Published:
Xiaoquan (Michael) Zhang and Chong (Alex) Wang, "Network Positions and Contributions to Online Public Goods: The Case of Chinese Wikipedia," Journal of Management Information Systems, 29(2), Fall 2012, 11-40.

Network Positions and Contributions to Online Public Goods: The Case of Chinese Wikipedia*

XIAOQUAN (MICHAEL) ZHANG[+] AND CHONG (ALEX) WANG[++]

[+] Department of ISOM, School of Business and Management, Hong Kong University of Science and Technology, Email: zhang@ust.hk

[++] Department of ISOM, School of Business and Management, Hong Kong University of Science and Technology and Department of Information Systems, College of Business, City University of Hong Kong, Email: wangch3.04@gmail.com

XIAOQUAN (MICHAEL) ZHANG is an Assistant Professor of Information Systems, Business Statistics and Operations Management in the School of Business and Management, Hong Kong University of Science and Technology. He holds a PhD in Management from MIT Sloan School of Management, an MSc in Management, a BE in Computer Science and a BA in English from Tsinghua University. His research interests are on issues related to creation, dissemination and processing of information in business and management contexts. His works study pricing of information goods, online word-of-mouth, online advertising, incentives of creation in open source and open content projects, and use of information in financial markets. His research has appeared in *American Economic Review*, *Management Science*, *MIS Quarterly*, *Journal of Marketing*, *Decision Support Systems*, and *Journal of Interactive Marketing*. He currently serves as an Associate Editor for *Information Systems Research* and is also on the editorial boards of *Production and Operations Management* and *Electronic Commerce Research and Applications.*

CHONG (ALEX) WANG is a PhD candidate in information systems in the School of Business and Management, Hong Kong University of Science and Technology. He will join the faculty of the Department of Information Systems, City University of Hong Kong. His research interests include online social network, IT facilitated mass collaboration, and economics of social media.

---

* An earlier version of this paper appeared as Zhang and Wang, Network Centrality and Contribution to Online Public Good – The Case of Chinese Wikipedia. In R. Sprague (ed.), Proceedings of the 45th Hawaii International Conference on Systems Science, Maui, HI, January 2012, IEEE Computer Society Press, Washington, DC, 2012.

ABSTRACT: In this paper, we study the effect of collaboration network structure on the contribution behavior of participating editors in Wikipedia. Collaboration in Wikipedia is organized around articles, and any two editors co-editing an article have a collaborative relationship. Based on economic theories about network games and social role theory, we propose that an editor's position in the collaboration network influences her decisions about her total contribution as well as the allocation of her efforts. By leveraging panel data collected from the Chinese language version of Wikipedia and a natural experiment resulting blocking it in Mainland China, we find strong support for the proposed effect of network position on contribution behavior. Our analysis further reveals that different aspects of an individual's network position have distinct implications. This research enhances our understanding about how collaboration network structure shapes individual behavior in online mass collaboration platforms.

KEY WORDS AND PHRASES: Effort allocation, mass collaboration, natural experiment, network centrality, online public goods, Wikipedia.

1. Introduction

Internet technologies enable virtual collaboration and play important roles in knowledge production. Open-source software (OSS) projects (e.g., Linux, Apache, and MySQL), Wikipedia, and collaborative projects based on content-sharing platforms (e.g., Flickr and YouTube) have attracted considerable attention from both business leaders and academic researchers. Mass collaborative projects are not only viable but also sustainable. For example, by the end of 2010, the English language version of Wikipedia had accumulated over 3.5 million articles by more than 34,000 registered editors. Knowledge about the organization of virtual collaboration in Wikipedia and other collaborative platforms is still limited, however [40]. Several studies examine the incentives that motivate contributions to these online public goods, with a focus on OSS projects [29,30,31,36,44]. Such studies typically take a utilitarian perspective and do not explicitly consider the effects of collaborative relationships. Online public goods provision, however, depends crucially on its editors' mass collaborative efforts, which are characterized by "free user-to-user assistance" [30]. Examining the structure of collaborative relationships organized as a social network is critical to understanding the production of online public goods.

Recently, researchers have started to explore the structural mechanisms through which outcomes of mass collaboration projects are determined [22,23,37,38]. This line of research characterizes the organization of collaborative relationships as networks of actors and discusses the effects of the structural properties of these networks. Most previous studies have been carried out in the context of OSS projects. While open-source software projects require programming expertise and creative content contribution, Wikipedia articles involve mostly informational contributions. The scope of collaborative efforts involved in creating Wikipedia content is much

broader than that of a small software development group. The editors who participate in Wikipedia bring expertise of various types. Collaborative efforts in Wikipedia take the forms of content contribution, writing refinement, referencing, identifying new topics, etc. [1]. The editors of an article are collaborators who interact with each other via "talk" pages –wiki-edited pages that are associated with their articles. Their shared goal is to maintain and extend the free online public encyclopedia. The collaboration network in Wikipedia is indispensable to bringing different types of efforts together for the refinement of articles. In spite of its importance, little prior research has studied the effect of collaboration network structure on editors' contribution behavior.

In this study, we extend current research on mass collaboration in the context of online public goods provision by adopting a structural view of the collaboration network. Our basic research question is how the position of a Wikipedia editor in the global collaboration network influences her future contribution behavior. An editor's position in the network refers to those with whom she connects, both directly and indirectly, through other editors. Specifically, we investigate how different aspects of an editor's position in the collaboration network, as reflected by centrality measures, influence her contribution behavior. We study the effects on both the total contribution and the allocation of effort. Although a few studies in the literature investigate the relationship between the total contribution and network centralities based on social capital theory [40], to our knowledge, the current study is the first to look at the allocation of effort. This extension makes it essential to also look at network positions from the *coordination* perspective, focusing on the structural determination of cooperation and behavior convergence. To establish the link between network position and contribution behavior, we refer to recent developments in network economics as well as the social role theory.

Empirical study of the effect of network position on contribution behavior suffers from an issue related to the endogenous determination of network position. That is, by definition, a collaboration network is the result of past collaborating behaviors. Combined with editors' unobservable characteristics, empirical identification of a causal relationship between network positions and contribution behaviors could be contaminated. Our study leverages a natural experiment to resolve this issue. For political considerations, Wikipedia was blocked a few times by the government of Mainland China. The blockage of editors from Mainland China removed some editors from the collaboration network of Chinese Wikipedia. As a result of the blockages, unaffected editors experienced changes in their network positions that were not related to their own contributions or other individual-level characteristics. Through the use of this natural experiment, our empirical analysis is able to establish a causal relationship between an individual's network position and contribution behavior. Our estimation further reveals the complexity of the effects of network positions, as reflected by the related centrality measures. We find that different aspects of network positions have different implications for editors' contribution behaviors, which indicate distinct role-taking behaviors.

Our major findings are: (a) *degree centrality* (a focal editor's number of ties as compared to the size of the network) is associated with a content-expert role: when an editor has more collaborators (an increase in degree centrality), she reduces her total contribution and focuses her efforts on articles that she created; (b) *closeness centrality* (the average distance between the focal user and users in the rest of the network) is associated with a generalist role: when an editor gets closer to others (an increase in closeness centrality), she reduces and diversifies her total contribution; and (c) *betweenness centrality* (the number of times a focal editor lies on the shortest path between other editors in the network) is associated with a role of a content

extender: when an editor becomes more critical in bridging collaborative relationships (an increase in betweenness centrality), she increases her total contribution as well as her focus on the articles she created.

This paper adds to the literature on IT-enabled mass collaboration by bringing a structural perspective to collaboration networks on such platforms as Wikipedia. This study's contribution is both confirmatory and exploratory. This paper extends our understanding about the effects of collaboration network structure on individual contribution behavior through the use of network game theory and the network perspective in organization theory. Second, by leveraging panel data collected from Chinese Wikipedia and the blockages of the site from Mainland China, we offer causal empirical evidence of the effects of network position on future contribution behavior. Our empirical analyses not only confirm the theoretical prediction of the relevance of network position in predicting an editor's total contribution and effort allocation, but also establish the differential behavior implications of the various centrality measures.

The rest of the paper is organized as follows. In the next section, we introduce the network perspective, following which we briefly review theories about network games and social roles. We develop three hypotheses based on these theories. In the next section, we describe our empirical context and the natural experiment. We then introduce our empirical study design and present the results.

2. Network Perspective - Centrality Measures

A *social network* consists of a collection of actors and the connections (a certain type of relationship, social ties) between them.[2] In the context of Wikipedia, collaboration ties are established once two editors edit the same page; and the collaboration network is the summation of all editors and collaboration ties. Network methods focus on the patterns and types of ties among actors or subsets of actors as a way to formalize the notion of social roles [41]. Every social decision that individuals make is embedded in the social context that could be represented by a network of connections [20]. A network perspective is critical for understanding the consequences of a network environment within which people and groups are embedded as they interact and engage in productive work. Over the past few decades, we have learned a great deal about what kinds of networks produce desirable outcomes and what situational characteristics shape the possible environments within which people and organizations construct their social networks [26].

Of all the network concepts, centrality measures are among the most extensively discussed. Overall, such measures gauge the position of an individual actor in the network [25,26,33,40] and capture the idea of an actor being important, well-connected, or influential within a group of people related to each other.

The social network literature offers a wide variety of conceptualizations of centralities with different focuses. Freeman synthesizes the literature and classifies proposed measures into three

[2] To be consistent with the social network analysis literature, we refer to people in the network as *actors*. In the context of Wikipedia collaboration, these actors refer to *editors* of Wikipedia pages. Thus, editors and actors are used interchangeably in the current paper.

categories, namely, degree centrality, closeness centrality, and betweenness centrality [17]. These measures capture different aspects of what it means for a position to be central in a network and, therefore, they have different implications for actors' behaviors.

Let us consider a network *N*, that is, a set of *actors* {A, B, C, D…} joined by *ties*. The *size* of a network is the number of actors in it. Two actors A and B are neighbors if there is a tie between A and B. A *path* between A and B is a sequence of actors, with A and B at the first and last places, respectively, such that there is a tie connecting every consecutive pair of actors in the sequence. The *length of a path* is the number of ties on the path. For example, if A connects to C, C connects to D, and D connects to B, then there is a path of length 3 between A and B, which goes through C and D. The *distance* between any two actors in a network is the length of the shortest path between them. It is possible for there to be no path between a pair of actors in a network. If there exists a path between every pair of actors, the network is *connected*, but if one or more pairs of actors are not connected by a path, the network is divided into disconnected *components*. A component of a network is a subset of actors such that every pair of actors in the subset is connected by some path that runs through other actors in the set. A connected network has one component.[3]

With these basic definitions, the *degree centrality* of actor *i* is defined as the number of her (direct) ties, normalized by the size of the network:

$$DegreeCentrality_i \equiv \frac{N_i}{N}.$$

[3] Wasserman and Faust [41] give a detailed introduction to networks.

Here N_i is the number of ties that user i has, and N is the total number of users in the network. If an actor has high degree centrality, she has direct connections to many others in the network and thus is well connected. Normalization with respect to network size enables the centrality measure to represent relative positions and to be comparable across networks. This is especially important for longitudinal network observations where an individual's position changes with both her direct connections and the overall network structure.

Closeness centrality, as suggested by the name, is based on relative closeness or distance. It is defined as the inverse of the average distance between actor i and the other actors in the network:

$$ClosenessCentrality_i \equiv \frac{N-1}{\sum_{j \neq i} D_{ij}},$$

where D_{ij} is the shortest distance between i and j. A higher value of closeness centrality indicates that an actor is easily accessible from any part of the network through the network connections. An actor with high closeness centrality thus enjoys the structural advantage of having an easy access to resources located in other parts (actors) of the network. High values of degree centrality and closeness centrality both indicate that an actor is well connected; these two measures are often positively correlated, but degree centrality focuses on local connectedness while closeness centrality focuses on global connectedness.

Interaction between non-adjacent actors might depend on other actors in the network, especially those who lie on the path between the two. *Betweenness centrality* captures the importance of an actor in bridging social connections. It is defined as the number of times that an actor lies on the shortest path between a pair of other editors, normalized by the number of all possible pairs of editors:

$$BetweennessCentrality_i \equiv \frac{\sum_{s,t\in\{1,\ldots,N\},s\neq t,s\neq i,t\neq i}\frac{NS_{st}(i)}{NS_{st}}}{\frac{1}{2}(N-1)(N-2)}.$$

Here, NS_{st} is the number of shortest paths between actor *s* and actor *t*, and $NS_{st}(i)$ is the number of shortest paths between *s* and *t* that go through *i*. Structural hole theory states that individuals who take positions of high betweenness centrality are critical to bridging information and resources, as well as coordinating interactions between other members of the network. These individuals (actors filling the structural holes) enjoy high structural social capital [10,11].[4]

These three centrality measurements are the most prevalent conceptualizations of an actor's centrality, and each one captures a different aspect of an actor being central in the network. Social network researchers extend centrality measurements in different ways, taking into consideration other structural features of the network, such as whether the tie is directed. Most of these extensions, however, follow the basic intuitions behind the above three measures. Considering that Wikipedia collaboration focuses on content contribution and page maintenance and puts little emphasis on individual identification, we focus on these three global measurements of individual centrality in this paper.

3. Relationship between Network Position and Contribution Behavior

Two theoretical perspectives are useful in characterizing the effects of network positions on actors' contribution behaviors.

[4] Consider a network of 4 actors, A, B, C, and D, located at the four corners of a square, respectively. In other words, there are 4 ties in the network, A-B, B-C, C-D, D-A. There are two paths between A and D: One goes through B, and the other goes through C. Thus NS_{AD}=2, and $NS_{AD}(B)$=1. $BetweennessCentrality_B = \frac{1/2}{0.5\times3\times2} = \frac{1}{6}$.

3.1 Network Games

The first perspective is from the economic theory of network games. It is well recognized that economic behaviors are embedded in their social contexts [20,21]. Traditional microeconomic theories, however, make simplifying assumptions about the structure of social interactions between actors. Only recently have economists begun to include the structure of interpersonal interactions (i.e., social networks) explicitly in their model specifications. A common feature of these emerging models is that interactions are structured in the form of networks. Connections in the network differ among players. Jackson [27] provides a comprehensive introduction to the theory of network games. Of specific relevance to our current study are games where economic agents make choices whose payoffs depend on the choices of their network neighbors. In this case, the network's structure is important because it determines the pattern of externalities that exist in the network.

Relevant to the Wikipedia context, in one study, Ballester et al. [2] develop a model where a player's marginal utility from engaging in an activity depends on the activity levels of others to whom the focal player is connected. For this game, a unique equilibrium exists where an individual's activity level is proportional to her Bonacich centrality in the network.[5] In another study, Bramoullé and Kranton [8] model contributions to public goods where social neighbors' contributions are perfect substitutes. They show that equilibrium contribution behavior is related to the network's maximal independent set, where an independent set refers to a group of users who are not connected to each other. Only those in the maximal independent set will make

[5] A social network could be represented by an *adjacency matrix* with the cell on the *i*th row and *j*th column representing the tie between actor *i* and actor *j*. Bonacich defines centrality based on the eigenvector of the adjacency matrix. The idea of *Bonacich centrality* captures the power of an actor and can be seen as an extension of the degree centrality [4].

positive contributions in equilibrium. The above model assumes that each player has complete knowledge of the network structure. Galeotti et al. [18] propose a general network game in which actors have no information about connections beyond their immediate friends and only know the distribution of individual degrees. Such a Bayesian game has unique pure-strategy equilibrium. A player's equilibrium strategy is increasing in her degree if neighbors' actions are strategic complements and decreasing in her degree if neighbors' actions are strategic substitutes.

In the case of Wikipedia, editors exert efforts to improve the quality of articles (pages). They derive utility from both enjoyment and self-fulfillment [29]. An editor can track the activity of others by examining the history of a page, which is automatically recorded by the system [35]. Editors thus know who else is also contributing to the article and collaborating. They can also interact through the talk pages. Depending on the article and their personal preferences, efforts from different editors may be complements or substitutes. On the one hand, when efforts from other editors are considered as substitutes, editors have an incentive to free-ride. On the other hand, if editors focus on reciprocal relationships and value social interactions, their contributions will be positively related to others' contributions, generating effort complementarity [39]. Either way, interactions between collaborating editors influence the total contribution as well as the allocation of efforts across different articles. Consequently, contribution behavior in a given situation depends on an editor's position in the collaboration network.

3.2 Social Role Theory

The second theoretical perspective is the *social role theory*, which relates structural positions to role-taking and role-playing behavior. Game theoretical analyses relate network positions to actions taken by individuals through rational optimization and equilibrium models. This line of research reveals the effect of network positions by discussing the fundamental trade-offs that

actors face. Any formal analysis, however, requires a high level of abstraction and simplifying assumptions about actors in the network. It unavoidably omits certain psychological processes, such as role-taking and role-playing behavior. Rational economic models should be complemented by the social role theory. In this theory, an individual's action in a given situation is explained by the situation-triggered role (role-taking) and the behaviors associated with the role (role-playing). For example, it has been shown in laboratory experiments that triggering a group identity could alter a game's equilibrium outcome [12].

Homans [24, p. 11] defines the term *role* as "the behavior expected of a person occupying a particular social position." "The total role of an individual in a social system has often been described as consisting of sets of relations of various types linking this person as ego to sets of others," as Lorrain and White [32] note. Role also is defined in terms of the pattern of social relationships that form resources and constraints for actions, leading to a set of prototypical behaviors. People adopt a set of roles in different social situations. Different role sets may be triggered depending on the context of social interactions.

Role analysis is one of the most important building blocks of social network theories. Social network analysis extends the role-making literature and tries to explain the commonality in attitudes and practices as results of similarity in network positions, with the latter often conceptualized in terms of centrality measures or other structural equivalents [6]. A network analysis of roles focuses on ties among actors and the regularities in patterns of relations that indicate regularities in roles of actors holding those social positions [15]. That is, actors occupying similar positions in the network behave similarly. The premise behind a network analysis of roles is that behavioral consistency will lead to a certain structural signature that then can be used to infer roles in certain positions. A significant amount of empirical evidence exists

showing that positions in social networks predict role-related psychological constructs. For example, researchers report that people taking similar positions in a social network report similar attitudes [5,9,13].

In addition, some recent studies use data collected from online communities to identify role-taking behavior [19,43]. Identity construction has been shown to be a major element in predicting sustained participation in OSS communities [14]. Gleave et al. [19], for example, identify two distinctive roles of Wikipedia editors, namely, substantive experts and technical editors. These roles are shown to be related to the different characteristics of editors' positions in the network. Combining this result on specific roles and structural hole theory, we postulate that, in the case of Wikipedia, network positions in the collaboration network (as indicated by various centrality measures) reflect editors' role-taking behaviors and, therefore, predict future contributions.

3.3 Hypotheses

Based on our review of the literature about network economics and social role theory, we have the following hypotheses.

- **Hypothesis 1 (The Total Contribution Hypothesis):** *An editor's total contribution is correlated with her centrality measures in the collaboration network.*
- **Hypothesis 2 (The Effort Allocation Hypothesis):** *An editor's effort allocation is correlated with her centrality measures in the collaboration network.*
- **Hypothesis 3 (The Differential Effect of Centrality Measures Hypothesis):** *Different centrality measures have differential effects on the total contribution and effort allocation, indicating the presence of different role-taking behaviors.*

Hypothesis 1 relates an editor's total contribution to network centrality measures. It is rooted in both game theoretic models as well as the social capital perspective [38,40]. Hypothesis 2 further relates effort allocation decisions to network positions. Informed by the social role theory, we propose that different centrality measures indicate differential role-taking (acting) behavior, thus they have differential effects on total contribution and effort allocation, which leads to Hypothesis 3. Existing theories provide limited guidance in forming expectations about the directions of these effects. For example, while social capital theory predicts more contributions from editors of high degree centrality, network game theory may suggest the opposite prediction. Further, an editor's behavior may also be constrained by the social role triggered at the position. In the following, we explore the directions of these effects in the context of the Chinese Wikipedia. Understanding how network position influences or predicts contribution behavior not only deepens our understanding of network theory; it also enables the design of interventions that could benefit both the editors and the collective product in general. For example, efforts could be taken to connect isolated editors to other parts of the collaboration network, increasing their closeness to other editors.

Although theoretical discussion leads to our speculation that editors' positions in the collaboration network influence their future contribution behavior, empirical studies about these effects of network position suffer from a reverse causal relationship and are plagued by unobservable confounding factors. The reverse causal relationship refers to the fact that network connections are developed based on editors' prior contributions. It is obvious that changes in network positions are correlated with editing behaviors. Other unobservable variables, such as an intrinsic preference for making contributions, further confound the causal interpretation of the observed correlation between network positions and contribution behaviors. Consequently, a

simple correlational analysis, such as cross-sectional regression, confounds the two directions of causal relationship. Such an analysis cannot give clear evidence of the causal relationship from network positions to contribution behaviors that would be theoretically more interesting and provide policy guidance. To establish a causal interpretation, we carefully designed our empirical analysis to leverage the panel structure of the data as well as the impact of an exogenous shock to the collaboration network structure resulting from the blockages of Chinese Wikipedia in Mainland China. We explain these empirical strategies in detail in the next section.

4. Data and Empirical Context

We collect data on revisions of pages from Chinese Wikipedia (zh.wikipedia.org). The Chinese Wikipedia is the Chinese version of the web-based encyclopedia project, Wikipedia, which was launched in January 2001. The Chinese version was launched in October 2002. Wikipedia is one of the most successful online public goods projects so far. Articles or pages on the site are written by volunteer editors. In this mass collaborative process, anyone at any time can edit any page by adding new information, revising existing paragraphs, or deleting irrelevant material. The system documents each and every edit of the articles and the data are publicly available. The Chinese Wikipedia is a perfect context to study online public goods provision on the Internet that occurs through mass collaboration [26,39].

Our full data set contains the complete editing history of the articles posted between October 2002 and February 2007. There are 196,130 articles. Each article has 4.5 different editors on average, with the highest number of 242.

4.1 Blockages of Wikipedia in the Mainland China

Besides the fit of context and the availability of data, the Chinese Wikipedia is an ideal empirical context for establishing a causal relationship. Articles on the Chinese Wikipedia are all written in Chinese. As a result, there are two major groups of editors that form the resource pool of contributors: those from Mainland China and those who are outside of Mainland China. The Chinese government implemented the Great Firewall System to censor mainlanders' access to the Internet. Wikipedia was one of the sites that got censored. The Chinese Wikipedia was blocked in Mainland China several times during the period from 2004 to 2008. For details of these blockages, please refer to Zhang and Zhu [39].

As a result of these blockages, editors from the Mainland China had no access to make additional edits to the articles on which they had collaborated. The blockages thus exogenously changed the network structure of editors who were not directly blocked. These blockages created an ideal empirical setting for us to establish causality. The blockages removed a subset of the editors (those from the Mainland China) from the collaboration network. The resulting changes in network positions of the unaffected editors were thus exogenous to other factors that could influence their contribution behavior. The implication is that, by comparing their contributions before and after the blockages, an identification of the causal relationship between network position and contribution behavior (network position → contribution behavior) can be achieved.

Specifically, we focus on the third blockage of the Chinese Wikipedia in the Mainland China that essentially became the beginning of a series of long-term blockages. It happened in October 2005. Before this blockage, two brief blockages ignited appeals from the Wikipedia community. These two short blockages served as overtures to the eventual complete blockage and informed the editors about the potential fallout. After the third blockage, the blockages were lifted for only

a few very brief periods until 2008. Thus, the third blockage is considered as the beginning of a series of long-term blockages that significantly altered the composition of Chinese Wikipedia's editors.

Figure 1 shows the impact of the third blockage. Each symbol in the figure represents the number of modifications during a particular week from IP addresses in one of the following four regions: China, Hong Kong, Taiwan, and the United States. The vertical dotted line indicates the third blockage. It can be seen that before the blockage, the editors from the Mainland China contributed the most. Immediately after the blockage, however, their participation dropped to almost zero, although there were still significant contributions from the other regions.

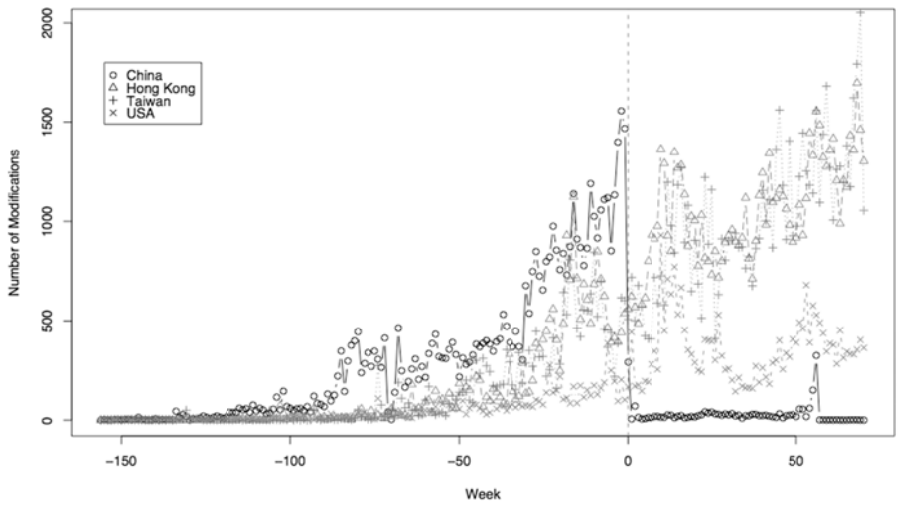


Figure 1. Impact of the Third Blockage

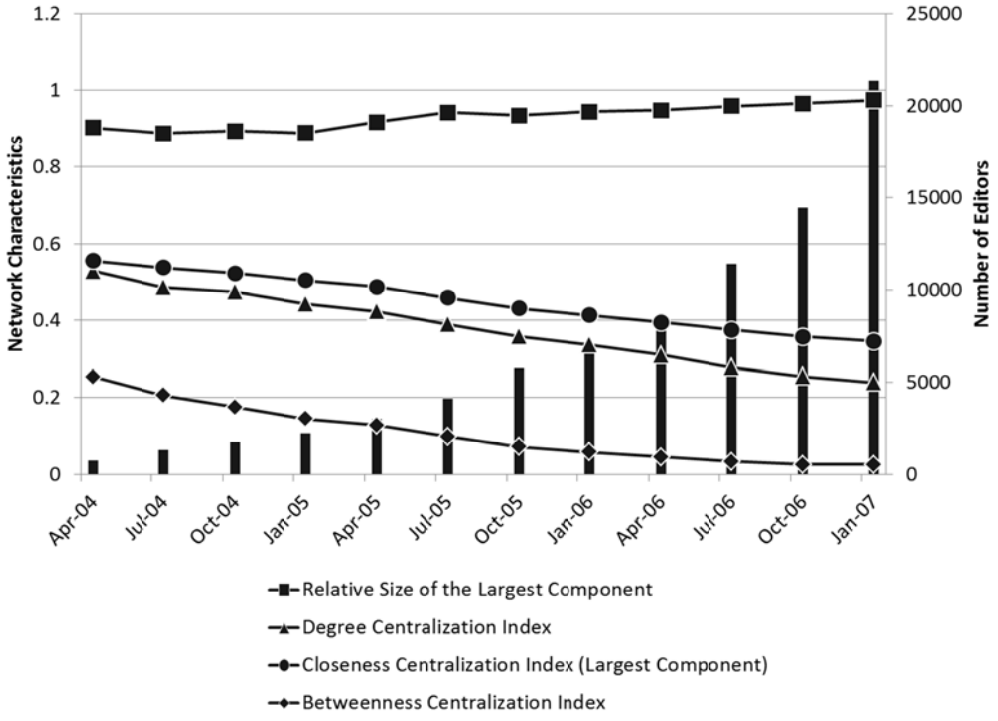
4.2 Collaboration Network on the Chinese Wikipedia

To assess the impacts of network positions on contributions, we construct the collaboration network of Wikipedia editors based on revision data. Because we want to study the effects of structural centrality measures, we focus on the collaboration network derived from the affiliation network based on co-authorship.

In composing the network, we exclude edits from unregistered editors, as well as those from robots and administrators. Two dynamic networks are constructed from the editing history: one with all the editors participating before the blockage, and the other with only those editors that were not influenced by the blockage. We follow the procedure as described in Zhang and Zhu [44] to identify the editors that were not influenced by the blockage and construct the network correspondingly.

Empirical measures characterizing the network are calculated with the social network analysis software package Pajek (Version 2.0). This software is designed to handle large networks with thousands of nodes [3,34], which is appropriate for analyzing collaboration networks in Wikipedia.

Panel A: Network Evolution (Ignoring the Blockage)



Panel B: Network after the Blockage

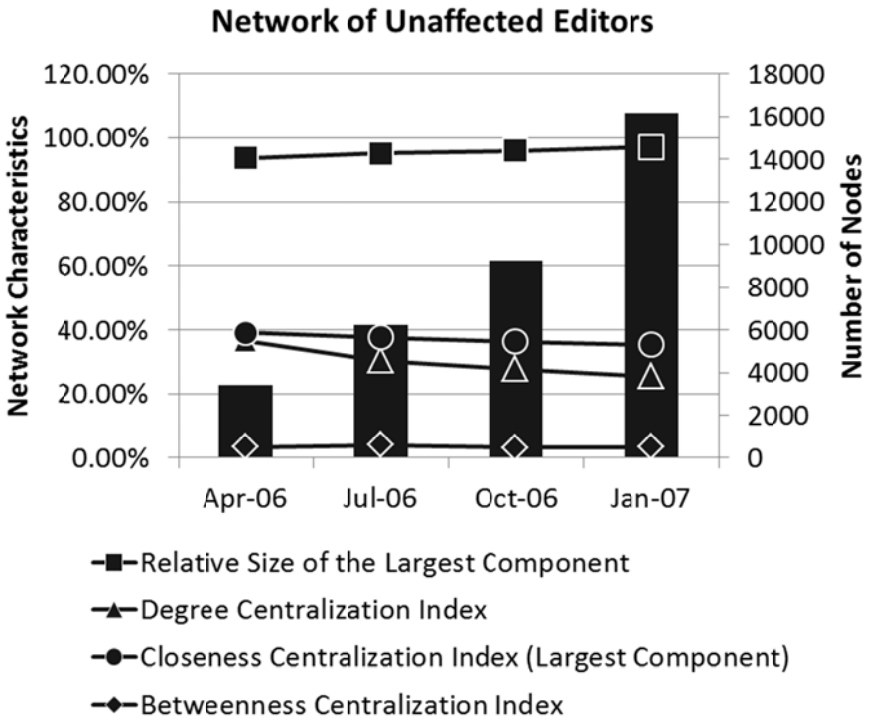


Figure 2. Size, Connectedness, and Centralization of the Collaboration Network

In addition, network centralization measures have steadily decreased over time. This indicates that collaboration has become less concentrated around a few initial editors as the numbers of editors and topics areas have increased. As new editors have gradually joined and started to contribute and collaborate, new centers of the collaboration network have emerged. On one hand, new editors with different knowledge expertise start new areas of collaborative editing. In these new territories of Wikipedia's knowledge space, new centers of collaboration network have emerged. On the other hand, new editors who actively participate in existing collaborations bridge existing collaborative relationships and become the new centers taking similar positions and roles as experienced editors. Generally, because the organization of collaboration in Wikipedia is decentralized, with no hierarchies among editors or central planners for

collaboration, the collaboration network decentralizes as the scale of the collaboration increases. Forte et al. [16] describe how governance in Wikipedia is becoming increasingly decentralized as the community grows. This necessitates a network perspective in understanding editors' behavior in Wikipedia.

5. Variables and Empirical Models

To test the effect of the individual's structural position in the collaboration network on her contribution, we calculated empirical measures to identify the contribution behavior as a function of the network centrality and other covariates (control variables). Table 1 lists the variable definitions.

Table 1. Variable Definitions

Notation	Definition
Con_{it}	The number of characters edited by editor *i* in period *t*.
$Edit_{it}$	The number of times an editor *i* edited Wikipedia articles in period *t*.
$Page_{it}$	The number of pages an editor *i* edited in period *t*.
$Ratio\_Con_{it}$	The ratio between editor *i*'s contribution to articles that she created and her contribution to articles that other editors created with respect to Con_{it}.
$Ratio\_Edit_{it}$	The ratio between editor *i*'s contribution to articles that she created and her contribution to articles that other editors created with respect to $Edit_{it}$.
$Ratio\_Page_{it}$	The ratio between editor *i*'s contribution to articles that she created and her contribution to articles that other editors created with respect to $Page_{it}$.
$Degree_{it}$	The degree centrality of editor *i* at the beginning of period *t*.
$Closeness_{it}$	The closeness centrality of editor *i* at the beginning of period *t*.
$Betweenness_{it}$	The betweenness centrality of editor *i* at the beginning of period *t*.
Age_{it}	The number of days from the first edit of editor *i* to the beginning of period *t*.
Cum_Con_{it}	Total number of characters editor *i* had edited by the beginning of period *t*.
Cum_Edit_{it}	Total number of times an editor *i* had edited Wikipedia articles by the beginning of period *t*.
Cum_Create_{it}	The number of pages an editor *i* had created by the beginning of period *t*.
Cum_Page_{it}	The number of pages an editor *i* had edited by the beginning of period *t*.

5.1 Contributions to Wikipedia

As indicated in the hypotheses, we are interested in two aspects of editors' contribution, their total effort and the allocation of their effort as individuals. We construct six measures to describe the contribution behavior for each user i in period t. First, to measure the overall contribution to the platform, we count the total number of characters edited (Con_{it}), the total number of edits ($Edit_{it}$), and the total number of different pages edited ($Page_{it}$). These three measures capture different aspects of the contribution behavior. Comparisons between these aspects reveal the concentration of an individual's contributions. To correct for non-normality, we take the logarithm of these variables in the following empirical analyses.

Second, to measure the allocation of contribution, we calculate the ratio between the contribution to one's own pages and the contribution to pages created by others. Similar to the findings in Gleave et al. [19], our preliminary data exploration suggests that there are distinctive patterns of contribution depending on whether the editor is the creator of a page (being the first to contribute to the page). Some recent studies also suggest that editors in Wikipedia take ownership of the pages that they focus on [39]. We designate the creator of a page based on the first edit of that page. We then differentiate between the editor's contributions to pages that she created and those to pages that others created. There are other ways of designating of ownership, and one may also allow "co-ownership" of a page. Wikipedia, however, does not officially designate any kind of ownership to its pages. As a result, our objective of an ownership definition is to capture the fact that an editor is more psychologically committed to a group of articles, while avoiding a rhetorical definition that relies too much on contribution patterns. Based on our definition of ownership, we calculate the ratio between the contribution to editors' own pages and the contribution to others' created pages with respect to the number of characters

edited ($Ratio\_Con_{it}$), the total number of edits ($Ratio\_Edit_{it}$), and the number of pages edited ($Ratio\_Page_{it}$). Editors who focus mainly on their own pages (having a higher value of the ratios) act as soloists and maintain the pages for which they have expertise. In contrast, editors who contribute more to the pages others created (having a lower value of the ratios) are collaborators who value the overall quality of articles in Wikipedia. Although both types of contributors are important to the sustainability of Wikipedia, the respective roles taken by the individual contributors have different implications. Content creators are of fundamental importance to enriching the coverage of the online encyclopedia, while editors engaging in editorial work help improve the accessibility of the article to a general audience and are, therefore, critical to maintaining a high quality of presentation.

5.2 Network Centralities

As discussed earlier, degree centrality ($Degree_{it}$) is defined as the focal editor's number of collaborations normalized by the size of the network. Closeness centrality ($Closeness_{it}$) is defined as the inverse of the average distance between the focal editor and others in the network. Betweenness centrality ($Betweenness_{it}$) is defined as the number of times that the focal editor lies on the shortest path between a pair of other editors divided by the number of all possible pairs of editors. All measures are calculated with Pajek.

The blockage of the Chinese Wikipedia in the Mainland China removed some editors from the collaboration network. As a result, centrality measures in the periods after the blockage are calculated based on the network composed of editors that were unaffected by the blockage.

As collaboration networks are constructed based on co-editing behavior, an editor’s contribution behavior also influences her network position. We use a temporal sequence of

events to alleviate the concern for any reverse causal relationship and calculate centrality measures at the beginning of each period. It is common to use a temporal sequence to ensure the direction of the casual chain in a situation where there is a loop of causal links [40].

5.3 Other Controls

We measure the experience of an editor, *i*, by the number of periods that the editor had been active (Age_{it}) before period *t*. To control for the influence of previous contributions, we calculate the cumulative number of characters edited (Cum_Con_{it}), the cumulative number of edits (Cum_Edit_{it}), the cumulative number of pages edited (Cum_Page_{it}), as well as the cumulative number of pages created (Cum_Create_{it}) before period *t*.

Summary statistics of these variables are reported in Table 2. We focus on the period between 2004 and 2007 to allow for initial diffusion. In total, we have 81,606 observations over 12 periods (quarters). On average, editors in our sample before the blockage (after the blockage) made 13 (17) revisions on 4.90 (6.30) pages and modified over 3,000 (4,300) characters per period. There is significant variation in contributions across both periods and individuals (as reflected by the standard deviation). The average degree centrality is low, which indicates that the network is relatively sparse; editors collaborate with only a few others, although the total number of contributors is large. The average distance between any two editors is short, however, as indicated by the closeness centrality measures. These characteristics resemble small-world networks [42]. It has been shown that such networks, which are commonly observed in social contexts, are robust to the removal of nodes.

Table 2. Summary Statistics

Panel A - Before Blockage

Variables		Mean (Std. Dev.)	1	2	3	4	5	6	7	8	9	10	11	12	13
1	Con_{it}	3.03×10^{3} (2.16×10^{4})													
2	$Edit_{it}$	13.39 (78.12)	0.76												
3	$Page_{it}$	4.90 (28.82)	0.62	0.88											
4	$Ratio\_Con_{it}$	0.11 (0.28)	0.17	0.19	0.17										
5	$Ratio\_Edit_{it}$	0.32 (0.39)	0.23	0.29	0.28	0.50									
6	$Ratio\_Page_{it}$	0.28 (0.35)	0.26	0.32	0.32	0.51	0.98								
7	$Degree_{it}$	7.97×10^{-3} (0.02)	0.19	0.24	0.27	0.03	0.08	0.10							
8	$Closeness_{it}$	0.24 (0.18)	0.01	0.01	0.02	-0.25	-0.52	-0.49	0.37						
9	$Betweenness_{it}$	4.25×10^{-4} (5.74×10^{-3})	0.15	0.19	0.24	0.03	0.08	0.09	0.79	0.11					
10	Age_{it}	2.02 (1.99)	-0.05	-0.06	-0.05	-0.26	-0.54	-0.52	0.24	0.62	0.09				
11	Cum_Con_{it}	9.09×10^{3} (6.11×10^{5})	0.53	0.46	0.42	0.11	0.16	0.19	0.69	0.13	0.73	0.10			
12	Cum_Edit_{it}	24.14 (159.90)	0.24	0.29	0.29	0.07	0.12	0.15	0.76	0.18	0.72	0.14	0.85		
13	Cum_Create_{it}	2.91 (19.15)	0.21	0.24	0.25	0.10	0.11	0.13	0.64	0.17	0.58	0.13	0.73	0.87	
14	Cum_Page_{it}	8.91 (54.32)	0.22	0.26	0.29	0.06	0.12	0.15	0.80	0.19	0.74	0.16	0.80	0.95	0.83

Panel B - After Blockage

Variables		Mean (Std. Dev.)	1	2	3	4	5	6	7	8	9	10	11	12	13
1	Con_{it}	6.47×10^{3} (6.87×10^{4})													
2	$Edit_{it}$	25.12 (176.64)	0.56												
3	$Page_{it}$	9.28 (82.26)	0.47	0.87											
4	$Ratio\_Con_{it}$	0.13 (0.31)	0.08	0.12	0.08										
5	$Ratio\_Edit_{it}$	0.47 (0.40)	0.12	0.18	0.14	0.39									
6	$Ratio\_Page_{it}$	0.41 (0.35)	0.14	0.22	0.18	0.40	0.97								
7	$Degree_{it}$	8.57×10^{-3} (0.028)	0.16	0.28	0.26	0.06	0.11	0.16							
8	$Closeness_{it}$	0.22 (0.20)	0.05	0.09	0.08	-0.15	-0.45	-0.39	0.40						
9	$Betweenness_{it}$	1.46×10^{-4} (1.13×10^{-3})	0.18	0.34	0.35	0.05	0.11	0.14	0.75	0.20					
10	Age_{it}	1.65 (2.26)	0.03	0.05	0.05	-0.09	-0.30	-0.26	0.43	0.66	0.23				
11	Cum_Con_{it}	2.07×10^{4} (1.50×10^{5})	0.70	0.60	0.44	0.11	0.13	0.16	0.43	0.15	0.49	0.16			
12	Cum_Edit_{it}	59.72 (374.99)	0.32	0.56	0.42	0.09	0.13	0.16	0.62	0.22	0.72	0.25	0.78		
13	Cum_Create_{it}	5.46 (45.50)	0.25	0.41	0.41	0.09	0.09	0.12	0.46	0.16	0.55	0.19	0.56	0.74	
14	Cum_Page_{it}	18.85 (112.52)	0.25	0.46	0.45	0.08	0.12	0.16	0.69	0.23	0.85	0.27	0.61	0.86	0.84

Comparing Panel A (Before Blockage) to Panel B (After Blockage), after the blockage removed Mainland editors from the network, the average degree centrality increases. This indicates that the impact of the blockage on individual editors' direct collaborators is smaller than the impact on the whole network. In contrast, the average closeness centrality decreases after the blockage, which suggests that the removal of editors forces network paths between unaffected editors to take longer routes. The average betweenness centrality of the editors also decreases, suggesting that editors are not bridging more connections as a result of the removal of editors. With respect to the contribution behavior, on average, the unaffected editors contribute more after the blockage. They also tend to focus more on articles that they created. Figure 3 further illustrates the relative change, with respect to the means of the variables in the pre-blockage period, in average centrality measures and behavior measures.

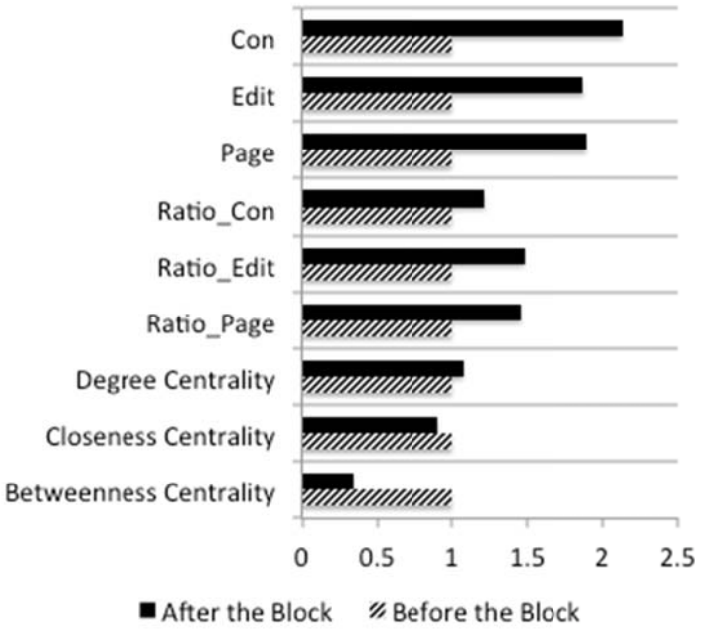


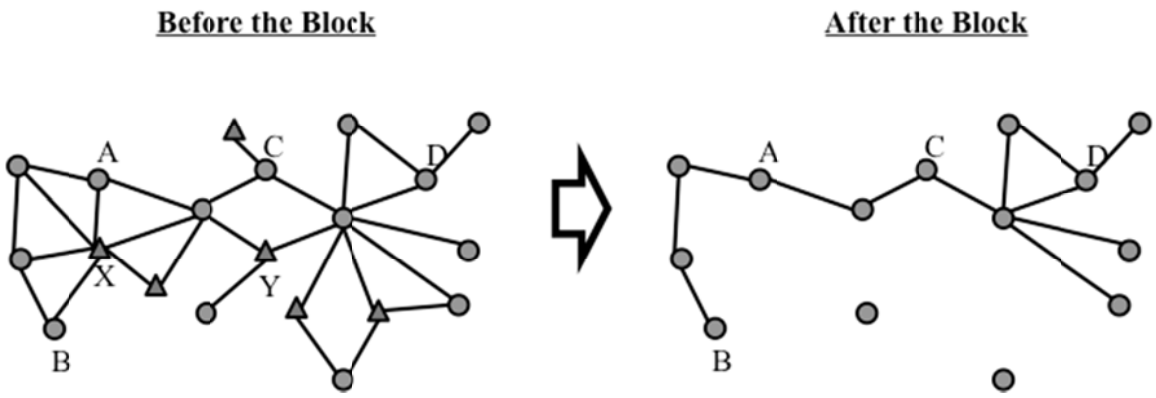
Figure 3. Comparison of Contribution Behavior and Centrality Measures

In both panels of Table 2, degree centrality and betweenness centrality show strong correlations with the total contribution (cumulative measures), while closeness centrality does not. For example, the correlation between Cum_Con_{it} and $Degree_{it}$ ($Betweenness_{it}$) is 0.69 (0.73) in Panel A (Before Blockage) while the correlation between Cum_Con_{it} and $Closeness_{it}$ is only 0.13 in the same table. This indicates that editors who have many direct collaboration ties or take a critical role in bridging connections between different parts of the network also contribute more. We also find that editors of high degree centrality tend to have high betweenness centrality. The correlation between $Degree_{it}$ and $Betweenness_{it}$ is 0.79 in Panel A (0.75 in Panel B). Closeness centrality shows a strong negative correlation with the ratio measures. For example, the correlation between $Ratio\_Edit_{it}$ and $Closeness_{it}$ is -0.52 in Panel A (-0.45 in Panel B), indicating that editors who concentrate on self-created articles are poorly connected in the collaboration network.

5.4 Natural Experiment Model

To establish a causal interpretation of our empirical results, we took advantage of the blockage of Mainland Chinese editors, which created a natural experiment about the effect of network position on contribution behavior. In the model, we compare periodic contributions right before and after the blockage. As already explained, the removal of editors caused by a regulatory blockage generates exogenous heterogeneous variations in the positions of individual actors that are not directly affected by the blockage. Hence, comparing their behavior before and after the blockage, we search for evidence about the effect of network position on contribution behavior. The formation of endogenous ties for each individual in the network is not an empirical concern in this natural experiment design, because changes in network positions that are pursued here do not relate to editors' behavior or other characteristics. As a result, a comparison of pre- and post-

blockage behaviors of unaffected actors provides empirical evidence as to whether there is indeed a causal relationship between the variables in question.



Notes:
In the figure, both triangle nodes and circle nodes denote editors. The links between nodes represent collaborative relationships. The triangle nodes denote editors that were removed by the blockage. Removing a node from a network also removes all links that connect to that node.

Figure 4. Illustration: Effect of the Blockage on the Collaboration Network

Figure 4 illustrates the effect of the blockage on the network of collaborators. In the figure, the round nodes represent actors who were not influenced by the blockage, and the triangle nodes represent those who were. It is apparent from the figure that an exogenous removal of some actors in the network unevenly influences those who remain, depending on the network positions before the blockage. For example, after removing a local hub (node X), node A becomes very important in bridging the communication between the two sides of the network and thus gets a high betweenness centrality after the blockage. At the same time, node B moves farther away from the right side of the network (resulting in a decrease in its closeness centrality). As another example, the removal of node Y makes its rival node C more central in terms of betweenness but not in the sense of closeness. Meanwhile, the blockage does not have much impact on node D because most close friends of D are not influenced.

We estimate the following model:

$$\{\Delta Con_i, \Delta Edit_i, \Delta Page_i, \Delta Ratio\_Con_i, \Delta Ratio\_Edit_i, \Delta Ratio\_Page_i\} = \alpha + \beta_1 \times \Delta Degree_i + \beta_2 \times \Delta Closeness_i + \beta_3 \times \Delta Betweenness_i + \varepsilon_i. \quad (1)$$

As discussed, we consider the six measures of the contribution behavior, including the total edits, number of edits, and number of pages edited, as well as the ratio between edits to own pages and total edits. In the natural experiment model above, α captures the average impact on the contribution behavior brought by the shock [44], while β_i's capture the effects of network position. The network positions were measured at the beginning of each period to control for the direction of the causal relationship.

5.5 Panel Regression Model

To corroborate our findings from the above model based on the natural experiment, we also conduct a panel regression analysis with data before the blockage (periods 1-7). Again, the dependent variables are the measures of contribution behavior as defined previously, and the network-centrality measures are used as independent variables. In the natural experiment model, individual-level factors are removed by differencing. Controlling for individual-level unobserved effects is achieved in the panel model by introducing fixed-effects (within estimator). Compared to the natural experiment model, the panel regression covers a longer period. It is thus more tolerant to unobserved changes that could have taken place together with the blockage that we leverage in the natural experiment design. In summary, although the fixed-effect approach provides weaker controls for the unobservable individual-level factors, it is more robust against random shocks. Confidence in the robustness of the findings as well as additional insights could be gained by comparing the results from these two empirical methods (and the two samples).

The following models are estimated.

$$\{Con_{it}, Edit_{it}, Page_{it}, Ratio\_Con_{it}, Ratio\_Edit_{it}, Ratio\_Page_{it}\} = \alpha_i + \beta_1 \times Degree_{it} + \beta_2 \times Closeness_{it} + \beta_3 \times Betweenness_{it} + \gamma' Z_{it} + \varepsilon_{it}. \quad (2)$$

Here, α_i's are the individual fixed effects representing unobserved editor characteristics. Z_{it} includes the control variables (cumulative contribution and editors' experience).

6. Results

6.1 Natural Experiment Model

Estimation results of the natural experiment model are reported in Table 3. The effect of degree centrality is negative on total contribution. The marginal effects of degree centrality on the total contribution (Model 1), number of edits (Model 2), and number of pages edited (Model 3) are -3.44 ($p < 0.1$), -4.13 ($p < 0.01$), and -3.95 ($p < 0.01$), respectively. When an editor gets more collaborators, she decreases her own effort and free-rides others' efforts on the articles. It appears that the effect of degree centrality on effort allocation is not significant (in Models 4, 5, 6), which means there is no statistically significant shift in effort allocation associated with changes in degree centrality. Our finding with respect to degree centrality indicates that the editor decreases her contribution to both her own articles and articles created by others when she connects to more collaborators in Wikipedia.[6] This finding confirms the prediction of network-game models

[6] Wikipedia articles offers more informational as opposed to "creative" content. It becomes hard to add more content when it nears a saturation point. This may also lead to the decrease in an editor's total contribution. As our later panel analysis suggests, an editor who has higher degree centrality tends to focus on articles she created and free-ride others for the other articles. This change in focus of contribution combined with accumulation of contents leads to a further decrease in total contribution. As for the natural experiment here, article status changes are minor. Consequently, we only find that editors who increase more in degree centrality tend to contribute less, indicating a general tendency for free-riding.

indicating that individuals may free-ride the efforts of their peers and apply less effort when there are more collaborators. Comparing coefficient estimations from Models 1, 2, and 3 further reveals that as an editor's degree centrality increases, her contribution per edits as well as edits per page all increase, indicating that she is less likely to make minor changes.

Table 3. Natural Experiment Model

DV:	Model 1 ΔCon_{it}	Model 2 $\Delta Edit_{it}$	Model 3 $\Delta Page_{it}$	Model 4 $\Delta Ratio\_Con_{it}$	Model 5 $\Delta Ratio\_Edit_{it}$	Model 6 $\Delta Ratio\_Page_{it}$
$\Delta Degree_{it}$	-3.440*	-4.133***	-3.953***	-0.00564	0.242	-0.0380
	(1.760)	(0.769)	(0.608)	(0.159)	(0.198)	(0.178)
$\Delta Closeness_{it}$	-6.568***	-2.442***	-1.602***	-0.163***	-0.815***	-0.728***
	(0.545)	(0.238)	(0.188)	(0.0491)	(0.0614)	(0.0551)
$\Delta Betweenness_{it}$	122.8***	34.77**	36.35***	-4.941	12.94***	13.35***
	(40.49)	(17.68)	(13.99)	(3.648)	(4.562)	(4.098)
Observations	1,599	1,599	1,599	1,599	1,599	1,599
R-squared	0.088	0.080	0.070	0.009	0.101	0.101

Notes:
In this model, we regress the change in contribution level on the change in network centralities as a result of the blockage of editors from Mainland China.
Standard errors are shown in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Closeness centrality has consistent negative effects on both the total contribution and allocation of efforts to self-created pages. The marginal effects of closeness centrality are -6.57 ($p < 0.01$), -2.44 ($p < 0.01$), and -1.60 ($p < 0.01$) in Models 1, 2 and 3, respectively. As an editor gets closer to the others, she becomes more prominent in the collaboration network. The social capital accumulated in the collaboration network makes it more likely that she gets help from others, which gives her the incentive to free-ride and decrease her own total contribution. Meanwhile, as she gets nearer to the other editors in the network, it becomes more likely that she finds articles not created by her, towards which she can make additional contributions. This means that she concentrates less on her own articles, which is supported by the estimates of the marginal effects of closeness centrality on $Ratio\_Con_{it}$ (-0.16, $p < 0.01$), $Ratio\_Edit_{it}$ (-0.82, $p <$

0.01), and $Ratio\_Page_{it}$ (-0.72, $p < 0.01$). From a role-taking perspective, it appears that an editor of high closeness centrality takes the collaborative role as a facilitator. The structural capital accumulated through closeness reduces her total contribution. Meanwhile, because of the substitutive effort exerted by the others, an editor who takes this role extends her contribution to articles created by others and helps improve the overall quality. In doing so, it appears that she also makes fewer changes per edited page, which can be seen by comparing the estimates from Model 1 (-6.57) and Model 2 (-1.60).

Betweenness centrality has positive effects on the total effort contribution as well as the proportion of effort allocated to articles that are created by the focal editor. With respect to total contribution measures, the marginal effects of $Betweenness_{it}$ on Con_{it}, $Edit_{it}$, and $Page_{it}$ are 122.8 ($p < 0.01$), 34.77 ($p < 0.05$), and 36.35 ($p < 0.01$). Its marginal effects on allocation measures are -4.94 (not significant) in Model 4 ($Ratio\_Con_{it}$), -12.94 ($p < 0.01$) in Model 5 ($Ratio\_Edit_{it}$), and -13.35 ($p < 0.01$) in Model 6 ($Ratio\_Page_{it}$). When an editor becomes structurally critical in bridging collaborative connections, she increases her contribution, while at the same time focusing more on her own pages. One interpretation of this finding is that when an editor is positioned on structural holes, she will be exposed to diverse sources of information, which leads her to keep improving the quality of the pages that she created and find new areas that deserve coverage.

To summarize, the estimation results of the natural experiment model support the effect of network position on both total contribution in the Total Contribution Hypothesis (H1) and on effort allocation in the Effort Allocation Hypothesis (H2). Comparing the estimates of different centrality measure reveals that each centrality measure has a unique influence pattern, which confirms our prediction in the Differential Effect of Centrality Measures Hypothesis (H3).

6.2 Panel Regression Model

Table 4. Panel Regression Model

DV:	Model 1 Con_{it}	Model 2 $Edit_{it}$	Model 3 $Page_{it}$	Model 4 $Ratio\_Con_{it}$	Model 5 $Ratio\_Edit_{it}$	Model 6 $Ratio\_Page_{it}$
$Degree_{it}$	3.147	-8.278***	-8.743***	0.780**	2.017***	0.898***
	(2.679)	(0.947)	(0.701)	(0.312)	(0.312)	(0.271)
$Closeness_{it}$	-9.944***	-2.987***	-1.954***	-0.346***	-1.346***	-1.144***
	(0.148)	(0.0521)	(0.0386)	(0.0172)	(0.0172)	(0.0149)
$Betweenness_{it}$	56.20***	50.54***	46.98***	0.939	2.608**	4.651***
	(9.335)	(3.298)	(2.443)	(1.089)	(1.088)	(0.944)
Age_{it}	-0.260***	-0.0714***	-0.0599***	-0.0228***	-0.0347***	-0.0314***
	(0.0124)	(0.00437)	(0.00324)	(0.00144)	(0.00144)	(0.00125)
Other Control	Yes	Yes	Yes	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of Editors	5,750	5,750	5,750	5,750	5,750	5,750
Observations	18,937	18,937	18,937	18,937	18,937	18,937
R-squared	0.488	0.471	0.466	0.122	0.539	0.542

Notes:
This table reports estimation results for the periods before the blockage based on a panel regression with editor fixed effects. We include cumulative measures of contribution activities as controls. Within-effect estimation is implemented to control for individual fixed effects, and within R-squared is reported.
Standard errors are shown in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

The results of the panel regression model are reported in Table 4. These results convey a very similar picture of the effects found in the natural experiment model, except we now find that degree centrality has a significant positive effect on effort allocation measures. Specifically, the marginal effects of $Degree_{it}$ on $Ratio\_Con_{it}$, $Ratio\_Edit_{it}$, and $Ratio\_Page_{it}$ are 0.78 ($p < 0.05$), 2.02 ($p < 0.01$), and 0.90 ($p < 0.01$)), respectively. This indicates that having more collaborators induces an editor to focus more on her own articles. An editor is more likely to allocate the contribution to articles for which she has expertise and wait for or free-ride on others to contribute to the other articles, about which she has less expertise, as she gets more collaboration. Through broader collaboration, an editor will take the role of a content expert who

concentrates on her own area of expertise. To support this explanation, we further include a variable, $Expertise_{it}$, that measures average edits per page edited ($Expertise_{it} = Cum\_Con_{it} / Cum\_Page_{it}$) in the model for effort allocation. Estimation results suggest $Expertise_{it}$ positively correlates with effort allocation. In Table 5, all coefficient estimates for $Expertise_{it}$ in Models 1, 2, and 3 are significant at 0.01. Confirming our interpretation that degree centrality is associated with expert status, after controlling for $Expertise_{it}$, the effect of degree centrality on effort allocation turns out to be insignificant or negative.

Table 5. Effect of Expertise

DV:	Model 1 $Ratio\_Con_{it}$	Model 2 $Ratio\_Edit_{it}$	Model 3 $Ratio\_Page_{it}$
$Expertise_{it}$	2.06e-06***	2.73e-06***	2.67e-06***
	(3.77e-07)	(4.74e-07)	(4.20e-07)
$Degree_{it}$	0.454	-1.659***	-1.749***
	(0.294)	(0.370)	(0.327)
$Closeness_{it}$	-0.101***	-0.0587	-0.0490
	(0.0336)	(0.0423)	(0.0374)
$Betweenness_{it}$	1.713**	7.797***	7.931***
	(0.837)	(1.054)	(0.933)
Other Control	Yes	Yes	Yes
Fixed Effects	Yes	Yes	Yes
Number of Editors	4,113	4,113	4,113
Observations	13,461	13,461	13,461
R-squared	0.022	0.058	0.068

Notes:
Standard errors are shown in parentheses.
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

We also find that an editor's age correlates negatively with her total contribution. From Table 4, the marginal effects of Age_{it} on total contribution measures are -0.26 ($p < 0.01$) in Model 1, -0.07 ($p < 0.01$) in Model 2, and -0.06 ($p < 0.01$) in Model 3. Growing in experience also negatively correlates with the concentration on own articles. The estimates of the marginal effects are -0.02

($p < 0.01$), -0.03 ($p < 0.01$), and -0.03 ($p < 0.01$) in Models 4, 5 and 6, respectively. This suggests that the editors tend to diversify their contributions over time.

7. Discussion

In this paper, we have investigated the relationship between the positions of Wikipedia editors in the collaboration network and their contribution behaviors. Based on our review of recent developments in network games and the literature on role-taking behavior, we proposed that centralities affect both an editor's decision about total contribution and the allocation of her effort in the Total Contribution Hypothesis (H1), and in the Effort Allocation Hypothesis (H2), respectively. Further, we proposed that different aspects of network positions, as represented by different centrality measures, have differential effects, indicating that they represent different role-taking behaviors, the Differential Effect of Centrality Measures Hypothesis (H3).

Our empirical context, the Chinese Wikipedia, provides an ideal opportunity for addressing a critical challenge that researchers testing the effect of an editor's position in the collaboration network on her contribution behavior face: contribution behavior naturally leads to the formation of network structure. Combined with other unobserved editor characteristics, changes in an editor's network position may be endogenous in cross-sectional regressions. As a result, an empirical identification of the causal relationship from network positions to contribution behavior could be contaminated. To provide clear evidence of this causal relationship, researchers need to find exogenous changes to the network positions while controlling for the unobserved individual effect. .

The blocks imposed on the Chinese Wikipedia by the Chinese government removed the Mainland editors from the network and thus created such an exogenous shock to the network

positions of unaffected editors (from Hong Kong, Taiwan, Singapore, the U.S., and other regions). Changes in network positions of the unaffected editors in this natural experiment were exogenous to their own editing behavior, and individual fixed effects are easily taken away by differencing. As a result, this natural experiment was capable of identifying the causal relationship from an individual editor's network position to her contribution behavior. We further supplemented the natural experiment model with a panel regression model in order to test the robustness of our findings from the natural experiment.

Our estimation results from the natural experiment and panel data analysis were consistent and indicate that there is a causal relationship between an editor's network position and her contribution behavior, supporting the Total Contribution Hypothesis (H1) and the Effort Allocation Hypothesis (H2). Further, our results revealed the complexity of behavioral implications of centralities, supporting the Differential Effect of Centrality Measures Hypothesis (H3). We found that each type of centrality is associated with a particular editing role, in ways that are consistent with network game theories, as follows.

First, degree centrality represents the number of direct collaboration relationships. Our results suggest that degree centrality is negatively associated with total contribution. Meanwhile, editors with higher degree centrality focused more on the articles that they have created. For example, from Table 4, the marginal effect of degree centrality on total number of edits ($Edit_{it}$) was -8.28 ($p < 0.01$), while its effect on the proportion of edits allocated to own articles was 2.02 ($p < 0.01$). Consistent with the prediction of network game theory, editors tended to free-ride on others' contributions when they have more collaborators. The potential to free-ride made them focus more on the articles they created and about which they have expertise. They tended to make more major edits as oppose to minor ones. This assertion is supported based on comparing

estimates in Model 1 and Model 2 in Table 3. In general, degree centrality was associated with the role of a content expert who makes major and focused contributions to Wikipedia.

Second, closeness centrality represents the average distance between an editor and the others in the collaboration network. Our estimation results suggest that an increase in closeness centrality reduces both an editor's total contribution and the focus on articles that she created. When an editor has high closeness to others in the collaboration network, she is more likely to receive help from a diverse group of people on the articles, which reduces her total contribution as a result of having a good prospect to free-ride. For example, the marginal effect of closeness centrality on total contribution was -6.57 ($p < 0.01$) in Model 1 of Table 3. Unlike in the case of degree centrality, it appears that editors reciprocated when they receive contributions from other editors on articles that they maintained by extending their contributions to other articles. In other words, editors seem to have focused less on their own articles as a result of an increase in closeness centrality. For example, the marginal effect of closeness centrality on the proportion of effort allocated to own articles was -0.16 ($p < 0.01$) in Model 4 of Table 3. Our findings suggest that closeness centrality is associated with the role of a generalist who contributes to a diversity of articles while receiving help from others on articles that she created.

Finally, betweenness centrality represents the importance of an editor in bridging the collaboration ties in the network. Our estimates of the marginal effect of betweenness centrality in both Table 3 and Table 4 were positive and significant in general. An editor of high betweenness centrality is located on the structural hole of the collaboration network, and thus has access to diverse parts of the network. Our result suggests that an increase in an editor's betweenness centrality is associated with an increase in her total contribution. For example, the marginal effect of $Betweenness_{it}$ on Con_{it} is 122.8 ($p < 0.01$) in Model 1 of Table 3. It also

suggests a concentration on articles that she created, as shown by the marginal effect of $Betweenness_{it}$ on $Ratio\_Page_{it}$ of 2.61 ($p < 0.01$) in Model 3 of Table 3. Having access to a diversity of editors enables learning, which contributes to the strengthening of expertise. It is also more likely for an editor with high betweenness centrality to find new areas that need coverage and create pages on those topics. Betweenness centrality is thus associated with the role of an expender who obtains expertise through learning from a diversity of editors and finds new areas for making contributions.

To our knowledge, this paper is among the first to document the differential effects of network positions, as described by different centrality measures, on contributions to online public goods. Through a natural experiment and a panel data set, we found clear evidence of the differential effects of network position on contribution behavior, which confirms our theoretical predictions based on network game theory and the social role theories. Our findings further call for an extension in network theory for mass collaboration platforms on the Internet and within organizations. Although the theories of network games and the structural determination of social roles are informative for understanding mass collaboration in general, more specificity is needed for network theories about the organization of online collaborative projects. For example, our study has revealed that free-riding behavior not only reduces an editor's total effort, but also shifts her focus to articles that she created. In other contexts, such as OSS projects, where a project's success is more dependent on the contributions, competition for resources between different projects should play an even more prominent role.

This paper has important implications for system designers and managers trying to take advantage of wiki platforms to harness the power of the crowd [28]. Given the voluntary nature of contributions, contributors value the social interactions. Consequently, the social structure of

collaboration is critical to managing these sites. For site managers of Wikipedia and the like, understanding the structural implications of the collaboration network enables informed interventions on voluntary contributions. For example, key players in the network could be identified based on their network positions. Site managers not only can target promotion activities on increasing users' contributions, but they also can try to alter their positions in the collaboration network, which may eventually help them shape the right role perception and create sustaining contributions. As our knowledge about the structural implication of networked collaboration extends, more innovative programs and site designs could be implemented to release a network's potential.

This paper is not without limitations. First, because of the current status of the theoretical literature and the limitations of our context, it is hard to pin down the exact mechanisms through which the network functions and whether the significance of each mechanism changes in other platforms. In general, network ties could be functioning as information channels or social bonds [7]. A position in the collaboration network is associated with both structural capital (access to resources) and structural constraints. Our study suggests that different aspects of a position may be associated with different mechanisms, and thus have differential implications on total contribution and effort allocation. Future research could extend the discussion and investigate the implication of network resources and network constraints in the emergence and stabilization of collaborative roles in online collaborative projects.

In this paper, we have focused on centrality measures. Although individuals' centralities in collaboration networks are important structural properties, other structural characteristics, such as the network's clustering and transitivity, also deserve investigation. For example, in this study we normalized the centrality measures with respect to the size of the whole network. In a sparse

network, however, it may make more sense to normalize with respect to some local clusters. Furthermore, theories about social capital, group identification, and social norms may provide additional insights about the collective and collaborative processes on mass collaboration platforms.

It is also important to notice that editors may have different levels of *central sensitivity*. While some editors have high social awareness and quickly adapt to new roles that are prescribed by their social position, others may be insensitive to the network environment and thus be slower in adopting new behavior patterns. In the current study, we focused our attention on the identification of the average effect of network position. Future studies could allow heterogeneity in editor sensitivity to social position and study its antecedents and consequences.

Last, although we choose the Chinese Wikipedia as our empirical context, it is imperative to extend the analysis to other online public goods contexts, such as Wikipedia in other languages, OSS, and content-sharing platforms. Extending the study of network structure to other contexts and comparing the results may reveal additional insights about the organization of mass online collaboration facilitated by modern information technologies.

Acknowledgments: We thank Rob Kauffman, Thomas Weber, DJ Wu, Debabrata Dey, Atanu Lahiri, Ravi Mantena, and three anonymous referees for their generous input to the development of this paper. Michael Zhang thanks the Hong Kong Research Grants Council (RGC) for generous support. All errors are our own.

REFERENCES

1. Arazy, O., Nov, O., Patterson, R., and Yeo, L. Information quality in Wikipedia: the effects of group composition and task conflict. *Journal of Management Information Systems*, *27*, 4 (Spring 2011), 71-98.

2. Ballester, C., Calvó-Armengol, A., and Zenou, Y. Who's who in networks. Wanted: the hey player. *Econometrica*, 74, 5 (September 2006), 1403-1417.

3. Batagelj, V. and Mrvar, A. Pajek – program for large network analysis. *Connections*, 21, 2 (1998), 1-11.

4. Bonacich, P. Power and centrality: a family of measures. *American Journal of Sociology*, 92, 5 (March 1987), 1170-1182.

5. Borgatti, S.P. and Everet, M.G. Notions of position in social network analysis. *Sociological Methodology*, 22, (1992), 1-35.

6. Borgatti, S.P. and Foster, P.C. The network paradigm in organizational research: a review and typology. *Journal of Management*, 29, 6 (December 2003), 991-1013.

7. Borgatti, S.P. and Halgin, D.S. On network theory. *Organization Science*, 22, 5 (September/October 2011), 1168-1181.

8. Bramoullé, Y. and Kranton, R. Public goods in networks. *Journal of Economic Theory*, 135, 1 (July 2007), 478-494.

9. Burt, R.S. Positions in networks. *Social Forces*, 55, 1 (September 1976), 93-122.

10. Burt, R.S. *Structural Holes: The Social Structure of Competition*. Harvard University Press, Boston, MA, 1995.

11. Burt, R.S. Structural holes and good ideas. *American Journal of Sociology*, 110, 2 (September 2004), 349-399.

12. Chen, R. and Chen, Y. The potential of social identity for equilibrium selection. *American Economic Review*, 101, 6 (October 2011), 2562-2589.

13. Erickson, B.H. The relational basis of attitudes. In B. Wellman (eds.) *Social Structures: A Network Approach*. Cambridge University Press, Cambridge, UK, 1988, pp. 99-121.

14. Fang, Y. and Neufeld, D. Understanding sustained participation in open source software projects. *Journal of Management Information Systems*, 25, 4 (Spring 2009), 9-50.

15. Faust, K. Comparison of methods for positional analysis: structural and general equivalences. *Social Networks*, 10, 4 (December 1988), 313-341.

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

17. Forte, A., Larco, V., and Bruckman, A. Decentralization in Wikipedia governance. *Journal of Management Information Systems*, 26, 1 (Summer 2009), 49-72.

18. Galeotti, A., Goyal, S., Jackson, M.O., Vega-Redondo, F., and Yariv, L. Network games. *Review of Economic Studies* 77, 1 (January 2010), 218-244.

19. Gleave, E., Welser, H.T., Lento, T.M., and Smith, M.A. A conceptual and operational definition of "social role" in online community. In R. Sprague (ed.), *Proceedings of the 42nd Hawaii International Conference on System Sciences,* January 2009, IEEE Computer Society Press, Washington, DC, 2009.

20. Granovetter, M. Economic action and social structure: the problem of embeddedness. *American Journal of Sociology*, 91, 3 (November 1985), 481-510.

21. Granovetter, M. The impact of social structure on economic outcomes. *Journal of Economic Perspectives*, 19, 1 (January 2005), 33-50.

22. Grewal, R., Lilien, G.L., and Mallapragada, G. Location, location, location: how network embeddedness affects project success in open source systems. *Management Science*, 52, 7 (July 2006), 1043-1056.

23. Hahn, J., Moon, J.Y., and Zhang, C. Emergence of new project teams from open source software developer networks: impact of prior collaboration ties. *Information Systems Research*, 19, 3 (September 2008), 369-391.

24. Homans, G.C. *The Nature of Social Science*. Harcourt, Brace and World, New York, NY, 1967.

25. Ibarra, H. and Andrews, S.B. Power, social influence, and sense making: effects of network centrality and proximity on employee perceptions. *Administrative Science Quarterly*, 38, 2 (June 1993), 277-303.

26. Ibarra, H., Kilduff, M., and Tsai, W. Zooming in and out: connecting individuals and collectivities at the frontiers of organizational network research. *Organization Science*, 16, 4 (August 2005), 359–371.

27. Jackson, M.O. *Social and Economic Networks*. Princeton University Press, Princeton, NJ, 2008.

28. Kane, G.C. and Fichman, R.G. The shoemaker's children: using wikis for information systems teaching, research, and publication. *MIS Quarterly*, 33, 1 (March 2009), 1-17.

29. von Krogh, G. and von Hippel, E. The promise of research on open source software. *Management Science*, 52, 7 (July 2006), 975-983.

30. Lakhani, K.R. and von Hippel, E. How Open source software works: "free" user-to-user assistance. *Research Policy*, 32, 6 (June 2003), 923-943.

31. Lerner, J. and Tirole, J. Some simple economics of open source. *Journal of Industrial Economics*, 50, 2 (June 2002), 197-234.

32. Lorrain, F. and White, H.C. Structural equivalence of individuals in social networks. *Journal of Mathematical Sociology*, 1, 1 (January 1971), 49-80.

33. Milton, L.P. and Westphal, J.D. Identity confirmation networks and cooperation in work groups. *Academy of Management Journal*, 48, 2 (April 2005), 191-212.

34. De Nooy, W., Mrvar, A., and Batagelj, V. *Exploratory Social Network Analysis with Pajek*. Cambridge University Press, Cambridge, UK, 2005.

35. Ransbotham, S. and Kane, G.C. Membership turnover and collaboration success in online communities: explaining rises and falls from grace in Wikipedia. *MIS Quarterly*, 35, 3 (September 2011), 613-627.

36. Roberts, J.A., Hann, I.-H., and Slaughter, S.A. Understanding the motivations, participation, and performance of open source software developers: a longitudinal study of the apache projects. *Management Science*, 52, 7 (July 2006), 984-999.

37. Singh, P.V. and Tan, Y. Developer heterogeneity and formation of communication networks in open source software projects. *Journal of Management Information Systems*, 27, 3 (Winter 2011), 179-210.

38. Singh, P.V., Tan, Y., and Mookerjee, V. Network effects: the influence of structural social capital on open source project success. *MIS Quarterly*, 35, 4, (December 2011), 813-829.

39. Thom-Santelli, J., Cosley, D.R., and Gay, G. What's mine is mine: territoriality in collaborative authoring. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems*, April 4-9, 2009, Boston, MA, ACM Press, New York, NY, 2009, 1481-1484.

40. Wasko, M.M. and Faraj, S. Why should I share? Examining social capital and knowledge contribution in electronic networks of practice. *MIS Quarterly*, 29, 1 (March 2005), 35-57.

41. Wasserman, S. and Faust, K. *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, UK, 1994.

42. Watts, D.J. and Strogatz, S.H. Collective dynamics of "small-world" networks. *Nature*, 393, 6684 (June 1998), 440-442.

43. Welser, H.T., Cosley, D., Kossinets, G., Lin, A., Dorkin, F., Gay, G., and Smith, M. In *Proceedings of the 2011 iConference*, University of Washington, Seattle, WA, February 8-11, 2011, ACM Press, New York, NY, 122-129.

44. Zhang, X. (Michael) and Zhu, F. Group size and incentives to contribute: a natural experiment at Chinese Wikipedia. *American Economic Review*, 101, 4 (June 2011), 1601-1615.