The impact of bifurcation on platform outcomes in a Q&A community

Xiaomeng Chen, Chris Forman

Cornell University

Michael Kummer

University of East Anglia

Abstract

While some platforms remain one united community that includes all subcommunities, others bifurcate subcommunities into spin-off platforms. Such bifurcation breaks the community in the original platform and forces users to reallocate between the home platform and the spin-off platform. In this paper, we investigate the impact of bifurcation on platform outcomes. We exploit the introduction of spin-off platforms in an online platform incubator, where users can propose to start a new spin-off platform, to identify the effects of bifurcation using a DID approach. We find that the bifurcation decreases user contribution in the home subcommunity. However, the two bifurcation communities generate more total user contribution and attract more new users, compared to a single united community. Moreover, the existing users decrease their contribution in the home platform and focus on new knowledge creation to the spin-off platform. Our results also show that subcommunities with higher level of interconnectivity experience a lower level of decrease in user contribution in the home subcommunity.

Keywords: Knowledge sharing, User contribution, Platform strategy

1. Introduction

Knowledge production platforms such as Wikipedia, GitHub and Stack Overflow have produced large volumes of knowledge as public goods to society. While some platforms remain one united platform that includes all subcommunities of users and content, others bifurcate into spin-off platforms. Such platform change can be driven by the need to separate different user groups into platforms that provide differentiated services, content or technologies. For example, Zhihu, the largest Q&A community in China, provides an editorial version, Zhihu Daily, to cultivate the preference of view-only users. In the open-source community, bifurcation (i.e. forking) is to start a new development in open source software based on existing code base (Nyman and Lindman 2013). Bifurcation can also allow the new spin-off platform to operate under different topics, format or regime of platform governance (Wikipedia's sister sites and Stack Exchange sites).

Previous literature has illustrated potential concerns over bifurcation (Karhu, Gustafsson and Lyytinen 2018). Bifurcation can create separated communities and a loss of economies of scale (Lerner and Tirole 2002). The spin-off platform might cannibalize the user base in the home platform and further decrease existing user contributions, as previous literature shows a positive effect of community size on user contributions (Zhang and Zhu 2011). It can also manifest challenges in community governance (von Krogh and von Hippel 2006) and failure in cooperation (Viseur 2012) when the bifurcated community has different interests from the original community. As bifurcation separates the community, it might cause inefficiency in knowledge production and can even lead to community failure (Rastogi and Nagappan 2016, Robles and González-Barahona 2012).

However, another stream of research portrays bifurcation more positively (Karhu, Gustafsson and Lyytinen 2018, Nyman 2014, Nyman et al 2012). In the context of open-source community, bifurcation need not be harmful (Simcoe and Watson 2019) and may create more advanced technology derived from the original community (Karhu, Gustafsson and Lyytinen 2018, Nyman 2014, Nyman et al 2012). Users may be better off with separated platforms when platforms are horizontally differentiated, and users have varying preferences (Farronato et al 2021). Moreover, when the home platform is overcrowded, bifurcating

into a spin-off platform may alleviate congestion issues stemming from competition and high search costs.

Despite ongoing work to understand the causes and consequences of bifurcation, there has been little work thus far examining the implications of bifurcation in online communities. In this paper, we take a first step in exploring the implications of bifurcation in communities, exploring the effects of bifurcation on the original platform, as well as the combined implications of bifurcation across the new and original platforms. We further explore the implications for new and existing users. While we frame how bifurcation might influences outcomes, at present most of our analyses are exploratory, examining the implications of bifurcation rather than testing hypotheses about how bifurcation influences outcomes and the conditions under which its effects are more positive or negative. At the end of this paper, we take further steps in this latter direction by exploring how the localization of network effects influences bifurcation outcomes. We plan to continue these explorations in future work.

To understand the effects of bifurcation, we exploit a unique setting in Stack Overflow (SO), the largest programming question and answer (Q&A) platform. SO hosts a platform incubator, area51.stackexchange.com (Area 51), where users can form a proposal to start a new Q&A site. Once the proposal for a new site gathers enough commitment from other users in Area 51, it will become a spin-off platform in Stack Exchange (stackexchange.com). In Stack Exchange, SO hosts established spin-off platforms in a range of diverse topics (including Stack Overflow itself) and the successful proposals in Area 51 will become one of the established sites on Stack Exchange. To identify the implications of bifurcation on the original platform, we first find the subcommunities in the home platform (SO) that are associated with proposals of new sites in Area 51. Then we identify the effect of successful bifurcation (i.e., establishing a site on Stack Exchange) by comparing subcommunities that successfully bifurcated and became a new spin-off platform on Stack Exchange and subcommunities that attempted to bifurcate but failed. Once the bifurcation succeeds, the original subcommunity splits into two parts, one is the spin-off platform and another one is the remaining subcommunity in the home platform (home subcommunity). And users of the bifurcated subcommunity need to decide how to allocate their contribution between these two places.

We define subcommunities that bifurcate from SO and become a new spin-off platform as treated subcommunities, and those that attempted to bifurcate but failed as control group. We compare both groups of subcommunities in a difference in differences (DID) approach to identify the effect of bifurcation. Our results show that bifurcation decreases user contributions to the home subcommunity. It also decreases the efficiency of knowledge exchange in the home subcommunity, which is measured by percentage of questions receiving an accepted answer within 16 hours. Moreover, after calculating the total contribution across the home subcommunity and the spin-off platform, we find that the combination of the two bifurcated communities provide more knowledge (i.e. more questions and answers) and attract more new users compared to a single united community. But the efficiency of knowledge exchange across two bifurcated communities is worsened off by the bifurcation. Overall, on average, when bifurcation occurs, while user contribution decreases in the home subcommunity, overall contribution increases across two bifurcated communities. But the efficiency of knowledge exchange decreases both in the home subcommunity alone and the two bifurcated communities combined.

We provide two exploratory findings. First, we look into the platform choice and contribution changes of existing users after the bifurcation. We examine how existing users change their contribution in the bifurcated subcommunity and other subcommunities in the home platform and whether they contribute to the spin-off platform. We find that existing users decrease their contribution in the home subcommunity and focus on knowledge creation in new topics in the spin-off platform. Second, we explore whether the effect of bifurcation can be moderated by the interconnectivity between bifurcated subcommunity and core subcommunities in the home platform. We define "interconnectivity" as the extent to which users in one cluster interact with users in other clusters on the global network (Zhu et al 2021). We focus on the interconnectivity between the bifurcated subcommunity and the core subcommunities in SO and find that subcommunities with higher level of interconnectivity experience a greater decrease in user contribution in the home platform. Moreover, subcommunities with lower level of interconnectivity experience a lower level of decrease in user contribution in the home platform.

Our research documents that bifurcations on a mature platform can have positive overall effects for

users. Although the spin-off platform indeed cannibalizes the home platform, two bifurcated communities lead to higher user contribution and attract more new users compared to one single united community. The insights are relevant for managers who are considering providing two differentiated platforms but are concerned about the impact of platform cannibalization. However, we note that the bifurcation can bring negative effect on knowledge exchange, which should be considered in creating a spin-off platform. Moreover, our exploratory results show existing users shift their effort from the home platform to the spin-off platform, focusing on knowledge creation in new topics. It indicates that the differentiated topics of knowledge in the spin-off platform might be effective in attracting existing users from the home subcommunities. Our results in subcommunities with high interconnectivity suggests that network structure is important in user's adoption decision of networks thus managers need to be aware of the interconnectivity of the bifurcated community when accessing the cannibalization from the spin-off platform.

After discussing how our work relates to prior research in the next section the paper proceeds as follows: Section 3 presents research context of this paper, Section 4 provides theoretical motivation of our empirical analysis, Section 5 describes the data and the identification strategy while Section 6 presents the empirical results. In Section 7, we conclude by discussing implications for platform strategy.

2. Literature Review

Our research contributes to a growing literature studying platform strategy, network effects and network adoption.

Early theoretical work has shown the presence of network effect (Katz and Shapiro 1985, Rochet and Tirole 2003). A big stream of more recent work characterizes the two-sidedness and network effect in platform business (Caillaud and Jullien 2003, Parker and Van Alstyne 2005, Weyl 2010). In their models, platforms attract multiple user groups (often with two sides such as buyers and sellers) and user's utility of joining a platform is characterized as an increasing function of the number of participating users in the platform. And these model focuses on pricing as a major platform strategy. Other studies have also looked into non-pricing strategies such as content differentiation (Seamans and Zhu 2014), quality (Zhu and Iansiti 2012) and platform merge (Farronato et al 2021). Our work extends existing literature by estimating

whether creating a spin-off platform brings net benefit and how users allocate their efforts between the original platform and the spin-off platform.

We also contribute to empirical literature in network externality and network adoption. Previous work shows evidence that network effects exist in technology adoption (Greenstein 1993, Saloner and Shepard 1995 and Tucker 2008). Several studies found that users' adoption decision and user participation can depend on other user's participation of the network, which suggested the presence of positive network effects and increasing return to scale (Zhu and Zhang 2010, Suarez 2005). However, some recent empirical studies find no evidence of increasing returns to scale in matching (Cullen and Farronato 2020, Li and Netessine 2020, Farronato et al 2021). We highlight one study, Farronato et al 2021, which examines a merger between two platforms. Their work is closely related to our paper because platform bifurcation is the reverse process of a platform merger. They find that users are not significantly better off after the merge when the network benefits are expected to increase. We contribute to this line of work by showing that although a single united community bifurcates into two communities, it doesn't necessarily reduce network benefits. On the contrary, the home platform and spin-off platform attract more new users and generate more knowledge creation after bifurcation.

3. Research Context

SO is a large programming Q&A platform in which users can pose and answer questions. In this section, we provide a brief overview of how the primary SO site functions and then offer an introduction to platform incubator hosted by Stack Overflow. Last, we provide contexts on the platform bifurcation process.

3.1 Stack Overflow Q&A Community

Users can ask and answer questions related to programming in the SO Q&A platform. The user who poses the question ("askers") can post it, and other users can answer the question ("answerers"). We summarize key features that are important to our analysis.

Questions, answers, and accepted answers. A question can have multiple answers. All users can vote up or down on a question or an answer, and those votes are then summarized as the question or answer's "score" (up votes minus down votes). The users who pose the questions can select only one answer as the

"accepted answer"—that is, the answer that they think successfully solves the question that was originally asked. The first answer is not always the accepted answer. In our analysis below, we will use questions and answers to quantify user contribution and we will use accepted answers to evaluate effective knowledge exchange.

Tags. The asker of each question can assign tags to it indicating the programming language, framework, or related module. For example, in our data the set of tags for one question includes "c#" and "asp.net-mvc", which shows that this question needs to be answered by someone who uses c# language and is familiar with the asp.net-mvc framework. Tags allow us to define the subcommunities of SO and identify which subcommunity is affected by a bifurcation.

3.2 Platform incubator and Stack Exchange

In 2010, SO hosted a platform incubator website, area51.stackexchange.com, to allow users to build new Q&A platforms (Atwood 2010).[1] Users can propose to start a new Q&A site. Figure 1a shows the ongoing proposals of sites in Area 51 and Figure 1b shows an example of a proposal. In the example proposal, the proposed site, "Proof Assistant", aim to generate knowledge in proof assistant for mathematicians and computer scientists. In the proposal, users discuss what the site should be used for and gather commitment from other users. Once the proposal gathers enough commitment (as indicated by the commitment percentage on the top right of Figure 1b), SO will establish a new site in Stack Exchange. Area 51 helps us to track all proposals that has been raised and the process from proposals to spin-off platforms. Over the past 10 years, Area 51 has successfully launched 172 sites from users' proposals with a wide range of topics. Those sites are hosted on Stack Exchange and become spin-off platforms of SO, covering a variety of topics that are different from SO. For example, typical sites related to technology topics are the Stack Exchange sites for Unix & Linux and the site of LaTeX. Sites related to other topics include the sites for Movies & TV and the site for Chess. These sites follow the same Q&A format and users need to sign up to participate in different sites.

[1] https://stackoverflow.blog/2010/09/28/factionalism-site-or-tag/

Figure 1a. Platform incubator: Area 51

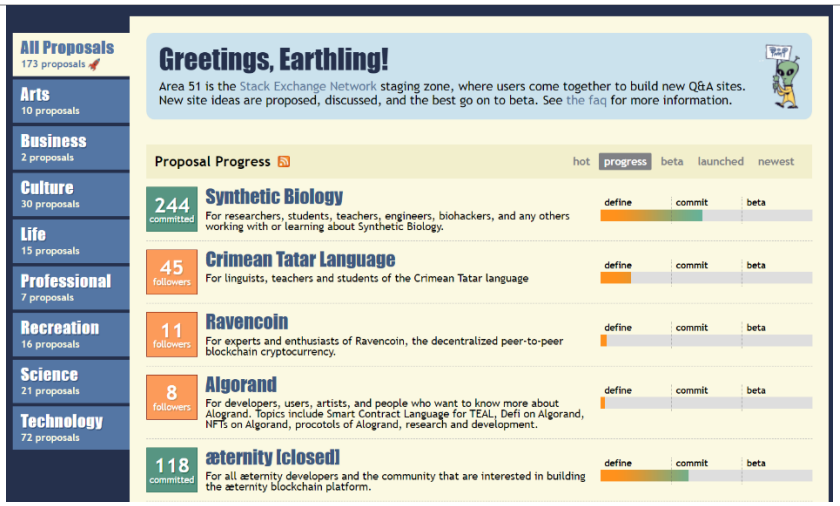
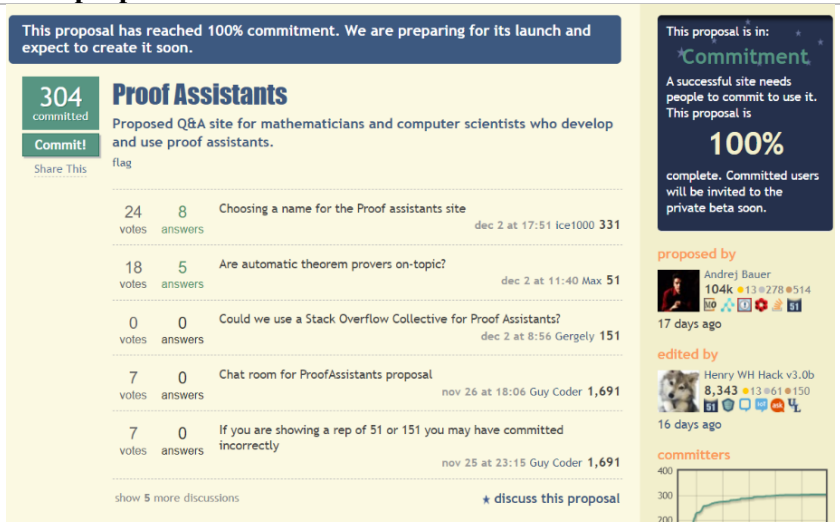


Figure 1b. A proposal of a new site in Area 51



We aim to understand the effect of bifurcation for both the home platform and spin-off platform. In the context of SO, the home platform is the original stackoverflow.com, which focuses on programming related questions and consists of subcommunities (i.e., tags) for programming languages such as python, c++ and Java. And the spin-off platforms are subcommunities that were proposed through Area 51, successfully bifurcated from SO and launched as a new Q&A site in Stack Exchange. In this paper we focus on proposed sites in Area 51 that are associated to subcommunities in the home platform (SO) and examine how a bifurcation of a subcommunity in SO can affect the platform outcomes.

3.3 Platform bifurcation

All users can propose to start a new Q&A platform in Area 51. Users will need to define the site, form the application and invite other users to participate in the proposal. Once the proposal gathers enough user participation in Area 51, SO will launch the site in Stack Exchange as a spin-off platform. Figure 2 provides a graphic of spin-off platforms hosted in Stack Exchange. We use question tag to define subcommunity in SO. For example, if a question contains an "android" tag, it belongs to "android" subcommunity and so do the answers of the question and the asker of the question. After the bifurcation, the original subcommunity in SO is split into two parts: one part is the remaining subcommunity in SO (home subcommunity) and another part is the spin-off platform on Stack Exchange. Users need to decide whether to sign up to the new spin-off platform and how to re-allocate their contribution between the home subcommunity and the spin-off platform if they multi-home. Figure 3 provides one example of a subcommunity after bifurcation: Figure 3a shows the remaining subcommunity in the home platform and Figure 3b shows the associated spin-off platform after bifurcation. If the proposal doesn't gather enough user commitment, SO will not support the new platform and the associated subcommunity in SO won't go through a bifurcation.

Figure 2. Spin-off platforms in Stack Exchange

For many sites, the bifurcation is driven by the need to generate knowledge under new topics that are not widely discussed in the home platform. This could happen for different reasons. For one, in the

home platform SO, users are encouraged to only ask/answer questions that are closely related to programming. But questions related to hardware, system installation or database administration are generally considered off-topic in the home platform[2]. Therefore, some subcommunities bifurcate in seek of

Figure 3. Subcommunity in spin-off platform vs. subcommunity in home platform

Figure 3a Latex subcommunity in home platform

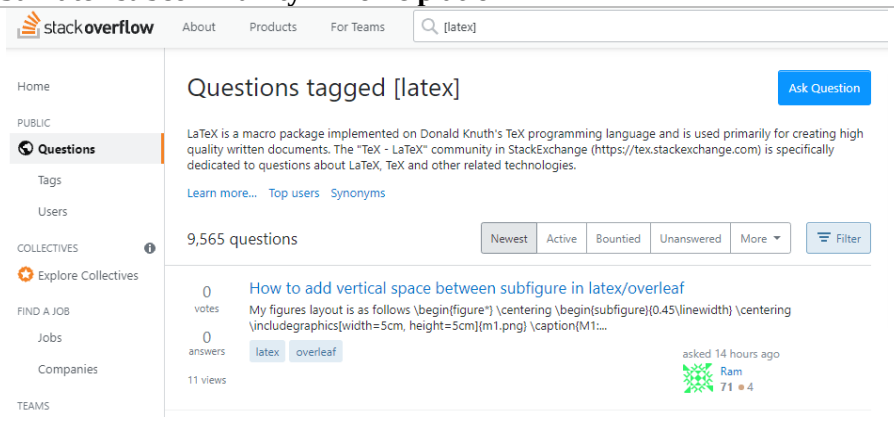
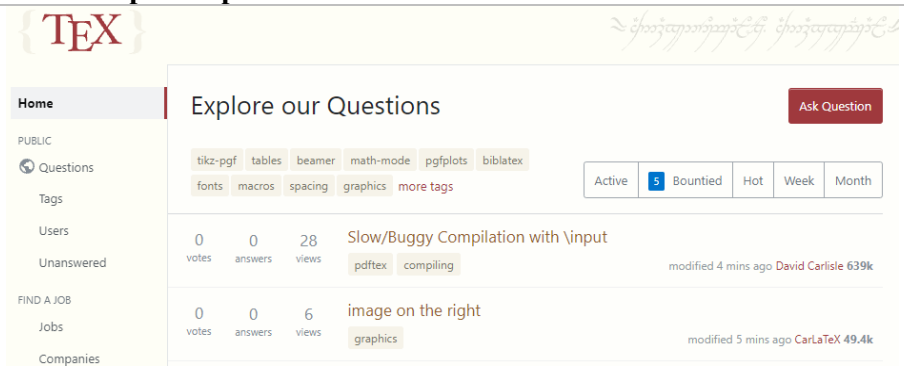


Figure 3b Latex spin-off platform



building a spin-off site that can help with questions related to these new topics. For example, in the home platform, the iOS subcommunity encourages questions related to programming in the iOS system, while in the bifurcated site, Ask Different[3], users are also encouraged to ask questions about Apple hardware. The difference in topics between home subcommunity and the spin-off platform is decided by the crowd, with a general guideline from community initiators, which leaves room for overlapping topics in both places.

After the bifurcation, the original subcommunity is split into two parts, one in the home platform

[2] https://stackoverflow.com/help/on-topic
[3] https://apple.stackexchange.com/

(SO) and another in the spin-off platform. Since we are interested in the total contribution in the remaining subcommunity in the home platform and the spin-off platform, two potential issues, duplicated questions and knowledge in new topics, may affect the measure in total contribution across two platforms. First, users may ask duplicated questions in two bifurcated communities to attract answerers from both places. However, such activity is prohibited by the community and will be removed by community moderator. In the data, we identified only a small percentage of duplicated questions and dropped such questions in calculating the total contribution[4]. So it won't contaminate the measures in the total contribution. Second, the spin-off platform allows users to generate knowledge in topics that are not encouraged in the home platform. As we add up total contribution to measure the aggregate outcome after bifurcation, we are concerned that the total contribution may be primarily driven by the knowledge generated in new topics. Therefore in this paper, we exclude the knowledge contribution in the new topics while calculating the total contribution. Details will be provided in section 5.2.

4. Theoretical Motivation

In this section we first contrast different theories of how network effects shape behavior. We define the concept of global and local network effects and how they have been explored in the literature. We then focus on our setting to describe how they influence outcomes when a platform bifurcates.

Prior work on network effects focused on the installed base of the entire network (Katz and Shapiro 1985, Farrell and Saloner 1986), i.e. global network effects. When global network effects are strong and platforms are undifferentiated, an early finding in the literature is that this can lead to "winner take all" outcomes (Arthur 1996, Shapiro and Varian 1999). Theoretical work that has focused on local network effects have argued that user adoption decisions can be influenced by the opinions and choices within his or her local network, such as friends and families (Lee et al 2006, Suarez 2005). For example, in communication technology adoption, other users in entire network (global network) may have a much

[4] We identify the duplicated questions by comparing the question titles. We compare two questions asked by the same user in the same day but one in the home platform and one in the spin-off platform. If the cosine similarity between titles of the two questions is greater than 0.7, the questions are identified as duplicated questions. In our baseline sample, only 108 questions (about 1.5%) are identified as duplicated questions and we dropped those duplicated questions. We will be further investigate in this issue and provide robustness results in the future.

smaller influence in user adoption compared to his close acquaintances (Lee et al 2006). Such local network bias may act as a brake on the winner-take-all process (Lee et al 2006).

Local vs. global network effects and users' platform adoption: We first extend the definition of local network as users with common interests. Then following previous literature in local networks (Zhu et al 2021, Lee et al 2006), we view the original platform as multiple local networks. In the context of SO, a local network is the subcommunity in which users contribute according to their common interests and we view SO as the global network that consists of multiple local networks. As mentioned in Section 3.1, we use tags to define subcommunities in SO. Therefore we view SO as multiple clusters of local network defined by tags, such as "c#", "python" and "java".

After a platform bifurcation, one subcommunity bifurcates away from the home platform and becomes a new spin-off platform. Users in the bifurcated subcommunity can choose to stay in the home platform and receive benefits in the global network. Alternatively, they can choose to multi-home and engage in both platforms, or even completely move to the spin-off platform and benefits only from local network provided by the spin-off platform. Note that as bifurcation creates two bifurcated subcommunities, one in the home platform and one in the spin-off platform, it provides two local networks for the users. If a user chooses to stay in the home platform only, the benefits from the local network in the remaining subcommunity in the home platform and from the global network by interacting with other local networks. While if he chooses to stay in the spin-off platform only, the benefits from the local network in the spin-off platform.

Therefore user's decision of contributing on the spin-off platform can be driven by comparing the benefits from the global network, from the local network in the home platform and the local network from the spin-off platform. In the global network, users benefit by interacting with users in other local networks and the global network itself such as greater visibility, better community governance. While in the local network in the spin-off platform, users benefit from differentiated topics of knowledge and a possibly greater community size, compared to the local network in the home platform. If all users benefit more from the global network and the local network in the home platform, they will choose to stay in the home platform,

and we will observe no significant change in user contribution in the home platform after bifurcation. However, if some users benefit more from the local network provided by the spin-off platform, they will decrease their contribution and direct a substantial share of their contributions to the new spin-off platform. In this case we would observe decrease in user contribution in the home platform after the bifurcation.

Moreover, it is unclear how the efficiency of knowledge exchange is affected by the bifurcation. We use the likelihood of a question receiving an accepted answer to measure the knowledge exchange between askers and answerers. Even if the bifurcation decreases the user contribution in the home platform, it is still not clear how it affects the exchange rate between these two groups of users, as bifurcation brings changes on both sides, question askers and answerers. On one hand, the bifurcation might disproportionally affect one side (answerers or askers) thus worsen off the knowledge exchange rate between the two groups. If question askers, especially new askers, are unaware of the new community, they will keep asking questions in the home platform. While answerers might move to the new platform according to their interests of contribution. On the other hand, the original subcommunity before bifurcation might be overcrowded and many questions remain unanswered because of high search cost. The bifurcation splits the original community into two places focusing on different topics and users interact in two less crowded communities. Therefore the bifurcation can improve the efficiency in knowledge exchange in the home platform.

We measure knowledge exchange by the ratio of the number of questions receiving an accepted answer within 16 hours (successful knowledge exchange) to the number of questions asked in the home subcommunity. Our first research question looks into both the effect of bifurcation on the user contribution and the knowledge exchange on the platform.

RQ1: How does bifurcation affect the user contribution and efficiency in knowledge exchange on the home platform?

Overall Contribution Generation: One major concern of bifurcation is that it might split communities and result in a loss of economies of scale (Lerner and Tirole 2002). The bifurcation splits one subcommunity into two parts, one of which stays in the home platform while the other one is in the spin-

off platform. If users can only single-home, with strong network effects (and large economies of scale) splitting the community will reduce the value of each network and might even result in an overall loss of value. In our application, the smaller economics of scale would translate to fewer user contributions and fewer new users even across both networks. We note though, that this effect is likely moderated if users can multi-home and simply contribute on both platforms. The opposite effect might ensue if the spin-off facilitates sizeable gains from differentiation that allow users to generate new knowledge in new topics that were not welcome in the relevant subcommunity of the home platform. If these gains outweigh the loss in scale users might ultimately be *better* off with two split and differentiated subcommunities. In that case we would expect to see that the two communities which result from the bifurcation combined attract more user contribution and more new users when considering the sum of contributions in both subcommunities - the subcommunity in the home platform and the spin-off platform.

To evaluate how bifurcation affects overall contributions on both platforms, we add up the user contribution in the home subcommunity and the spin-off platform and examine whether the overall contribution of bifurcated communities is greater than that of a single united community. Because sometimes bifurcation is driven by the requirement of creating knowledge in new topics that are not encouraged in the home platform, the spin-off platform is expected to generate more contribution in the new topics. Therefore we include only knowledge contribution in the old topics to obtain a more conservative calculation of the total contribution. We define contributions in old topics as questions/answers of which all tags appeared in the home subcommunity before the bifurcation. Details will be provided in Section 5.2.

Besides overall contribution, it is unclear how the bifurcation would change the efficiency of knowledge exchange in the two bifurcated communities. On one hand, bifurcation may create confusion in the boundary of these two places: askers might place a question in the home subcommunity while they should place it in the spin-off platform, or some questions are on the overlapped area of these two places and can fit in both. Answerers also need to decide to locate in two places according to their expectation of how askers post questions. Therefore the bifurcation can bring misalignment in askers and coordination

problem between askers and answerers, causing inefficiency in knowledge exchange. Such negative effect of bifurcation on knowledge exchange can be alleviated by user multi-homing and contributing on both places. On the other hand, the bifurcation might improve the knowledge exchange under certain circumstances. If the original subcommunity is overcrowded with many unanswered questions because of high search cost, the bifurcation can alleviate such congestion problem by creating two places focusing on different topics and allowing users to interact in two less crowded communities. With well-defined boundary between the two bifurcated communities, askers and answerers can self-select into two places after bifurcation with successful coordination and the knowledge exchange can be better off.

We calculate the efficiency of knowledge exchange across the home subcommunity and the spin-off platform and examine whether the overall efficiency of bifurcated communities is greater than that of a single united community. More specifically, we use the ratio of questions receiving an accepted answer within 16 hours to questions asked as measure of efficiency in knowledge exchange. Following the same principle of calculating the overall contribution, the numerator is calculated by the number of questions receiving an accepted answer and only in the old topics from both the spin-off platform and home subcommunity. And the denominator is calculated by the number of questions asked only in the old topics and from both the spin-off platform and home subcommunity.

RQ2: How does bifurcation affect the aggregate contribution and the overall efficiency in knowledge exchange on both the home and the new platform?

5. Identification and Data Strategy

5.1 Identification of the Main Effect

We seek to understand how platform bifurcation affects user contribution in the home subcommunity and the spin-off platform. We identify the subcommunities that bifurcated from SO and became a new spin-off platform as treated groups and subcommunities that attempted to bifurcate but failed as counterfactuals. To address the concern of selection in the treated groups, we use coarsened exact matching (CEM) to match on control groups as robustness checks. We use difference in difference (DID) approach to identify the effect of bifurcation.

We first define a subcommunity in Stack Overflow by question tag. For example, if a question contains an “android” tag, it belongs to “android” subcommunity and so do the answers to the question and the asker of the question. Then we associate subcommunities in SO with proposed sites in Area 51 by matching the site name and description with tag name and description. We identify the tags that bifurcated successfully from SO as treated tags and select a 32-week time window, 16 weeks before and 16 weeks after the bifurcation.

The major challenge in this identification strategy is to find counterfactuals for treated groups. We use two approaches to select on control tags. In our baseline results, we select tags that attempted to bifurcate but failed as control tags. We select the tags that tried to create a spin-off site in Area 51 at some point in the sample. These tags failed in bifurcation because they didn’t gather enough user commitment in the proposal of a new site in Area 51. As successful bifurcation depends on whether the proposal gathers enough user commitment, it can be driven by the size and user activity of the treated tag in the home platform before bifurcation. If the trends in our treated tags are shaped by these factors in a way that is systemically different from those of the control tags, they have the potential to shape our results. Therefore in the robustness checks, we use CEM to identify control tags with similar characteristics, which helps to mitigate concerns about differences in unobservable trends between control tags and treated tags. We choose to present the sample before CEM as baseline results because the differences in most matching variables are insignificant and it is a more representative sample. The matching slightly improves the sample balance but drops half of treated tags, so we use the after CEM sample as robustness checks. We will provide more details about these two samples in Section 5.2.

For each treated tag, we select 32-week window and select tags that exist in this time window and once attempted to bifurcate but failed as control tags. For the control tags, we use the same 32-week window as the treated tag. It is possible that some control tags are selected to multiple treated tags. So we keep only one occurrence for control tags that are selected to multiple treated tags and have overlapped selected time window. After identifying the control group, we construct a panel data set to run a DID regression. Our baseline empirical specification is as follows:

$$Y_{jt} = \beta_0 + \beta_1 Bifurcation_{jt} + \alpha_j + Week_t + \varepsilon_{ijt} \quad (1)$$

Y_{jt} indicates the outcome variable of tag j on time t. We examined user contribution (questions and answers) and contributors (users who answer questions) along with other variables to measure community outcomes. We looked both at the home subcommunity outcome and the combined outcome. Our primary outcome variables Y_{jt} denote the total answers and contributors in the focal tag. For treated tag j, $Bifurcation_{ijt} = 1$ when the tag j bifurcates from Stack Overflow and starts a new site at time t. For control tag j, $Bifurcation_{ijt} = 0$ during the 32-week time window. Tag level fixed effects α_j and weekly dummies $Week_t$ are included.

5.2 Data

Our primary source of data is publicly available through the Stack Exchange API. We supplement the publicly available information with additional data from area51.stackexchange.com and archive.org to obtain the current and historical record of the proposed sites.

First, we identify all tags that attempted to bifurcate by comparing the name and description of proposed sites in Area 51 and the name and description of tags in SO. Next, we identify treated and control tags. From October 2010 to December 2020, we identified 224 tags that attempted to bifurcate from SO, of which 50 tags succeeded and became a new spin-off platform in Stack Exchange. In our baseline sample ("before CEM"), we use the other 174 tags that attempted to bifurcate but didn't succeed as our control tags. In the robustness check ("after CEM"), we use CEM to identify control tags that are similar to the treated tags in terms of pre-period characteristics. We use tag-level characteristics 10 weeks before the bifurcation as matching covariates: question count, answer count, tag age, percentage of days that have at least one question and percentage of programming related questions and the growth rate of the question count. Appendix Table A1 describes each of the matching covariates.

Tables 1 and 2 show the balance check on matching covariates before and after CEM. Before CEM, the differences in means between the treatment and control groups are insignificant, except for active percentage in day -30. It indicates that the differences in means between the treatment and control groups

are small before CEM. After CEM, differences across all matching covariates are insignificant but the sample only contains 32 of the treated tags. The matching improves the sample balance between treated and control groups, with the sacrifice of dropping almost half of treated tags. Therefore we decide to use the more representative sample-all treated tags with control tags before CEM-in our baseline regressions and leave the sample after CEM as robustness checks. The results remain robust to control tags that are selected by CEM (Appendix Tables A2, A3, A4, A5 and Figure A2).

Table 1 Balance check before CEM

	treat	control	Difference	Se	(p-value)
Question count from day -1 to day -30	187.00	138.76	48.24	82.42	0.56
Question count from day -30 to day -60	177.76	126.72	51.04	74.61	0.49
Total question up to day -30	3,009.06	2,994.90	14.16	1,431.74	0.99
Active percentage in day -30	0.51	0.31	0.20	0.06	0.00
Tag age (day) in day -30	1,399.00	1,212.20	186.80	134.54	0.17
Answer count from day -1 to day -30	315.52	274.85	40.67	179.44	0.82
Answer count from day -30 to day -60	297.44	248.64	48.80	158.27	0.76
Total answer up to day -30	6,296.86	6,227.44	69.42	2,997.67	0.98
Question growth rate	0.11	0.26	-0.15	0.15	0.32
Ratio of program questions	0.19	0.16	0.02	0.02	0.30
N. of tags	50	174			

Note: The column "treat" represents the average value of variables in the first column across treated tags in the week before bifurcation (matching period), and the column "control" represents the average value of variables in the first column across control tags in the week before bifurcation (matching period). The column "Difference" represents the sample difference between treated tags and control tags.

Table 3 provides summary statistics for the outcome variables in our regressions, based on the before CEM sample in Table 1. We focus on three aspects of platform outcomes: user contribution, knowledge exchange and new user attraction. We use log of question count, log of answer count and log of contributor count (users who provide answers) to measure user contribution in the platform. We also use log of question count that get an accept answer within 16 hours (Log(Accept16Hour)) to measure user contribution with effective knowledge exchange. To understand the efficiency of knowledge exchange, we construct a variable (Accept16HourRatio) to measure the ratio of successful knowledge exchange between askers and answerers within 16 hours. This variable is calculated by the number of questions that get an

accepted answer within 16 hours divided by the number of questions asked and there will be missing value when the denominator is zero. We use number of new users (Log(NewUser)) to measure new user attraction.

Table 2 Balance check after CEM

	treat	control	Difference	Se	(p-value)
Question count from day -1 to day -30	139.56	136.09	3.47	53.08	0.95
Question count from day -30 to day -60	134.66	135.42	-0.76	52.29	0.99
Total question up to day -30	3,340.78	4,578.91	-1,238.13	1,889.79	0.51
Active percentage in day -30	0.57	0.45	0.12	0.07	0.11
Tag age (day) in day -30	1,432.59	1,398.71	33.89	175.81	0.85
Answer count from day -1 to day -30	245.75	220.94	24.81	85.07	0.77
Answer count from day -30 to day -60	232.13	224.12	8.00	85.28	0.93
Total answer up to day -30	6,849.69	8,939.00	-2,089.31	3,815.95	0.59
Question growth rate	0.06	-0.04	0.10	0.06	0.12
Ratio of program questions	0.17	0.17	0.00	0.02	0.88
N. of tags	32	65			

Note: The column "treat" represents the average value of variables in the first column across treated tags in the week before bifurcation (matching period), and the column "control" represents the average value of variables in the first column across control tags in the week before bifurcation (matching period). The column "Difference" represents the sample difference between treated tags and control tags.

Table 3. Summary statistics

	N	Mean	St.Dev	min	max
Log(Question) Home Platform	7165	1.7943	1.7010	0	7.3218
Log(Answer) Home Platform	7165	2.0833	1.8942	0	8.1271
Log(Contributor) Home Platform	7165	2.0311	1.8411	0	7.7502
Log(Accept16Hour) Home Platform	7165	1.2660	1.4992	0	6.8794
Accept16HourRatio Home Platform	5058	0.4054	0.2783	0	1
Log(NewUser) Home Platform	7165	0.8651	1.1360	0	5.5175
Log(Question) Home+spin-off Platform	7165	1.8614	1.7143	0	7.3218
Log(Answer) Home+spin-off Platform	7165	2.1656	1.9121	0	8.1271
Log(Contributor) Home+spin-off Platform	7165	2.1103	1.8573	0	7.7502
Log(Accept16Hour) Home+spin-off Platform	7165	1.3543	1.5461	0	6.8794
Accept16HourRatio Home+spin-off Platform	5191	0.3775	0.2793	0	2.1111
Log(NewUser) Home+spin-off Platform	7165	0.8982	1.1473	0	5.5175

In Table 3, the first six variables are platform outcomes in the home subcommunities only. The last six variables are calculated across both home subcommunity and spin-off platform but only including the knowledge generated in old topics in the spin-off platform. As mentioned in section 3.3, the spin-off platforms allow users to generate knowledge in new topics that are now encouraged in the home platform. As we add up total contribution to measure the aggregate outcome after bifurcation, we are concerned that

the total contribution may be primarily driven by the knowledge generated in new topics. Therefore we exclude the knowledge contribution in new topics and measure the total contribution across two platforms in old topics only. We identify tags from questions asked before bifurcation as old topics and tags that have never appeared in any question before bifurcation as new topics. For a given question created in the spin-off platform, if all its tags have appeared before bifurcation, it is identified as knowledge contribution in old topics. Otherwise, it is knowledge contribution in new topics. We calculate total contribution as questions/answers in old topics only and exclude the contribution in the new topics. For example, in Table 3, "Log(Question) Home+spin-off Platform" is the log of total questions in the home subcommunity and in the spin-off platform but only in the old topics. Any questions in the new topics are excluded when calculating this measure.

6. Results

6.1 Main Effect of the bifurcation

Tables 4 and 5 present our main results. Table 4 shows that the bifurcation decreases user contribution (number of questions, answers and contributors) in the home subcommunity. Column 2 shows that, the log of weekly answers (Log(Answer)) declines by 21.31% ($e^{-0.2396} - 1$). The results for log of question count, log of answer count and log of questions that get an accepted answer within 16 hours are qualitatively similar. Moreover, Column 5 shows that bifurcation reduces the ratio of questions that receive an accepted answer within 16 hours by 2.9 percentage points, or 7% (0.029/0.41). However, the bifurcation doesn't have significant negative impact on the new users joining in the home platform. One possible explanation is that the spin-off platform is less known than the home platform SO and some new users are unaware of the bifurcation. So some new users' decision of joining the platform is not affected by the bifurcation.

In Table 5, we examine the effects of treatment on the total (combined) outcomes of home subcommunity plus the spin-off platform. We use the same outcome variables but aggregate the contribution across home subcommunity and spin-off platform, including only knowledge contribution in the old topics. For example, in Column 2, we aggregate answers from the home subcommunity and answers from the associated spin-off platform. We include the answers in the new spin-off platform only if all tags

of the answer appeared in the home platform before bifurcation (i.e. old topics). Such measure excludes answers in the spin-off platform with tags that are newly generated (i.e. new topics). Therefore, our aggregate measure is not driven primarily by the new topics encouraged in the spin-off platform. We follow the same principle to calculate other total outcomes in Table 5.

Table 5, column 2 shows that the total number of answers across the home subcommunity and the spin-off platform in the old topics increases by 33.03% ($e^{0.2854} - 1$). It indicates that after bifurcation, there is a net increase in total user contribution across the home subcommunity and the spin-off platform even when we exclude the knowledge in new topics that are encouraged in the spin-off platform. Column 3 shows that the home subcommunity and spin-off platform together attract more new users to contribute in the old knowledge topics. However, Column 5 shows that the bifurcation decreases the ratio of questions that receive an accepted answer within 16 hours. It indicates that the efficiency of knowledge exchange within the old topics of knowledge across two bifurcated communities is worsened off after bifurcation.

6.2 Robustness Checks

Our identification relies on the assumption that the trend in user activity for treated tag is similar to that for control tag until after the incidence of treatment. To explore the validity of this assumption, we include leads and lags to illustrate the difference in the trend (before and after the treatment, here called pre-trend and post-trend) between the control and treated tags.

We present the pre- and post-trend analysis in Figure 4. We construct dummy variables for tag-week observations with treated tags, based on the relative weeks to the treatment day. To construct pre-trend variables, we create dummy variables that equal one for tag-week observations affiliated with treated tags that were before treatment (week -15, week -14, …, and week -1 respectively in figure 4). We similarly construct dummy variables that equal one for treated tags based on the number of weeks after the bifurcation. The reference group consists of tag-week observations in week -16.

Figure 4 presents the week-by-week estimates of pre- and post-trend dummies. It shows that in the pre-period, there is no evidence of a statistically significant effect from affiliation with a treated tag before

Table 4. The bifurcation reduces user contributions in the home platform

	Log(Question) Home platform	Log(Answer) Home platform	Log(Contributor) Home platform	Log(Accept16Hour) Home platform	Accept16HourRatio Home platform	Log(NewUser) Home platform
	(1)	(2)	(3)	(4)	(5)	(6)
Bifurcation	-0.1481**	-0.2396***	-0.2086***	-0.1426**	-0.0290**	-0.0200
	(0.0697)	(0.0810)	(0.0755)	(0.0680)	(0.0146)	(0.0376)
R2	0.0180	0.0199	0.0176	0.0114	0.0037	0.0046
N	7,165	7,165	7,165	7,165	5,058	7,165
N. of tag	224	224	224	224	213	224
Mean	1.79	2.08	2.03	1.27	0.41	0.87

* p<0.1; ** p<0.05; *** p<0.01. Heteroskedasticity robust standard error clustered at the tag level in parentheses. Includes tag level fixed effects and weekly dummies.

Table 5 The bifurcation increases total user contribution across the combined home platform and spin-off platform but reduces the efficiency in knowledge exchange.

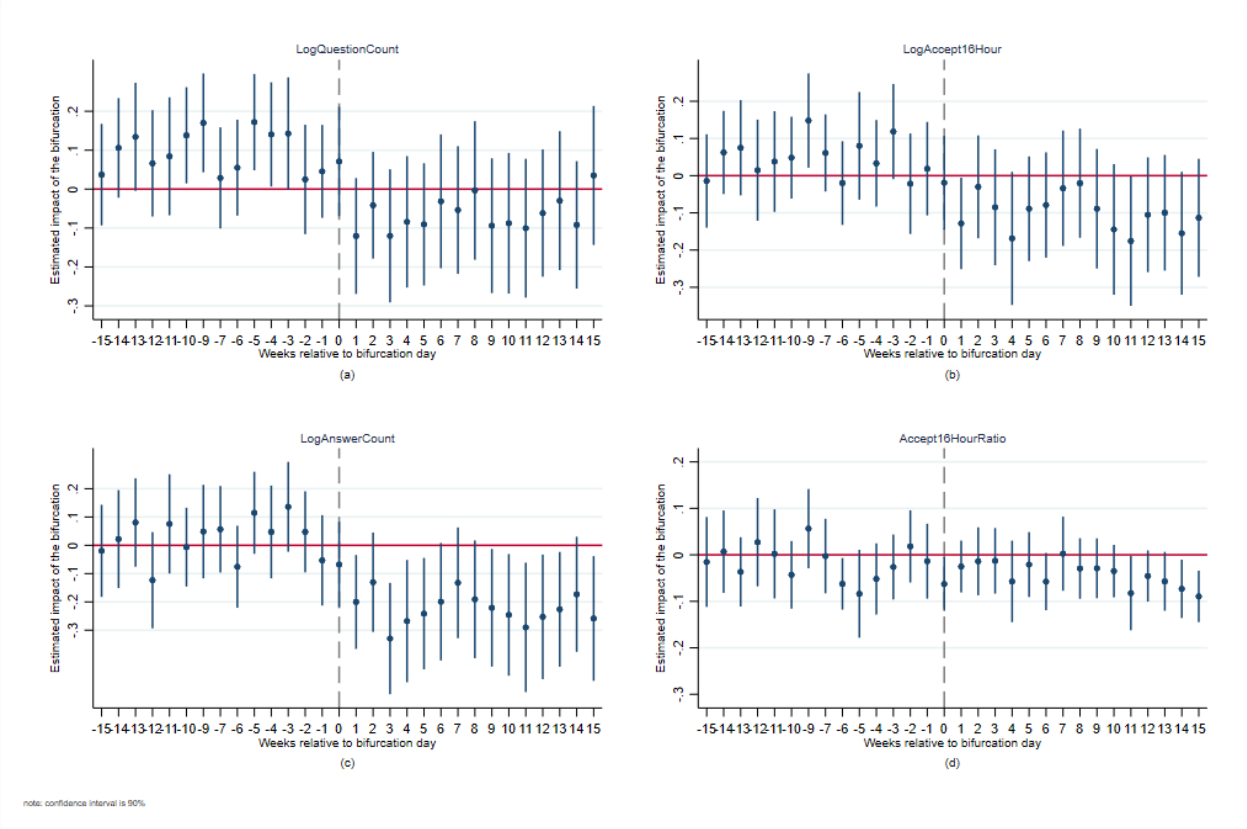
	Log(Question) Home + spin-off platform	Log(Answer) Home + spin-off platform	Log(Contributor) Home + spin-off platform	Log(Accept16Hour) Home + spin-off platform	Accept16HourRatio Home + spin-off platform	Log(NewUser) Home + spin-off platform
	(1)	(2)	(3)	(4)	(5)	(6)
Bifurcation	0.2774***	0.2854***	0.3030***	0.2774***	-0.1314***	0.1641***
	(0.0654)	(0.0714)	(0.0674)	(0.0654)	(0.0225)	(0.0387)
R^2	0.0664	0.0507	0.0553	0.0664	0.0298	0.0232
N	7,165	7,165	7,165	7,165	5,191	7,165
N. of tags	224	224	224	224	213	224
Mean	1.86	2.17	2.11	1.86	0.38	0.90

* p<0.1; ** p<0.05; *** p<0.01. Heteroskedasticity robust standard error clustered at tag level in parentheses. Includes tag level fixed effects and weekly dummies.

treatment in answer count, ratio of questions receiving an accepted answer within 16 hours and number of questions receiving an accepted answer within 16 hours. After the bifurcation, the treated tags have fewer user participation. While in the pre-period, the treated tags have more question count than the control tags, after the bifurcation, there is no significant difference in question count between the treated tags and control tags. The difference in log of question count between treated and control tags drops after the bifurcation.

We present robustness checks for the after CEM samples in the Appendix. Both the main results and pre-trend analysis are consistent with the sample before CEM. Tables A2 and Table A3 replicate the analysis in Table 4 and Table 5 respectively. Figure A1 is the pre-trend analysis using the after CEM sample.

Figure 4 pre-trend analysis: home platform



6.3 Exploring existing users' platform choice after bifurcation

In this section, we present results that describe existing users' platform choice and how they shift their contribution after bifurcation.

The bifurcation affects the platform outcome by making users change the extent of their contributions on the home platform and reallocate effort to the spin-off platform. To track users' platform

choice and change in contribution, we examine the existing users as their activity before and after bifurcation is observable. By comparing the contribution before and after bifurcation for existing users, we can infer their decisions in platform adoption and effort allocation, which helps us to understand their choice between the home platform and the spin-off platform. To understand their changes in contribution, we look into how existing users change their contribution in the focal subcommunity, in other subcommunities in the home platform. We also examine the overall contribution of the existing users across the spin-off platform and the home subcommunity to indicate whether they join and shift their contribution to the spin-off platform.

RQ3: *Do existing users reallocate their contribution to other subcommunities in the home platform or do they contribute to the spin-off platform?*

Table 6 The existing users reduce contribution in the home platform and reallocate knowledge contribution to new topics.

	Log(OldUserAnswer) Focal Tag Home platform (1)	Log(OldUserAnswer) Other Tags Home platform (2)	Log(OldUserAnswer) Old Knowledge Home + spin-off platform (3)	Log(OldUserAnswer) Old & New Knowledge Home + spin-off platform (4)
Bifurcation	-0.6630***	-0.0994***	-0.3016***	0.2447**
	(0.0894)	(0.0311)	(0.0864)	(0.1107)
R^2	0.3499	0.1446	0.2764	0.1831
N	7,165	7,165	7,165	7,165
N. of tags	224	224	224	224
Mean	1.66	4.98	1.71	1.79

* p<0.1; ** p<0.05; *** p<0.01

Heteroskedasticity robust standard error clustered at tag level in parentheses. . Includes tag level fixed effects and weekly dummies.

Table 6 shows contribution from existing users and their reallocation of effort between the home platform and the spin-off platform. Column 1 shows that the existing users reduce their contribution in answers after bifurcation by 48.46% ($e^{-0.6630} - 1$) in the remaining subcommunity in the home platform. Column 2 shows that they also reduce contribution to tags other than the focal tag in the home platform by 9.46% ($e^{-0.0994} - 1$). It indicates that existing users reduce their effort not only in the bifurcated subcommunity but also in other subcommunities in the home platform. Column 3 shows that in knowledge creation of old topics, the existing users *decrease* total contribution across home subcommunity and spin-

off platform by 26.03% ($e^{-0.3016} - 1$). However, Column 4 shows that they *increase* the total contribution across home platform and spin-off platform in both old topics and new topics by 27.72% ($e^{-0.3016} - 1$). It suggests that after the bifurcation, existing users shift to the spin-off platform but instead of contributing to old topics, they focus more on the knowledge creation in new topics. Table 6 suggests that the existing users value benefits from local network effect and they may be attracted by the new topics provided in the spin-off platform. Table A4 in Appendix shows robustness check using after CEM sample and is consistent with Table 6.

6.4 Exploring interconnectivity to core subcommunities as a moderating factor

In this section, we present further explorations that might be helpful to uncover the mechanisms of platform bifurcation. We provide theoretical motivation and empirical results in how interconnectivity to core subcommunities can moderate the effects of bifurcation.

As mentioned in Section 4, we extend the definition of local network as users with common interests and view the subcommunities in the home platform (SO) as multiple local networks. As one subcommunity bifurcates away from the home platform and becomes a new spin-off platform, the users need to allocate effort in the home platform and the new spin-off platform. Once a user contributes primarily to the spin-off platform, he/she loses the benefits of interacting with users in other local networks from the global network. We use "interconnectivity" to measure the degree to which users in one cluster interact with users in other clusters on the global network (Zhu et al 2021). Interconnectivity characterizes the inter-relationship between a local network cluster to other clusters in the global network. In this paper, we focus on the interconnectivity between the bifurcated subcommunity and the core subcommunities in SO. We rank all tags by number of total questions one month before the bifurcation and identify the top 50 tags as core subcommunities in SO. These core subcommunities represent large communities in SO such as "Python", "c++" and "java". We argue that users in subcommunities that have higher interconnectivity with the core subcommunities in the home platform benefit more from the global network because they value interactions with users in the core clusters of networks. As a result, they are less likely to engage with the spin-off platform after bifurcation but instead remain engaged with the home platform. So we hypothesize

that subcommunities with higher interconnectivity exhibit lower level of decrease in user contribution after the bifurcation.

RQ4: *Will subcommunities (tags) with a higher level of interconnectivity in the home platform have lower level of decrease in user contribution in the home platform after the bifurcation?*

Table 7 shows heterogeneity effects of bifurcation based on the interconnectivity between the focal tag and the core tags in the home platform. We use the percentage of questions in the focal tag that also contains one of the top 50 tags to measure interconnectivity with the core tags. We define tags that have more than 30% questions[5] associate with the top 50 tags as tags with high interconnectivity, *HighIntercon*. We interact this variable with the treatment variable *Bifurcation.*

Table 7 Tags with higher interconnectivity have lower level of decrease in user contribution in the home platform

	Log(Question) Home platform (1)	Log(Answer) Home platform (2)	Log(Contributor) Home platform (3)	Log(Accept16Hour) Home platform (4)
Bifurcation*HighIntercon	0.2584*	0.3031*	0.2792*	0.2050
	(0.1330)	(0.1548)	(0.1444)	(0.1319)
Bifurcation	-0.2774**	-0.3912***	-0.3482***	-0.2451**
	(0.1076)	(0.1234)	(0.1149)	(0.1057)
R^2	0.0241	0.0263	0.0234	0.0156
N	7,165	7,165	7,165	7,165
N. of tags	224	224	224	224
Mean	1.79	2.08	2.03	1.27
Linear combination	-0.0189	-0.0880	-0.0690	-0.0401
	(0.0796)	(0.0948)	(0.0885)	(0.0797)

* p<0.1; ** p<0.05; *** p<0.01. Heteroskedasticity robust standard error clustered at tag level in parentheses. . Includes tag level fixed effects and weekly dummies.

The results show that tags with higher interconnectivity have a lower level of decrease in user contribution in the home subcommunity, compared to tags with low interconnectivity. For example, in column 2, the coefficient on Bifurcation*HighIntercon is significantly positive, indicating that the tags with higher interconnectivity has a lower level of decrease in answers compared to tags with lower

[5] In the treated tags, 30% is the 50 percentile of the distribution of interconnectivity.

interconnectivity. The results for log of question count, log of contributor count and log of questions that get an accepted answer within 16 hours are qualitatively similar. Table 7 suggests that users in subcommunities have a stronger connection with the core subcommunities are more likely to keep engaging in the home subcommunity when a new spin-off platform is available. Table A5 in Appendix shows robustness check using after CEM sample and is consistent with Table 7.

7. Discussion and Conclusions

We examine the effects of platform bifurcation on the platform outcomes of both home platform and the spin-off platform. We analyze this question using the platform incubator set up in the Q&A platform Stack Overflow. We identify the causal effect of bifurcation by exploiting the multiple incidents of spin-off platform in SO, which causes subcommunities to bifurcate into two parts. We provide two main findings. First, the bifurcation reduces the user contribution (questions, answers, contributors, etc) in the home subcommunities. Second, we show that after bifurcation, the two split communities, one in the home platform and another in the spin-off platform, generate more user contribution and attract more new users, even when we only include knowledge in the old topics in spin-off platforms. It indicates that the two bifurcated communities generate better platform outcome compared to a single united community. In addition to the main findings, we also provide two exploratory findings. First, we provide evidence on existing users' platform choice after bifurcation. Our results show that existing users reduce their contribution in both the bifurcated subcommunity and in other subcommunities in the home platform. They shift their effort from the home platform and focus on generating knowledge in new topics in the spin-off platform. Second, we show that subcommunities with higher interconnectivity experience a lower level of decrease in user contribution in the home subcommunity, compared to the subcommunities with lower interconnectivity.

This paper has several limitations. First, our primary identification strategy relies on the using the tags that attempted to bifurcate but failed as counterfactuals. It requires tags that bifurcated successfully do not change over time in ways that are different from those in the control group. Although balance checks support that in the pre-period, control tags and treated tags do not different significantly in tag-level

characteristics, and our findings are unchanged if we use control tags generated by CEM, our selection in the control tags might still have tag-level unobservables that are different from the treated tags and can affect our estimates.

Our findings have important managerial implications. First, our results main findings show that the spin-off platform indeed cannibalizes the home platform but two bifurcated communities lead to higher user contribution and attract more new users compared to one single united community, indicating that platform bifurcation does not necessarily lose in economics of scales. The managerial implication is that providing a spin-off platform might be an attractive option for managers who are considering providing two differentiated platforms but are concerned about the impact of platform cannibalization. Second, our exploratory results show existing users shift their effort from the home platform to the spin-off platform, focusing on the knowledge creation in new topics. Therefore platform differentiation might be effective in attracting existing users in creating spin-off platforms. Moreover, our results in subcommunities with high interconnectivity suggests that network structure is important in user's adoption decision of networks thus managers need to be aware of the interconnectivity of the bifurcated community when accessing the cannibalization from the spin-off platform.

References

Armstrong, M. (2006). Competition in two-sided markets. The RAND journal of economics, 37(3), 668-691.

Atwood J (2010) Factionalism: Site or Tag? Accessed November 29, 2021, https://stackoverflow.blog/2010/09/28/factionalism-site-or-tag/.

Caillaud, B., & Jullien, B. (2003). Chicken & egg: Competition among intermediation service providers. RAND journal of Economics, 309-328.

Cullen, Z., & Farronato, C. (2020). Outsourcing tasks online: Matching supply and demand on peer-to-peer internet platforms. Management Science, 67(7), 3985-4003.

Farronato, C., Fong, J., & Fradkin, A. (2020). Dog eat dog: Measuring network effects using a digital platform merger (No. w28047). National Bureau of Economic Research.

Greenstein, S. M. (1993). Did installed base give an incumbent any (measureable) advantages in federal computer procurement?. The RAND Journal of Economics, 19-39.

Karhu, K., Gustafsson, R., & Lyytinen, K. (2018). Exploiting and defending open digital platforms with boundary resources: Android's five platform forks. Information Systems Research, 29(2), 479-497.

Katz, M. L., & Shapiro, C. (1985). Network externalities, competition, and compatibility. The American economic review, 75(3), 424-440.

Lerner, J., & Tirole, J. (2002). Some simple economics of open source. The journal of industrial economics, 50(2), 197-234.

Li, J., & Netessine, S. (2020). Higher market thickness reduces matching rate in online platforms: Evidence from a quasiexperiment. Management Science, 66(1), 271-289.

Lee, E., Lee, J., & Lee, J. (2006). Reconsideration of the winner-take-all hypothesis: Complex networks and local bias. Management science, 52(12), 1838-1848.

Nyman, L., & Lindman, J. (2013). Code forking, governance, and sustainability in open source software. Technology Innovation Management Review, 3(1).

Nagaraj, A., & Piezunka, H. (2020). How Competition Affects Contributions to Open Source Platforms: Evidence from OpenStreetMap and Google Maps (p. 58). Working Paper.

Parker, G. G., & Van Alstyne, M. W. (2005). Two-sided network effects: A theory of information product design. Management science, 51(10), 1494-1504.

Rastogi, A., & Nagappan, N. (2016, March). Forking and the Sustainability of the Developer Community Participation--An Empirical Investigation on Outcomes and Reasons. In 2016 IEEE 23rd international conference on software analysis, evolution, and Reengineering (SANER) (Vol. 1, pp. 102-111). IEEE.

Robles, G., & González-Barahona, J. M. (2012, September). A comprehensive study of software forks: Dates, reasons and outcomes. In IFIP International Conference on Open Source Systems (pp. 1-14). Springer, Berlin, Heidelberg.

Rochet, J. C., & Tirole, J. (2003). Platform competition in two-sided markets. Journal of the european economic association, 1(4), 990-1029.

Seamans, R., & Zhu, F. (2014). Responses to entry in multi-sided markets: The impact of Craigslist on local newspapers. Management Science, 60(2), 476-493.

Saloner, G., & Shepard, A. (1995). Antitrust Law and Intellectual Property: Intersection or Crossroad. The Computer & Internet Lawyer, 28(3), 479-501.

Suarez, F. F. (2005). Network effects revisited: The role of strong ties in technology selection. Academy of Management Journal, 48(4), 710-720.

Tan, G., & Zhou, J. (2021). The effects of competition and entry in multi-sided markets. The Review of Economic Studies, 88(2), 1002-1030.

Tucker, C. (2008). Identifying formal and informal influence in technology adoption with network externalities. Management Science, 54(12), 2024-2038.

Von Krogh, Georg, and Eric Von Hippel. "The promise of research on open source software." Management science 52.7 (2006): 975-983.

Viseur, R. (2012). Forks impacts and motivations in free and open source projects. International Journal of Advanced Computer Science and Applications, 3(2), 117-122.

White, A., & Weyl, E. G. (2016). Insulated platform competition. Available at SSRN 1694317.

Weyl, E. G. (2010). A price theory of multi-sided platforms. American Economic Review, 100(4), 1642-72.

Zhang XM, Zhu F (2011) Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. American Economic Review, 101(4):1601-1615.Zhu, F., & Iansiti, M. (2012). Entry into platform-based markets. Strategic Management Journal, 33(1), 88-106.

Zhu, F., Li, X., Valavi, E., & Iansiti, M. (2021). Network interconnectivity and entry into platform markets. Information Systems Research.

Appendix

Table A1. CEM: matching covariates

Question count from day -1 to day -30	Question count under associated tag from 1 to 30 days before bifurcation
Question count from day -30 to day -60	Question count under associated tag from 30 to 60 days before bifurcation
Total question up to day -30	Total question up to 30 days before bifurcation
Active percentage in day -30	Percentage of days with at least 1 question up to 30 days before bifurcation
Tag age (day) in day -30	Tag age up to 30 days before bifurcation
Answer count from day -1 to day -30	Answer count under associated tag from 1 to 30 days before bifurcation
Answer count from day -30 to day -60	Answer count under associated tag from 30 to 60 days before bifurcation
Total answer up to day -30	Total answer up to 30 days before bifurcation
Question growth rate	Question count (day -1 to day -30) minus Question count (day -30 to day -60)
Ratio of program questions	Ratio of programming related questions as 1 month before bifurcation

Table A2 Robustness check: The effect of bifurcation in home tag (after CEM)

	Log(Question) Home platform (1)	Log(Answer) Home platform (2)	Log(Contributor) Home platform (3)	Log(Accept16Hour) Home platform (4)	Accept16HourRatio Home platform (5)	Log(NewUser) Home platform (6)
bifurcation	-0.1095**	-0.1840***	-0.1598***	-0.1498**	-0.0311*	-0.0031
	(0.0540)	(0.0621)	(0.0564)	(0.0614)	(0.0165)	(0.0392)
R^2	0.0114	0.0146	0.0122	0.0157	0.0100	0.0022
N	3,104	3,104	3,104	3,104	2,703	3,104
N. of tags	97	97	97	97	96	97
Mean	2.37	2.76	2.70	1.71	0.42	1.15

* p<0.1; ** p<0.05; *** p<0.01. Heteroskedasticity robust standard error clustered at tag level in parentheses. Includes tag level fixed effects and weekly dummies.

Table A3 Robustness check: The bifurcation increases total user contribution across the combined home platform and spin-off platform but reduces the efficiency in knowledge exchange (after CEM)

	Log(Question) Home + spin-off platform (1)	Log(Answer) Home + spin-off platform (2)	Log(Contributor) Home + spin-off platform (3)	Log(Accept16Hour) Home + spin-off platform (4)	Accept16HourRatio Home + spin-off platform (5)	Log(NewUser) Home + spin-off platform (6)
Bifurcation	0.2501***	0.2404***	0.2519***	0.3447***	-0.1465***	0.1642***
	(0.0649)	(0.0784)	(0.0753)	(0.0844)	(0.0227)	(0.0481)
R^2	0.0717	0.0497	0.0548	0.0757	0.0597	0.0211
N	3,104	3,104	3,104	3,104	2,743	3,104
N. of tags	97	97	97	97	96	97
Mean	2.46	2.87	2.80	1.84	0.39	1.20

* p<0.1; ** p<0.05; *** p<0.01. Heteroskedasticity robust standard error clustered at tag level in parentheses. Including tag level fixed effect

Table A4 Robustness check: The existing users reduce contribution in the home platform and reallocate knowledge contribution to new topics (after CEM)

	Log(OldUserAnswer) Focal Tag Home platform (1)	Log(OldUserAnswer) Other Tags Home platform (2)	Log(OldUserAnswer) Old Knowledge Home + spin-off platform (3)	Log(OldUserAnswer) Old & New Knowledge Home + spin-off platform (4)
Bifurcation	-0.4699***	-0.0746**	-0.1843*	0.2980**
	(0.0990)	(0.0337)	(0.1069)	(0.1270)
R^2	0.4841	0.2988	0.4023	0.2907
N	3,104	3,104	3,104	3,104
N. of tags	97	97	97	97
Mean	2.14	6.23	2.25	2.36

* p<0.1; ** p<0.05; *** p<0.01. Heteroskedasticity robust standard error clustered at tag level in parentheses. Includes tag level fixed effects and weekly dummies.

Table A5 Robustness check: Tags with higher interconnectivity have lower level of decrease in user contribution in the home platform (after CEM)

	Log(Question) Home platform	Log(Answer) Home platform	Log(Contributor) Home platform	Log(Accept16Hour) Home platform
Bifurcation*HighIntercon	0.2363**	0.2720**	0.2320**	0.1832
	(0.0928)	(0.1051)	(0.0953)	(0.1118)
Bifurcation	-0.2202***	-0.3115***	-0.2685***	-0.2357**
	(0.0798)	(0.0959)	(0.0860)	(0.0969)
R^2	0.0199	0.0228	0.0187	0.0201
N	3,104	3,104	3,104	3,104
N. of tags	97	97	97	97
Mean	2.37	2.76	2.70	1.71
Linear combination	0.0160	-0.0395	-0.0365	-0.0525
	(0.0523)	(0.0499)	(0.0487)	(0.0590)

* p<0.1; ** p<0.05; *** p<0.01. Heteroskedasticity robust standard error clustered at tag level in parentheses. Includes tag level fixed effects and weekly dummies.

Figure A1. Robustness check: The estimated impact of bifurcation on the home platform contribution (after CEM)