sampling

Randomly selecting	10%	5%	1%	10%	5%	1%
	log(Upvoted posts)			**log(Accepted answers)**		
Community-level						
Community age	-.0139***	-.0140***	-.0129***	-.0056***	-.0054***	-.0053***
	(.0030)	(.0031)	(.0031)	(.0008)	(.0010)	(.0013)
Community membership	-.0000	-.0000	-.0000	-.0000	-.0000	.0000
	(.0000)	(.0000)	(.0000)	(.0000)	(.0000)	(.0000)
Community crowding	.4431***	.4419***	.3353**	.1021***	.1230***	.0765*
	(.1258)	(.1246)	(.1092)	(.0219)	(.0278)	(.0326)
Community maturity	.0164	.0174	.0267	-.0098**	-.0093*	-.0090
	(.0130)	(.0135)	(.0140)	(.0036)	(.0044)	(.0055)
Member-level						
Concurrent communities	-.0067***	-.0068***	-.0104***	-.0036***	-.0036***	-.0047*
	(.0016)	(.0016)	(.0022)	(.0007)	(.0009)	(.0019)
Pseudonymity	-.0087*	-.0069	-.0070	-.0224***	-.0232***	-.0266***
	(.0036)	(.0036)	(.0048)	(.0020)	(.0023)	(.0051)
Comment rank	-.0084***	-.0087***	-.0083***	-.0011*	-.0011*	-.0006
	(.0016)	(.0016)	(.0019)	(.0005)	(.0005)	(.0008)
Answer rank	-.0236***	-.0230***	-.0232***	-.0316***	-.0298***	-.0335***
	(.0044)	(.0047)	(.0052)	(.0061)	(.0068)	(.0062)
Total contributions	.0057***	.0055***	.0055***	.0048***	.0047***	.0048***
	(.0006)	(.0007)	(.0009)	(.0004)	(.0005)	(.0006)
Tenure	.0004	.0037	.0077	.0243***	.0306***	.0378***
	(.0054)	(.0062)	(.0095)	(.0030)	(.0033)	(.0056)
Links per post	.1823***	.1893***	.1968***	.0373***	.0398***	.0391**
	(.0204)	(.0200)	(.0353)	(.0040)	(.0044)	(.0119)
Mentions per post	.0028	-.0002	.0045	.0007	.0011	.0093**
	(.0021)	(.0029)	(.0076)	(.0021)	(.0026)	(.0035)
Readability of posts	.0035***	.0036***	.0041**	.0060***	.0061***	.0053***
	(.0009)	(.0009)	(.0014)	(.0006)	(.0006)	(.0010)
Language prototypicality	.0041**	.0042**	.0043**	.0003	.0000	.0009
	(.0013)	(.0014)	(.0017)	(.0004)	(.0006)	(.0009)
Positive sentiment	.0746***	.0806***	.0775**	-.0184**	-.0123	-.0182
	(.0125)	(.0147)	(.0274)	(.0067)	(.0073)	(.0130)
Social embeddedness	.4208***	.4097**	.4411***	-.2003***	-.2178***	-.1316*
	(.1157)	(.1247)	(.1190)	(.0317)	(.0347)	(.0641)
Epistemic marginality	-.3260**	-.3534**	-.2712	-.4415***	-.5162***	-.4206**
	(.1108)	(.1205)	(.1658)	(.0721)	(.0778)	(.1302)
Social embeddedness	3.1280***	3.2155***	3.1145***	1.7556***	1.8561***	1.6867***
× Epistemic marginality	(.2221)	(.2424)	(.2865)	(.1237)	(.1401)	(.2154)
Constant	-.3252	-.3721	-.5974	.6217***	.5875***	.5843**
	(.4530)	(.4776)	(.5067)	(.1397)	(.1754)	(.2166)
Random effects						
$\sigma^2(v_{0c})$	-1.0767***	-1.0905***	-1.1768***	-2.6613***	-2.6650***	-3.0342***
	(.0951)	(.0965)	(.1054)	(.1431)	(.1488)	(.2903)
$\sigma^2(u_{0mc})$	-1.0671***	-1.0726***	-1.1758***	-1.5194***	-1.5483***	-1.6168***
	(.0496)	(.0561)	(.2014)	(.0452)	(.0499)	(.1886)
Observations	157200	78505	15657	73380	36668	7318
Log likelihood	-181513.16	-90712.00	-18283.98	-43425.23	-21806.78	-4378.94
χ^2	12630.02	11295.62	4605.70	6039.92	5610.79	4365.27

Robust standard errors in parentheses, * $p<0.05$, ** $p<0.01$, *** $p<0.001$, Probability > χ^2 = 0.0000 for all models

Table K: Sample reduction by the systematic exclusion of top 1% of members (based on # of contributions)

	log(Upvoted posts)			log(Accepted answers)		
Community-level						
Community age	-.0156***	-.0133***	-.0132***	-.0051***	-.0042***	-.0042***
	(.0027)	(.0026)	(.0026)	(.0006)	(.0006)	(.0006)
Community membership	-.0000	-.0000	-.0000	-.0000	-.0000	-.0000
	(.0000)	(.0000)	(.0000)	(.0000)	(.0000)	(.0000)
Community crowding	.3001**	.3559**	.3615**	.0758***	.0774***	.0778***
	(.1131)	(.1153)	(.1140)	(.0172)	(.0202)	(.0177)
Community maturity	.0228*	.0179	.0176	-.0017	-.0049	-.0045
	(.0114)	(.0113)	(.0112)	(.0023)	(.0028)	(.0026)
Member-level						
Multi-community	-.0024*	-.0042***	-.0044***	-.0009*	-.0022***	-.0023***
	(.0010)	(.0011)	(.0011)	(.0003)	(.0004)	(.0003)
Pseudonymity	-.0092*	-.0063	-.0059	-.0256***	-.0179***	-.0170***
	(.0041)	(.0034)	(.0033)	(.0021)	(.0014)	(.0014)
Comment rank	-.0147***	-.0066***	-.0066***	-.0056***	-.0008*	-.0015***
	(.0024)	(.0015)	(.0015)	(.0007)	(.0003)	(.0003)
Answer rank	-.0235***	-.0196***	-.0193***	-.0267***	-.0259***	-.0238***
	(.0048)	(.0040)	(.0039)	(.0057)	(.0056)	(.0050)
Total contributions	.0432***	.0285***	.0297***	.0099***	.0083***	.0097***
	(.0038)	(.0030)	(.0029)	(.0007)	(.0006)	(.0012)
Tenure	-.0274***	-.0139**	-.0145**	.0058**	.0068***	.0104***
	(.0052)	(.0049)	(.0048)	(.0020)	(.0019)	(.0018)
Links per post	.1963***	.1852***	.1865***	.0336***	.0348***	.0322***
	(.0180)	(.0171)	(.0169)	(.0026)	(.0027)	(.0024)
Mentions per post	.0069**	.0024	.0022	.0044***	.0020**	.0016*
	(.0022)	(.0023)	(.0023)	(.0008)	(.0007)	(.0006)
Readability of posts	.0023	-.0027**	-.0034***	.0095***	.0051***	.0027***
	(.0019)	(.0010)	(.0009)	(.0009)	(.0005)	(.0005)
Language prototypicality	.0043***	.0029*	.0029*	.0005*	.0002	.0003
	(.0013)	(.0011)	(.0011)	(.0002)	(.0002)	(.0002)
Positive sentiment	.0873***	.0767***	.0780***	-.0069*	-.0173***	-.0138***
	(.0105)	(.0097)	(.0097)	(.0034)	(.0043)	(.0040)
Social embeddedness		.8892***	.7449***		.1517***	.0014
		(.1126)	(.1330)		(.0233)	(.0253)
Epistemic marginality		1.0576***	.6793***		.5173***	.1997**
		(.0736)	(.1389)		(.0378)	(.0764)
Social embeddedness			.5873*			.3537*
× Epistemic marginality			(.2608)			(.1409)
Constant	.1150	-.6305	-.5538	.4893***	.2376*	.2992**
	(.4030)	(.3865)	(.3864)	(.0817)	(.0954)	(.0944)
Random effects						
$\sigma^2(v_{0c})$	-1.1044***	-1.1589***	-1.1567***	-2.9442***	-2.7754***	-2.8586***
	(.0915)	(.0920)	(.0922)	(.2552)	(.1442)	(.1598)
$\sigma^2(u_{0mc})$	-1.2000***	-1.2729***	-1.2807***	-1.6483***	-1.7196***	-1.8394***
	(.0590)	(.0565)	(.0568)	(.0420)	(.0411)	(.0438)
Observations	1556434	1556434	1556434	731961	731961	731961
Log likelihood	-1808314.60	-1713572.47	-1706597.29	-413485.89	-384118.66	-336501.29
χ^2	6616.96	8248.63	8290.87	5120.55	9099.24	7097.85

Robust standard errors in parentheses, * p<0.05, ** p<0.01, *** p<0.001, Probability > χ^2 = 0.0000 for all models

Table L: Last quarter analysis including the moderator role (quarter = 2016Q4)

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
	log(Upvoted posts)			**log(Accepted answers)**		
Community-level						
Community age	-.0928*	-.0367	-.0386	-.0628*	.0007	.0002
	(.0404)	(.0219)	(.0224)	(.0300)	(.0032)	(.0032)
Community membership	-.0005**	-.0002**	-.0002**	-.0004**	.0000	.0000
	(.0002)	(.0001)	(.0001)	(.0001)	(.0000)	(.0000)
Community crowding	-.4979	-.2776	-.2790	-.0365	.0870	.0862
	(.5422)	(.2652)	(.2787)	(.4173)	(.0525)	(.0543)
Community maturity	-.0486	-.0058	-.0092	-.0636	-.0090	-.0096
	(.0573)	(.0316)	(.0324)	(.0424)	(.0068)	(.0067)
Member-level						
Concurrent communities	-.0037	-.0081***	-.0085***	-.0023*	-.0042***	-.0047***
	(.0019)	(.0021)	(.0022)	(.0009)	(.0009)	(.0009)
Pseudonymity	-.0067	-.0027	-.0015	-.0184***	-.0142***	-.0143***
	(.0045)	(.0039)	(.0040)	(.0026)	(.0024)	(.0024)
Comment rank	-.0309***	-.0167***	-.0167***	-.0046***	.0003	-.0006
	(.0032)	(.0025)	(.0024)	(.0006)	(.0006)	(.0007)
Answer rank	-.0634***	-.0480***	-.0466***	-.0387***	-.0365***	-.0346***
	(.0089)	(.0073)	(.0071)	(.0041)	(.0040)	(.0039)
Total contributions	.0108***	.0074***	.0060***	.0065***	.0056***	.0051***
	(.0011)	(.0008)	(.0006)	(.0006)	(.0005)	(.0005)
Tenure	.0028	.0019	-.0007	.0225***	.0183***	.0180***
	(.0050)	(.0047)	(.0047)	(.0033)	(.0031)	(.0031)
Links per post	.2267***	.2104***	.2195***	.0308***	.0310***	.0344***
	(.0235)	(.0227)	(.0222)	(.0046)	(.0047)	(.0046)
Mentions per post	.0042	.0010	.0006	.0012	.0004	-.0002
	(.0022)	(.0016)	(.0016)	(.0018)	(.0016)	(.0015)
Readability of posts	.0212***	.0065***	.0045***	.0141***	.0068***	.0056***
	(.0020)	(.0011)	(.0011)	(.0014)	(.0008)	(.0008)
Language prototypicality	.2774***	.1325***	.1403***	.1928***	.0037	.0055*
	(.0264)	(.0146)	(.0156)	(.0199)	(.0025)	(.0026)
Positive sentiment	.0958***	.0834***	.0867***	-.0128*	-.0135*	-.0088
	(.0119)	(.0117)	(.0117)	(.0062)	(.0067)	(.0069)
Moderator role	.4328**	.3508**	.3109**	.1959	.1750	.1548
	(.1656)	(.1222)	(.1110)	(.1078)	(.0924)	(.0874)
Social embeddedness		.9137***	.2492**		.2113***	-.1261**
		(.1070)	(.0844)		(.0325)	(.0388)
Epistemic marginality		1.3465***	-.5226***		.5665***	-.2880***
		(.0937)	(.1153)		(.0601)	(.0753)
Social embeddedness			3.0453***			1.3341***
× Epistemic marginality			(.2076)			(.1527)
Constant	3.9153	.6984	1.1733	2.7973	.1899	.4038
	(2.0708)	(1.1297)	(1.1693)	(1.5651)	(.2999)	(.3024)
Random effects						
$\sigma^2(v_{0c})$	.1737*	-.5211***	-.4811***	-.1192	-2.5329***	-2.5487***
	(.0733)	(.0766)	(.0780)	(.0861)	(.1908)	(.2235)
$\sigma^2(u_{0mc})$	-.2633***	-.3568***	-.3660***	-.8345***	-.9242***	-.9326***
	(.0253)	(.0165)	(.0395)	(.0207)	(.0308)	(.0091)
Observations	79092	79092	79092	31128	31128	31128
Log likelihood	-96635.87	-89501.59	-88540.21	-18948.42	-17515.32	-17235.04
χ^2	1339.98	2737.36	3603.05	1813.32	2450.18	2440.11

Robust standard errors in parentheses, * p<0.05, ** p<0.01, *** p<0.001, Probability > χ^2 = 0.0000 for all models

Table M: Heckman specification to correct for self-selection bias

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
		log(Upvoted posts)			**log(Accepted answers)**	
Community age	-.0203***	-.0145***	-.0143***	-.0071***	-.0054***	-.0052***
	(.0003)	(.0002)	(.0002)	(.0002)	(.0002)	(.0002)
Community membership	-.0000***	-.0000***	-.0000***	-.0000**	-.0000***	-.0000***
	(.0000)	(.0000)	(.0000)	(.0000)	(.0000)	(.0000)
Community crowding	.4634***	.4591***	.4598***	.0979***	.0940***	.0986***
	(.0119)	(.0120)	(.0119)	(.0067)	(.0061)	(.0059)
Community maturity	.0222***	.0147***	.0129***	-.0035**	-.0080***	-.0092***
	(.0013)	(.0011)	(.0011)	(.0011)	(.0010)	(.0010)
Concurrent communities	-.0027***	-.0063***	-.0068***	-.0018***	-.0037***	-.0043***
	(.0003)	(.0003)	(.0003)	(.0003)	(.0002)	(.0002)
Pseudonymity	-.0051***	-.0005	-.0019**	-.0209***	-.0099***	-.0118***
	(.0010)	(.0007)	(.0007)	(.0009)	(.0008)	(.0008)
Comment rank	-.0246***	-.0078***	-.0086***	-.0076***	.0001	-.0011***
	(.0003)	(.0001)	(.0001)	(.0001)	(.0001)	(.0001)
Answer rank	-.0368***	-.0242***	-.0236***	-.0337***	-.0315***	-.0296***
	(.0004)	(.0002)	(.0002)	(.0007)	(.0007)	(.0006)
Total contributions	.0116***	.0072***	.0058***	.0070***	.0058***	.0050***
	(.0007)	(.0004)	(.0004)	(.0003)	(.0003)	(.0003)
Tenure	-.0098***	-.0057***	-.0064***	.0337***	.0223***	.0231***
	(.0012)	(.0009)	(.0009)	(.0011)	(.0009)	(.0009)
Links per post	.2181***	.1950***	.2017***	.0377***	.0412***	.0438***
	(.0060)	(.0054)	(.0056)	(.0015)	(.0016)	(.0017)
Mentions per post	.0144***	.0041***	.0032***	.0071***	.0028***	.0016**
	(.0016)	(.0007)	(.0006)	(.0013)	(.0006)	(.0005)
Readability of posts	.0244***	.0060***	.0034***	.0164***	.0081***	.0062***
	(.0008)	(.0003)	(.0002)	(.0006)	(.0003)	(.0002)
Language prototypicality	.0063***	.0034***	.0035***	.0006***	.0002*	.0003**
	(.0002)	(.0001)	(.0001)	(.0001)	(.0001)	(.0001)
Positive sentiment	.0769***	.0682***	.0731***	-.0108***	-.0245***	-.0175***
	(.0020)	(.0017)	(.0017)	(.0013)	(.0012)	(.0012)
Social embeddedness		1.1546***	.4328***		.2555***	-.2010***
		(.0099)	(.0137)		(.0115)	(.0142)
Epistemic marginality		1.6127***	-.3219***		.6901***	-.4267***
		(.0150)	(.0433)		(.0132)	(.0458)
Social embeddedness × Epistemic marginality			3.1293***			1.7537***
			(.0871)			(.0884)
Constant	1.4444***	-.0833	.3539***	.9180***	.3790***	.6498***
	(.0542)	(.0428)	(.0429)	(.0447)	(.0388)	(.0397)
Inverse mills ratio	-.30***	-.20***	-.15***	-.22***	-.23***	-.20***
	(.005)	(.005)	(.004)	(.009)	(.010)	(.010)
Posting / Answering						
Tenure	.2340***	.2298***	.2290***	.2340***	.0483***	.0477***
	(.0017)	(.0017)	(.0017)	(.0017)	(.0013)	(.0013)
Pseudonymity	-.1563***	-.1541***	-.1537***	-.1563***	-.0650***	-.0645***
	(.0010)	(.0010)	(.0010)	(.0010)	(.0011)	(.0011)
Constant	1.2996***	1.2952***	1.2946***	1.2996***	-.1271***	-.1276***
	(.0038)	(.0038)	(.0038)	(.0038)	(.0038)	(.0038)
Observations	1906434	1906434	1906434	1906434	1906434	1906434

Log likelihood	-2886441	-2671632	-2653675	-1767278	-1716752	-1706995
χ^2	177009.12	465525.50	578394.94	71311.01	87914.26	95240.59

Clustered robust standard errors in parentheses, community dummies included but not shown, * $p<0.05$, ** $p<0.01$, *** $p<0.001$, Probability > χ^2 = 0.0000 for all models