ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Slivko, Olga

Working Paper

Peer effects in collaborative content generation: The evidence from German Wikipedia

ZEW Discussion Papers, No. 14-128 [rev.]

Provided in Cooperation with:
ZEW - Leibniz Centre for European Economic Research

Suggested Citation: Slivko, Olga (2015) : Peer effects in collaborative content generation: The evidence from German Wikipedia, ZEW Discussion Papers, No. 14-128 [rev.], Zentrum für Europäische Wirtschaftsforschung (ZEW), Mannheim

This Version is available at:
https://hdl.handle.net/10419/110597

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Discussion Paper No. 14-128

Peer Effects in Collaborative Content Generation: The Evidence from German Wikipedia

Olga Slivko

ZEW
Zentrum für Europäische Wirtschaftsforschung GmbH
Centre for European Economic Research

Discussion Paper No. 14-128

Peer Effects in Collaborative Content Generation: The Evidence from German Wikipedia

Olga Slivko

Previous version: December 22, 2014
This version: March 3, 2015

Download this ZEW Discussion Paper from our ftp server:

http://ftp.zew.de/pub/zew-docs/dp/dp14128.pdf

Die Discussion Papers dienen einer möglichst schnellen Verbreitung von neueren Forschungsarbeiten des ZEW. Die Beiträge liegen in alleiniger Verantwortung der Autoren und stellen nicht notwendigerweise die Meinung des ZEW dar.

Discussion Papers are intended to make results of ZEW research promptly available to other economists in order to encourage discussion and suggestions for revisions. The authors are solely responsible for the contents which do not necessarily represent the opinion of the ZEW.

Peer Effects in Collaborative Content Generation: The Evidence from German Wikipedia

Olga Slivko *

ICT Research Department, Centre for European Economic Research (ZEW)

Previous version: December 22, 2014
This version: March 3, 2015

Abstract

On Wikipedia, the largest online encyclopedia, editors who contribute to the same articles and exchange comments on articles' talk pages work in collaborative manner sometimes discussing their work. They can, therefore, be considered as peers, who are likely to influence each other. In this article, I examine whether peer influence, measured by the average amount of peer contributions or by the number of peers, yields spillovers to the amount of individual contributions. The partially overlapping groups in the peer network structure allow to identify peer effects and to use the number of the indirect peers as an instrument for the activity and the number of direct peers.

The results suggest the presence of modest peer effects in knowledge generation even in the absence of "friendship" ties. While controlling for observable editor and peer characteristics, an increase in the monthly average peer contribution by 1 per cent increases the amount of individual monthly contributions to Wikipedia, among individuals who contribute every month, by up to 0.1 per cent and are enabled by collaboration on article talk pages.

Keywords: Peer Effects; Network of Editors; Direct and Indirect Peers; User-generated Content; Wikipedia.

JEL Classification Numbers: D83, D85, J46.

*Address: L 7, 1, 68161 Mannheim, Germany, telephone: (0)621123-358, e-mail: *slivko@zew.de*. I am grateful to Irene Bertschek, Michael Kummer, Sisley Maillard, Frank Nagle, Marianne Saam, Michael Ward, Michael (Xiaoquan) Zhang and seminar and conference participants in Mannheim, Milan, Barcelona for helpful suggestions and comments. I appreciate helpful advice of Rodion Permin on the Wikipedia data processing. The financial support from the WissenschaftsCampus Tübingen is gratefully acknowledged.

1 Introduction

The emergence of participatory web applications based on digital technology transformed the users of online information into the active producers of knowledge (Lerner and Tirole (2002)). As a result, a significant amount of knowledge and open-source software generated on online platforms is produced by the participants of online communities. Prominent examples of such peer production communities are technical support forums (Stackoverflow, Quora), open source software (for example, the operating system Linux[1]) or the online encyclopedia Wikipedia.[2] The volunteer activity of individuals with heterogenous backgrounds results in a socially valuable output. Since Wikipedia appeared it demonstrated a new way to organize knowledge generation processes. The idea of such a platform was adopted by some firms with the aim to organize internal knowledge accumulation, although this proved to be challenging.

The voluntary provision of public goods on the Internet crucially depends on how effectively the large-scale human interaction systems will be designed in order to motivate voluntary participation. Benkler (2002) compares peer production with traditional production by firms in the markets and suggests gains from peer production in terms of information collection cost and improved allocation due to availability of large sets of resources, agents and projects. Recent economic research has advanced the understanding of the role of social motivation for contributions to Wikipedia (Algan et al. (2013), Zhang and Zhu (2010)). Zhang and Zhu (2010) find that the size of the recipient audience matters for the amount of knowledge contributed to Wikipedia. Algan et al. (2013) focus on the impact of social image and reciprocity for the size of charity donations to Wikipedia. My paper goes further in understanding how social mechanisms work on Wikipedia by analyzing whether the performance of peers has an effect on individual knowledge contributions. The empirical analysis

[1]Linux runs on more than 100K machines and 71M Linux users (LinuxCounter web-site)

[2]Wikipedia has over 1.8M users and 31.2M articles (Stats.wikimedia site). All data on the use of open source platforms are as in May 2014

is based on a sample that tracks contributions of more than 730 editors[3] on 518 pages in selected article categories during the period from January 2005 to January 2011. The full revision history allows one to identify the set of peers for each editor, which varies across articles. I construct the network of peers that are considered to be connected with each other if they contributed to the same article and commented on the talk page of the article. Using the panel structure of the data (editors' monthly contributions) and the structure of the editor network, I analyze whether there are spillovers to content generation by an individual from the amount of content generated by her peers or from the number of peers.

For identification of peer effects, I apply an econometric approach based on De Giorgi et al. (2010), which allows disentangling peer effects from exogenous characteristics of peers and correlated effects within groups. By groups, in which peers interact, I mean articles on Wikipedia written by editors in a collaborative manner. Using information on collaborative writing of articles, I construct the editor network, in which peer groups have partially overlapping structure. The econometric approach applied here takes advantage of this network feature. The property of partially overlapping groups enables variation of the group mean across individuals and thereby generates enough observations for the identification of the coefficient on peer effects. I use robustness checks to address several potential threats to identification. To disentangle the impact of exogenous shock to the content of articles I exclude from the sample all articles with characteristics of breaking news, recently created and very popular among contributors. Another potential problem, the endogenous network formation, is addressed as follows. Assuming that the arrival of new editors is related to the network of articles in Wikipedia (content driven) and is exogenous to the network of editors (since an individual has no instruments of visualization of the contributor network) I perform the same regressions as in the baseline model focusing exclusively on newcomers to see whether other editor activity has an effect on their contributions.

The results show that, while controlling for observable editor and peer characteristics,

[3]Hereafter, I will use the term "editors" or "contributors" for users who contribute voluntarily by editing articles on Wikipedia.

an increase in the average peer contribution by 1 per cent has a positive effect of 0.11 per cent on individual contributions. Interpreting this effect based on the median values for peer and individual contributions across all articles allows to evaluate the magnitude peer effects have on the content generated. An increase in peer contributions to all articles by 1 per cent would amount to approximately 1.2 kbytes, which it would correspond to a 0.11 per cent increase in individual contributions to all articles, or 1.04 byte. Assuming that 1,000 bytes could represent, approximately, half of a page of A4 format, a total increase in peer contributions on all articles by one page would yield the spillover of 2 bytes (about two characters) to individual contributions to Wikipedia. Similarly, spillovers from an increase in the number of peers contributing to the articles, yields a positive effect on individual contributions of 0.06 per cent. This evidence suggests that even in the absence of explicit "online-friendship" ties between individuals (as those established on Facebook and other social networking platforms) peer effects are present. These effects are both observed among individuals who contribute at least monthly to Wikipedia and also who have peers during some consequent time intervals, contributing to articles and engaging in discussions on the article talk pages. In addition, the amount of individual contributions is affected by an interest, or an expertise, in a special category of articles, which suggests the presence of the interest-based motivation for individual online contributions.

The robustness of obtained results is supported by a number of alternative specifications. Firstly, I apply an alternative measure for individual and peer contributions. Instead of using the amount of contributions in bytes, I use a number of revisions and find peer effects of a similar magnitude. Secondly, I address a potential threat to identification in networks, which is endogenous selection of individuals into networks. Fortunately, due to institutional features of the platform Wikipedia, the concept of "friendship ties" differs significantly from that of social networks. Therefore, as opposed to the studies of adoption through social networks (Aral et al. (2009)), individuals join the network of editors without observing the characteristics of their potential peers, and the only reason for homophily could be

common interests or expertise in the topic. I assume that editors choose Wikipedia articles to contribute randomly with respect to the intensity of their peers' edits on the rest of articles. Rather, they choose articles according to the topics of their interests or curiosity. While contributing to their first articles in Wikipedia, editors might get to know their peers and then decide to stick to them coordinating contributions with one another. Therefore, the only source of potential selection into the network could come from continuous contributing to some articles with some Wikipedians whom an individual has already met and talked to on the article talk pages. If this assumption is correct, I can exclude editors with more than one month's experience from the sample and concentrate exclusively on newcomers. Thereby, I examine the extent of peer effects on newcomer edits in the cross-section framework.

The results suggest that, during their first months on Wikipedia, editors are affected by their peers as well, with a similar magnitude to that in the baseline model. Overall, these results suggest that communications between most active community members encourages building-up and promoting new online communities and enhances knowledge generation in the existing on-line communities.

This paper is organized as follows. Section 2 presents a review of the relevant literature. Section 3 describes the data. Section 4 presents the econometric model. The main results are discussed in Section 5, and the robustness checks are presented in Section 6. Section 7 concludes.

2 Background and hypotheses

2.1 Preferences in contributions

Peer productive knowledge platforms can be distinguished in several important aspects. The specific feature of Wikipedia is the way in which content is generated. The content in Wikipedia can be very sensitive to the events happening outside it and, therefore, important instruments for enhancing attention spillover are exogenous shocks to the content (Kummer

(2013)). The newly created empty articles can be considered as signals to experienced contributors that there is a demand for that type of content (Gorbatai (2011)).

The organizational structure of content generation, potential rewards, and the usage of output in Wikipedia differ significantly from open source software as well. While the output of open-source projects is often aimed at sophisticated users, an online encyclopedia has a high value for the vast range of users, therefore representing a public good. Due to the modular structure, little communication between developers of open-source software is needed. On the contrary, encyclopedic content is sometimes a subject of discussion between contributors with several contradicting opinions. Since any revision can be reverted, contributors have to agree on the content (explicitly or implicitly) in order that the content remains on the page for a longer time. In contrast to open-source projects where monetary incentives are implicitly present through the future expectations of project participants for better-paid jobs, in Wikipedia, social and psychological incentives (reciprocity, socialization) can instead play a very important role (Osterloh and Rota (2007), Algan et al. (2013)).

Contrary to social networks, Wikipedia does not have explicit friendship ties. Individuals become peers in the process of collaborative content generation. Do social effects, nevertheless, matter on Wikipedia, given such a structure? Studies focusing on Wikipedia point out that when the group of individuals is sufficiently large, private benefits dominate free-riding incentives, thus enabling the provision of a public good (Zhang and Zhu (2010)). Voluntary contributions might breed recognition in the community or improve social image of an individual (Lacetera and Macis (2010), Algan et al. (2013)), or contributions might be affected by the feeling of reciprocity. Algan et al. (2013) find that reciprocity matters for donations to Wikipedia, while Shriver et al. (2013) and Harper et al. (2010) find this phenomenon in other social networks, correspondingly, for wind-surfing and movielens. However, to the best of my knowledge, there is still no analysis of an impact which peers might have on individuals regarding the amount of contributions. Peer effects arise when individuals interact in groups and the average outcomes of peers affect individual outcomes. The present study fills this

gap in the literature by showing that the interactions with other editors indeed matter for the core of the most productive contributors.[4]

Theoretical and empirical studies provide confronting views on the mechanisms that underlie the success or the failure of productive online communities. On the one hand, individuals contributing to online communities might have incentives to free-ride, meaning that as a group expands, individual contributions would decline (Andreoni (1988), Bilodeau and Slivinski (1996)). In these models, a contributor receives utility from the total provision and her private consumption of a public good. With an increase in the group size, an average contribution level falls to zero and only individuals with the lowest costs of contributing or the highest income will contribute. According to Andreoni (2007), an individual's utility depends also on the number of recipients of the public good. When the recipient group size is sufficiently large, the relative importance of private benefits, as compared to free-riding incentives, dominates and positively affects individual contributions.

In the case studies of successful open source software projects, Lerner and Tirole (2002) stress the importance of a new organizational structure, which requires low capital investments to the projects and relies on the collaboration between individuals. In the project Apache, the organizational structure that enables success of the project is represented by the core of responsible editors and a large number of volunteer participants.

The empirical literature on Wikipedia suggests individual interests and/or expertise as one of the main reasons for contributions. Panciera et al. (2009) show that only a small fraction of editors, so-called "Wikipedians", contribute more intensely than others from the moment of their initiation, and all contributors reduce activity over time, with the only distinction being that "Wikipedians" end up at higher levels of contribution. Nov (2007) surveys Wikipedia contributors and finds that the top motivations were "Fun" and

[4]In the recent economic literature, the influence of peers on individual behaviour has been already addressed in a number of contexts, for instance, in individual decisions on housing area (Hanushek et al. (2003)), schooling or degree (De Giorgi et al. (2010)), health attributes such as obesity or smoking (Fowler and Christakis (2008)). The definition of peers also differs depending on the context. Peers could be individuals who interact in groups while studying (school mates or students), live in the neighborhood or produce some output together (co-authors, colleagues, open-source software developers).

"Ideology" (individuals support open-source). Laniado and Tasso (2011) find the presence of a nucleus of very active contributors who spread their contributions over the whole of Wikipedia, and interact with inexperienced users. In this case, individual preferences would affect the amount of contributions to Wikipedia. Together, these findings provide a strong support to the following hypothesis:

Hypothesis 1. An interest in a specific topic, or an expertise in it, positively affects individual contributions to Wikipedia.

Contributions can also be induced by the characteristics of Wikipedia articles. For instance, Keegan et al. (2012) suggest that pages that appear due to some exogenous shock ("breaking news") initially experience different patterns of contribution, with highly clustered and centralized editors' interactions. In their approach, tighter collaborations are rather caused by shocks to pages. To avoid capturing the impact of exogenous shocks to pages, I exclude pages that have breaking news properties and control for the fact that the page to which an editor contributed is among the most popular articles on Wikipedia during a given time period. Aaltonen and Seiler (2014) suggest that the article size, which is a measure for accumulated editor activity, triggers further contributions due to knowledge spillovers. Therefore, I control the page size in order to capture this potential source of spillover. Overall, the above-mentioned studies suggest an impact of exogenous editor and page characteristics on contributions to Wikipedia.

2.2 Existence and nature of peer effects

There is a range of studies that examine the existence of potential peer effects in social networks and Q&A forums. Bapna and Umyarov (2012) show that, on Spotify, an exogenous adoption of a premium subscription by peers increases individual adoption by about 50%. Notably, this effect is stronger for users with fewer friends. Hahn et al. (2008) study collab-

oration ties in open-source software development projects and show that prior collaborative ties and the perceived status of project members in the network matter for developers' choosing to join new projects. Shriver et al. (2013) use the variation in wind speeds at surfing locations in Switzerland as an exogenous shifter of content generation about surfing activity onto an online social network. The local network effect in content generation is suggested to cause an increase in content and, as a result, stronger ties between users, which, in turn, breeds more visits and browsing on the website. Moon and Sproull (2008) highlight the role of feedback in producing and sustaining high-quality contributions: in groups where systematic quality feedback systems are implemented (for example, a rating system) question askers return over a longer duration, answer providers contribute more often.

Several empirical studies on Wikipedia reexamine the existence of social effects for the case of an online encyclopedia where neither explicit friendship ties nor organizational structure are present. In Wikipedia, the size of the potential recipient audience matters. When the group of individuals is sufficiently large, private benefits from contributing to a public good dominate free-riding incentives (Zhang and Zhu (2010)). Another reason is that voluntary contributions breed recognition in the community or improve social image of individuals (Lacetera and Macis (2010), Algan et al. (2013)). Together with the social image, the feeling of reciprocity to peers (expectation that they will also contribute if she does) positively affect individual money donations to Wikipedia (Algan et al. (2013)). These reasons are also supported by psychological literature (Burke et al. (2010); Kittur and Kraut (2010); Faulkner et al. (2012)), documenting that, in Wikipedia, numerous direct communications occur on user-talk pages and talk pages of articles. These studies describe socialization strategies of individuals in online communities, including requests for participation or information and expressions of similarity to others. Their findings suggest that personalized moderation is effective in order to increase the number of contributing members and their commitment, while community-level moderation increases commitment alone.

There are several studies that are closest to the present study in that they analyze the

mechanism underlying collaborations on Wikipedia. Gorbatai and Piskorski (2012) suggest that editors involved in high-density structures in the network of editors are less likely to abandon contributing.[5] Gorbatai (2011) proposes to consider collective contributions to an online public good in the absence of price mechanisms as the following three-stage process. Firstly, consumers express the demand for the public good by occasional contributions. Then, at the third stage, producers observe the unsatisfied demand for knowledge and become willing to improve these collective goods. In addition to the demand-supply model, social effects in Wikipedia have been addressed in the two articles mentioned earlier: Algan et al. (2013) and Zhang and Zhu (2010). However, until now, not much has been known about peer effects in Wikipedia and their role in motivating individual contributions.

In this paper, I examine another potential factor of social influence on contributions, i.e. the effect of peer performance on individual performance. In sociological literature, Sassenberg (2002) suggests that individuals may feel psychologically connected to a group and hence act according to the norms and the standard behavior of the group. Moreover, social learning theory argues that individuals follow the behavior of relevant peers if they face uncertainty about norms, as this strategy maximizes their expected payoffs given the chosen strategy (Bercovitz and Feldman (2008)). There are also a number of education studies (De Giorgi et al. (2010), Contreras et al. (2012)) that suggest the presence of peer effects on the individual performance. In line with previous studies, I expect that individuals involved into contributing to Wikipedia observe their peers' activity and, in response, change their activity. As a result, peer activity could positively affect individual contributions in Wikipedia.

Hypothesis 2-1. The amount of individuals' contributions is possibly affected by the average amount of peer activity.

[5]The two editors are connected in the networks if they contributed to the same article and posted messages on the article's talk page within one month.

Hypothesis 2-2. The amount of individuals' contributions is possibly affected by the number of peers.

This paper adopts the econometric framework for peer effect analysis, which was developed in the empirical studies of academic performance (Contreras et al. (2012)), researcher collaboration with industry (Kacperczyk (2013), Aschhoff and Grimpe (2014)), career choices (De Giorgi et al. (2010)), and health-related attributes, such as obesity, smoking (Fowler and Christakis (2008)). This methodology is based on partially overlapping groups of peers (De Giorgi et al. (2010); Contreras et al. (2012)). De Giorgi et al. (2010) present an empirical analysis of students' choices of major (Economics or Business) as affected by their peers' choices after controlling for individual characteristics of students (age, gender, schooling grade). The characteristics of excluded peers (for an individual, the set of peers who are in the same groups with her direct peers but unconnected directly with her) are used as instruments. A two-stage least squares estimator is used to find the peer effect (the choices of peers) on the outcome (students' own choices of major between Economics and Business). Contreras et al. (2012) study the peer influence on students' grades in the public University College of Business at the Unites States. In order to estimate the endogenous peer effect, they use the exclusion restriction approach (similar to De Giorgi et al. (2010)). They find that a student's classroom performance has a significant demotivating effect on her peers. Furthermore, they classify excluded peers by ability on four groups according to percentiles and examine their effect on low-ability and high-ability students' performance. Low-ability excluded students are shown to have a negative effect on other students. At the same time, high-ability excluded students have a negative effect on low-ability students, while high-ability excluded students have a positive effect on high ability students. Hanushek et al. (2003) also investigate peer effects on student achievements. In order to separate peer effects from other confounding influences and to address the reciprocal nature of peer interactions, they apply past achievement as a measure of peer group quality.

In the case of Wikipedia, I use a definition of peers according to which editors are getting connected by contributing to Wikipedia articles together within a short time period. The composition of peer groups of an individual varies across pages. This gives rise to partially overlapping peer groups, which are the key to solve the "reflection problem" (Manski (1993)). The excluded peers of an editor are those editors who do not collaborate with her directly but work together with her direct peers on other articles.

3 Data

The dataset is obtained from a publicly available dump of the German Wikipedia provided by Wikimedia Deutschland. It is currently the second largest Wikipedia and accounts for about 1,500,000 articles. The dump contains meta-information on articles' revisions including the time stamps and the contributors' identifiers. The empirical analysis is based on the sample, which tracks contributions of more than 730 editors (in some estimations, up to 1649 editors) on 518 pages in some selected categories of Wikipedia articles during the period from January, 2005 to January, 2011. To reduce the size of the data set, I use the meta revision history only for articles in the following categories:[6] Alcohol, Astrology, China, Druids, Economics, India, Islands, Medicine, Reptiles, Soccer. The data identify contributors who edited articles at given moments. They enable constructing an editor network where editors are connected by contributions to the same articles and comments on the articles' talk pages.

Some contributions in Wikipedia are made anonymously and so they are identified in Wikipedia by the IP addresses of the contributors. Since the contributions of the same editor in Wikipedia's revision history might have different IP addresses, they provide misleading information on the intensity of contributors' monthly activity and are excluded from the data sample. Bots, e.g. automated scripts, can be identified from the data and are also excluded from the editors' network. Overall, the final sample contains only registered users with at

[6]The meta data dump does not contain the information about article categories. The tree of article categories should be additionally extracted and processed, therefore, this study uses only the categories available in our database.

least five edits on Wikipedia and at most 3,000 monthly edits, which allows the exclusion of occasional contributors and overly active contributors who might be remunerated for their contributions or unregistered automated bots. Furthermore, in order to avoid taking into account vandalism (and consequent reverts of the pages) I exclude from the editing history those revisions for which the revision length varies from some positive value to zero and back to the same positive value within the consequential periods. The analysis of individual and peer activity amounts involves only contributions to the articles in the main Wikipedia name space.

3.1 Dependent variables

In order to measure the activity of individual contributors on Wikipedia, this study considers the logarithms of the total number of bytes changed, which is the sum of the absolute values of bytes added and deleted. All measures of individual activity are computed as the activity of individual i on page j in time t for the analysis at the editor-article level, and then aggregated for individual i at time t in the analysis at the editor level. I also examine the robustness of my measures in capturing editor activity by using the number of revisions made by individual i at time t at the editor level in the alternative specifications.

3.2 Independent variables

3.2.1 Peer effects

Peer effects can arise from interaction with peers. Individuals are likely to observe their peers' activity as measured by the total peer contributions. Consequently, peer effects can be captured by the average amounts of peer contributions measured in bytes or by the number of peers. The definition of peers rests on the collaboration mechanisms provided in Wikipedia. Beyond contributing to the same article, editors can leave messages on the talk page of each article. Therefore, my measure of peers relies on co-authorship of an article in Wikipedia and

coordination involving talk pages of each article. More precisely, two editors are connected on the article if they have collaborated on it within a month (four weeks) *and* left comments on the talk page of the article. In order to bring the definition of links between editors closer to the notions of collaborative content generation, I consider two editors as peers only if they made at least one revision of an article and one revision of the article's talk page during a month. Once the link is set up, it expires in four weeks unless both editors contribute to the article again in the next period. Then, monthly snapshots of the editor network are taken to construct the final data set. This definition of peers is similar to the definition in the study of academic entrepreneurship (Aschhoff and Grimpe (2014)), where co-authors of academic papers are regarded as peers. This definition also considers tighter collaborations between individuals than Gorbatai and Piskorski (2012). The average amount of peer contributions is a weighted average, where the weights are defined by the intensity of communication, i.e. the product of the number of revisions made by the two connected editors.

The definition of the editor network in Wikipedia ignores occasional contributors who make revisions once and never come back (less than 2% of the initial sample of contributors), such that only contributors with more than five revisions during the years 2005 - 2011 are in the sample. Registered bots (the editor accounts which are registered on Wikipedia as automated programs) as well as editors with suspiciously high numbers of monthly edits[7] (which could be unregistered bots) are excluded from the sample to avoid blowing up the human activity on Wikipedia.

3.2.2 Editor characteristics

The independent variables are characteristics of editors and articles that can be extracted from the revision history dump of German Wikipedia.

The editor characteristics are the most important control variables. From the data, I can compute the editor experience measured as the length of the period in months since

[7]We assume that a human being cannot contribute to German Wikipedia more than 9,000 edits within four weeks (less than 1% of the initial sample of contributors).

the individual's first contribution to Wikipedia. It captures the impact of the editor's life cycle on Wikipedia. Further, I can infer to which article category (from available Economics, Medicine, Alcohol, Astronomy, China, Druids, India, Reptiles, Soccer)[8] an individual contributed mostly and what is the share of his contributions to this category in his total contributions to Wikipedia. The share of contributions to the most interesting category indicates how specialized is the interest of an individual in one specific topic, which could explain the size of the contribution to Wikipedia.

Editors connected in the network might share interests; for example, they might be both interested in reading and contributing to articles about famous economists. Empirical literature documented that individuals tend to associate with one another due to homophily, i.e. similarity in characteristics, such as gender, age, education and religion. In Wikipedia, among scarce individual characteristics observed, one particularly important characteristic that can be extracted from the data I dispose is individual interest or expertise in the topic (or article category). I define interest, or specialization in the topic, by the share of the editor's contributions among her total contributions in the category to which she contributed most.

Additionally, I control for the potential preference of the editor to contribute only to very popular articles. To account for the behavior according to which an editor usually browses popular articles and sometimes introduces minor changes, I compute the share of popular articles in the articles to which the editor contributed. Five per cent of articles that got the highest editing activity over a given period are assumed to receive high attention due to clicks and, therefore, popular articles. The share of these articles in the total number of the editor's articles accounts for the preference for popular topics rather than a specialized interest or expertise in the topic. Individuals who often browse Wikipedia's most popular pages [9] might contribute small pieces of knowledge or correct typos. Then, such a behavior

[8]Articles in these categories (except Soccer) are not likely to experience everyday updating activity. As the articles about Soccer are likely to experience constant updating about recent events, in some further robustness checks I exclude all contributions and links of this category from the sample, and this doesn't affect the obtained results. These further robustness checks, excluding Soccer category, are available upon request.

[9]For instance, the starting page of Wikipedia every day advertises a new article of the day

could be a potential reason for generally smaller contributions to Wikipedia.

3.2.3 Article characteristics

An important characteristic that could be extracted from the data dump in the absence of the full-text revision history is the average page size in kilobytes during each month. Aaltonen and Seiler (2014) suggest that a page needs to grow to a certain size in order to attract intensive editing activity. According to this finding, I would expect individuals to contribute more to longer articles. Conversely, the size of the article can thwart someone's adding further information once the article is a nearly complete. Then, I would expect a negative effect of the article length on individual contributions to this article.

Furthermore, extensive attention to some emergent topic in the media (e.g. release of a new mobile application) may cause the creation of a related page on Wikipedia as well as a high number of clicks on the article and articles related to the topic that are connected by hyperlinks. More attention, as measured by clicks to an article, would breed more edits to the pages. Therefore, following the description of individual behaviour on pages that are breaking news in Keegan et al. (2012)[10] I exclude all activity on breaking news pages from the sample.

Table 1 displays summary statistics on the editor contribution size, editor and article characteristics for the data used for the analysis at the editor-article level. As the values of the variables for individual and peer contributions are skewed, in the regressions I normalize them by using natural logarithms of each value plus one, to preserve observations with a value equal to zero. From table 1, one can notice that most pages in the final sample are quite long (for example, 12 kbytes at the 25th percentile) and are mostly those articles attracting high editing activity (more than 50% of articles in the sample are those attracting the number of edits above 99th percentile over the whole of German Wikipedia). Similarly,

[10] As in Keegan et al. (2012), I define breaking news pages as recently created articles that attract a higher attention (articles with a number of edits above 95th percentile over the time period) during the first month since their creation.

the descriptive statistics for the variables and observations used in the analysis at the editor level are presented in table 2.

Table 1: Descriptive statistics at the editor-article level

	Mean	S.D.	Minimum	25th	50th	75th	90th	Maximum
Contribution per article (bytes)	4608	19200	0	270	1047	3423	9657	749872
Peer contrib. per article (bytes)	1989	2433	34	745	1355	2413	3997	40179
# revisions per article	11	21	2	3	5	11	23	425
Peer # revisions per article	3.3	2	1	2	2.8	4.2	5.7	34
Interest in the category (%)	14	15	.063	2.8	8.2	22	37	93
Editor experience (months)	27	20	.059	10	22	39	56	109
Edits on popular pages (%)	53	20	0	39	52	67	80	100
Page size (bytes)	36615	35757	0	12045	25236	49944	80266	232921
Page is popular	.7	.46	0	0	1	1	1	1
# editors per page	12	13	2	5	8	14	22	157
# peers per page	2.7	2.8	1	1	2	3	6	24
Av. # indirect peers per page	1	.038	1	1	1	1	1	2.3
Av. # indirect editors per page	10	11	2	5	7.4	11	17	157
Observations	4407							

NOTES: The table shows the descriptive statistics for all monthly observations of the editing activity on Wikipedia used in the regressions at the editor-article level.

Table 2: Descriptive statistics at the editor level

	Mean	S.D.	Minimum	25th	50th	75th	90th	Maximum
Contribution all articles (bytes)	4336	18690	0	226	946	3300	9291	750582
Peer contrib. all articles (bytes)	196010	270886	0	49354	122135	243769	434944	5385174
# revisions of all articles	10	19	2	3	5	10	22	425
Peer # revisions to all articles	382	375	1	126	278	514	842	2971
Interest in the category (%)	10	13	.021	1.8	4.3	13	30	93
Editor experience (months)	28	20	.0031	12	25	42	58	109
Edits on popular pages (%)	47	20	0	33	45	60	75	100
# peers on all pages	3.3	3.9	1	1	2	4	7	63
Av. # indirect peers on all pages	3.2	3.7	1	1	1	4	7.6	52
Page size (bytes)	27989	32654	0	6546	17079	36918	67612	273676
Page is popular	.5	.49	0	0	.53	1	1	1
Observations	8492							

NOTES: The table shows the descriptive statistics for all monthly observations of the editing activity on all articles on Wikipedia used in the regressions at the editor level.

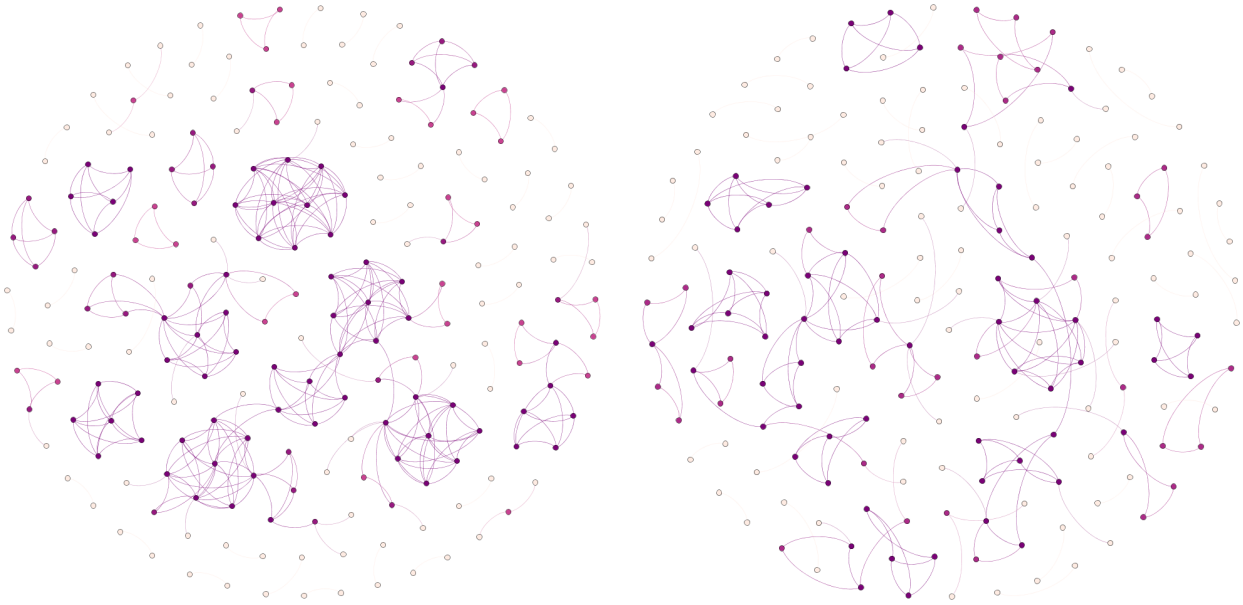
4 Empirical analysis

4.1 The network of peers

Wikipedia articles have talk pages, which provide the contributors with a mechanism for coordinating their efforts. When a disagreement on an article's content or layout arises, editors can have a discussion on this article's talk page. Therefore, my definition of peers is based on the co-authorship of Wikipedia articles involving communication on the talk pages. Precisely, *individuals are considered to be connected in the network of editors on Wikipedia if they contributed to the same article and commented to the talk page of the same article within a short time span.* This definition is meant to capture the network of contributors to Wikipedia who work collaboratively on new content generation and interact with each other. In order to bring the definition of links closer to the notions of collaborative content generation, two editors are considered peers only if they make at least *one revision of an article* and at least *one revision on an article talk page* each during this time period. The time span taken in this paper is equal to four weeks. Once the link is set up, it expires in four weeks unless both editors contribute to the article in the next period as well.

The network of editors obtained under such a definition is depicted in the two consequent periods, with high number of observations, in figures 1a and 1b. The nodes represent the contributors and the edges are the pages which they edited together. The nodes are colored according to their degrees with darker nodes standing for larger numbers of collaborators per contributor. The edges between contributors are coloured according to the intensity of collaboration measured by the number of pages they jointly edited. The figures show that there is no evidence of stable productive clusters where the most productive editors in every period collaborate with selected counterparts.

The network of editors considered in this study is defined similarly to Jackson and Wolinsky (1996), where the finite set of players $N = 1, 2, .., n$ are connected in the network and are represented by the nodes. Their pairwise relations are represented by the arcs of the network.



(a) Jan, 2007 (b) Feb, 2007

Figure 1: The network of editors in Wikipedia from the used sample, displayed in two consecutive months of observation

Network G can be expressed by an NxN adjacent matrix, and $g_{i,j}$ is a link between nodes i and j. It takes the value 1 if nodes i and j are connected, and 0 otherwise.[11] In what follows, the set of links of a node i will be denoted by G_i. The equilibrium in such a network is based on the concept of pairwise stability proposed by Jackson and Wolinsky (1996), meaning that the link is formed if both parties involved are consent, while a unilateral decision is needed for the link severance.[12]

The effect of peer contributions on the performance of focal editors could be analyzed on two levels. First, I analyze article-specific peer effects, i.e. the productive pressure experienced by an individual from her peers on a particular article. This peer effect would indicate how an individual activity on this article would change if she met there more active peers on the article. This average "activeness" of the individual's peers on an article, according

[11]Note that $g_{i,i} = 0$ and $g_{i,j} = g_{j,i}$ by definition.

[12]See Jackson and Wolinsky (1996) and Bloch and Jackson (2006) for more details on equilibrium stability and efficiency.

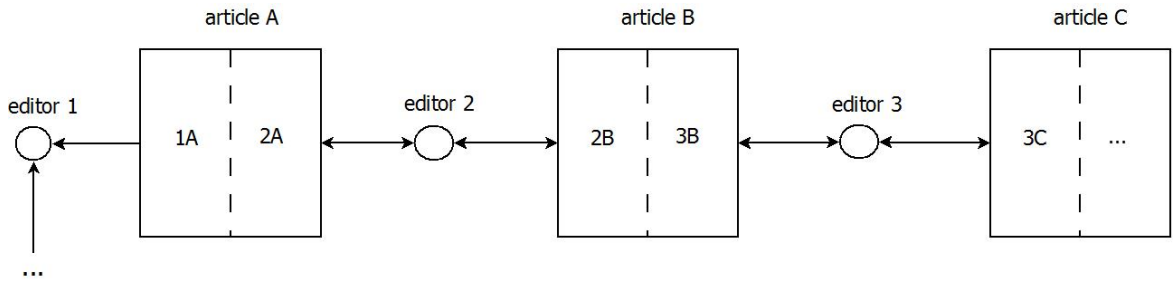


Figure 2: The network of editors connected due to collaborations on Wikipedia articles: article-specific peer effects

to the linear-in-means model (Manski (1993)) described below, would be expressed by the average peer contribution across articles other than the focal article. The structure of the peer network on Wikipedia is displayed in Figure 2. Editors are denoted by numbers above the circles while articles are denoted by letters within the squares. Each editor, say, editor 1, has a set of direct peers with whom she is connected due to collaboration on article A. This set of peers varies across articles for each individual. If there was a peer effect, the activity of 1's peers on article A would be affected by the average activity of editor 2 on article B due to interactions on the same article A.

The peer pressure mechanism might function in a different way. A contributor might observe her most important peers on the set of articles to which she contributes. The interaction with more engaged peers might affect this contributor in a way such that she feels also more engaged with Wikipedia and checks more articles in order to add some more content. As opposed to the first mechanism, once the article where the editor is currently working is filled with information, she might find it reasonable to switch her effort to other articles. Therefore, beside an analysis of article-specific peer effects, the potential peer effect should be also analyzed as the impact of average peer total contributions on individual total monthly contributions to Wikipedia. Figure 3 displays the corresponding data structure. Here, I consider editors connected due to collaborations on some sets of articles. The overall peer effect then would be expressed as how an individual activity of editor 1 on all articles, which in our example consists of contributions only to article A, would be affected by the

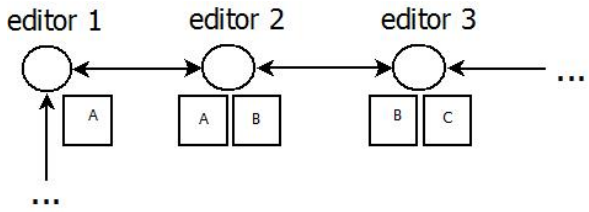


Figure 3: The network of editors connected due to collaborations on Wikipedia articles: overall peer effects

average contributions from editor 2 to all her articles except article A, in this example contributions to article B.

4.2 Econometric methodology

To address the research question whether peers' productivity affects contributors' outcomes in Wikpedia, I adopt the linear-in-means model introduced by Manski (1993):

$$y_{ijt} = \alpha_{ij} + \beta E(y|G_{it}) + X_{it}\gamma + E(X|G_{it})\delta + Z_{jt}\theta + \epsilon_{ijt} \tag{1}$$

where a contribution of editor i on article j at time t is affected by the average amount of peer contributions ($E(y|G_{it})$) as well as by the vector of her peers' exogenous characteristics ($E(X|G_{it})$), and G_{it} denotes the peer group of an individual i at time t. This can be rewritten as:

$$y_{ijt} = \alpha_{ij} + \beta \frac{\sum_{k \in P_{-ijt}} y_{k-jt}}{N_{P_{-ijt}}} + X_{it}\gamma + \frac{\sum_{k \in P_{-ijt}} X_{kt}}{N_{P_{-ijt}}}\delta + Z_{jt}\theta + \epsilon_{ijt} \tag{2}$$

where y_{ijt} is the logarithm of the contribution length (in bytes) by editor i on article j at time t and X_{it} is the vector of characteristics of editor i. $\beta \frac{\sum_{k \in P_{-ijt}} y_k}{N_{P_{-ijt}}}$ is an endogenous effect of peers' productivity (measured as a logarithm of the average amount of peers' contributions), where $k \in P_{-ijt}$ is a member of individual i's peer group composed of $N_{P_{-ijt}}$ members. $\frac{\sum_{j \in P_{it}} X_{jt}}{N_{P_{it}}}$ is an exogenous or contextual effect of peers characteristics and preferences on the individual outcomes, aimed at capturing a homophily, i.e. the property capturing that

the connected individuals can be similar in some observed characteristics, such as interests or experience. Finally, Z_{jt} is the vector of observable article characteristics (or, in the terminology of education studies, group characteristics).

If the coefficient β is positive, equation 2 shows the extent to which an individual editor is willing to contribute more to an article if her peers also contribute more on average. In Wikipedia, it is technically possible to check who are the peers in the revision history of an article, and then one can go further by clicking on any peer in order to check how active she has been. The latter effect is captured by equation 2. However, the former action is less technically sophisticated than the latter. In some specifications of the model, I also check this former mechanism. Concretely, I examine whether spillovers due to a higher number of peers affect individual performance. Then, the model estimated is:

$$y_{ijt} = \alpha_{ij} + \beta N_{P_{-ijt}} + X'_{it}\gamma + \frac{\sum_{k \in P_{-ijt}} X_{kt}}{N_{P_{-ijt}}}\delta + Z'_{jt}\theta + \epsilon_{ijt} \tag{3}$$

Peer pressure might also be important for the overall level of the engagement in knowledge generation on the Wikipedia platform. Once the article where the editor is currently working is filled with information, she might find it reasonable to switch her effort to other articles. Therefore, beside an analysis at the editor-article level, the potential peer impact on individual contributions is also analyzed at the level of overall individual contributions per time period aggregated across articles. Then, the empirical model is given by:

$$y_{it} = \alpha_i + \beta \frac{\sum_{k \in P_{-it}} y_{k-jt}}{N_{P_{-it}}} + X'_{it}\gamma + \frac{\sum_{k \in P_{-it}} X_{kt}}{N_{P_{-it}}}\delta + \epsilon_{it} \tag{4}$$

The positive peer effect in this model would indicate that there are positive spillovers due to collaboration with other contributors that affect an individual motivation to provide more knowledge to Wikipedia overall. Similarly to the editor-article level, at the editor level peer effects can also be expressed through the number of peers on Wikipedia across all articles, analogically to equation 3.

4.3 Identification issues and instrumental variables

In the linear-in-means model, the "reflection problem" and correlated effects are usually considered the major threats to identification of peer effects (Manski (1993)). Since the network structure in Wikipedia is based on partially overlapping peer groups, this solves the reflection problem and allows the identification of peer effects. Then, correlated effects (the shocks that are common to groups, in the context of Wikipedia to articles) could be addressed by using characteristics of indirect peers as instruments for endogenous outcomes of direct peers (as discussed in Bramoullé et al. (2009), De Giorgi et al. (2010)). In the case of Wikipedia, these could be shocks of attention to article content. To eliminate the impact of these shocks, I use the number of indirect peers (and its second order polynomial) as an instrument for the peer effects coming from direct peers.

The most important concern in the analysis of peer effects is the potential endogeneity of the network formation. This problem arises since individuals choose endogenously counterparts with whom they become peers. In Wikipedia, individuals come to read articles and their decision to contribute is most likely related to the content of an article rather than because of other editors' characteristics. However, individuals can hardly observe other contributors' individual characteristics because few contributors have extensive information in their user profiles and they are at least three clicks away from the article itself. What individuals mainly observe when entering Wikipedia are articles, their length, quality and how well they cover the topic. Therefore, while attraction of readers and potential contributors is not random with respect to the network of articles (more popular articles have more hyperlinks and, therefore, are more central in the hyperlink network, which, in turn, attracts more readers) it can still be initially considered random with respect to the network of editors. Later, when contributing to articles and observing the contributions of others, individuals can choose whether to remain peers with those other contributors. However, learning about "key" productive users takes some time and some reactions be it on Wikipedia articles or on articles talk pages. Under the assumption that individuals enter in the editor network

without prior knowledge about this network, I make a robustness check and examine the peer impact on individuals during only their very first month on Wikipedia.

Finally, in the case of an online community, such as Wikipedia, individuals might engage in discussions on article talk pages or in "editing wars". This activity is directly caused by the personal appeal and is beyond peer effects in performance. Therefore, for the direct peers of an individual the average amount of contributions excludes the page shared with this individual.

5 Results

This section discusses the main results of the analyses of peer effects at the editor-article and editor levels presented in tables 3 and 4. The first stage regressions for all tables containing IV estimations are in the Appendix (see tables 8 and 9). All results in the tables include year and month dummies and heteroscedasticity robust standard errors in the parentheses.

The instruments significantly affect the endogenous regressor in the first stage estimation and have large partial F statistics for testing the weakness of instruments (Kleibergen-Paap or Wald rk statistics Kleibergen and Paap (2006)), varying from 18 to 68 for the editor-article level and 60-355 for the editor level. Tables 3 and 4 represent specifications with ordinary least squares (columns 1 and 4) and fixed effects (columns 2 and 5) estimations. Columns 3 and 6 in each table contain specifications, in which peer effects are estimated using the instrumental variable approach. In each specification, I examine peer effects using one of the peer activity indicators, the log of average amount of bytes contributed or the number of peers.

The IV specifications do not reveal any significant impact of average peer contribution on the number of peers on the individual per article contributions (table 3). According to the "endogeneity test", also called Sargan-Hansen J-test[13], both peer activity indicators should

[13]The corresponding χ^2 test statistics ranges from 0.933 to 1.077.

Table 3: Peer effects at the editor-article level

	Log length of contribution (in bytes)					
	OLS	FE	2SLS	OLS	FE	2SLS
Log peer contrib. per article (bytes)	0.255***	0.241***	0.586			
	(0.040)	(0.052)	(0.364)			
# peers per page				0.117***	0.133***	0.361
				(0.017)	(0.023)	(0.228)
Interest in the category (%)	0.014***	0.048***	0.049***	0.014***	0.048***	0.050***
	(0.002)	(0.008)	(0.009)	(0.002)	(0.008)	(0.009)
Editor experience (months)	-0.010***	-0.013	-0.019	-0.010***	-0.047	-0.112
	(0.002)	(0.147)	(0.151)	(0.002)	(0.147)	(0.166)
Edits on popular pages (%)	0.009***	-0.002	-0.002	0.009***	-0.002	-0.003
	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
# editors per page	0.009***	0.023***	0.021***	-0.009**	0.003	-0.033
	(0.002)	(0.004)	(0.004)	(0.004)	(0.005)	(0.036)
Page size (bytes)	-0.000**	-0.000***	-0.000***	-0.000	-0.000**	-0.000**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Page is popular	0.350***	-0.386	-0.385	0.338***	-0.338	-0.255
	(0.075)	(0.290)	(0.300)	(0.075)	(0.287)	(0.304)
Peer interest in the category (%)	0.000	0.002	0.003	0.000	0.002	0.004
	(0.002)	(0.003)	(0.004)	(0.002)	(0.003)	(0.004)
Peer experience (months)	0.002	0.005	0.006*	0.001	0.005	0.006*
	(0.002)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
Peer edits on popular pages (%)	0.000	-0.001	-0.007	0.004**	0.002	0.000
	(0.002)	(0.003)	(0.006)	(0.002)	(0.003)	(0.003)
Observations	4407	4407	4398	4407	4407	4398
Kleibergen-Paap Wald F statistic			52.03			14.71

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

NOTES: The table shows the results of the reduced form regressions to estimate peer effects. Columns (1)-(3) show the results for the peer average contribution and Columns (4-6) for the number of peers. Specification (1) and (4) show OLS results; (2) and (5) show FE results. In Columns (3) and (6), I assume that peer effects are endogenous and estimate them in two steps. All regression coefficients are presented with heteroscedasticity robust standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The unit of observations is a monthly contribution of an editor to an article on Wikipedia. All month and year dummies are included.

not be considered endogenous. Most likely, this is because the instrument performance in the first stage is rather poor. Therefore, I rely on the OLS and FE estimation results for this model. For the editor contributions per article, a one per cent increase in average contributions of peers on that article would increase individual contributions by 0.24 per cent, while one more peer on the article would yield a positive effect of 0.13 per cent. The results suggest that the personal interest in the article topic matters for the amount of contributions as well as spillovers coming from the vast number of editors who also edited the article. These editors are not peers, their number is exogenous to the focal individual and might reflect the general level of attention to the article, which is not captured by the number of clicks (readership).

As compared to contributions per article, overall individual contributions to Wikipedia experience strong peer effects (see table 4), which are lower in magnitude but seem to be more robust and better identified due to the powerful instruments. The results show that, while controlling for observable editor and peer characteristics, an increase in the average peer contribution by 1 per cent has a positive effect of 0.11 per cent on individual contributions. Taking the median values for peer and individual contributions across all articles from table 2, we can interpret the peer impact as follows. An increase in peer contributions to all articles by 1 per cent would amount to, approximately, 1.2 kbytes, and it would correspond to a 0.11 per cent increase in individual contributions to all articles, or 1.04 byte. Assuming that 1,000 bytes are, approximately, half of a page of A4 format, a total increase in peer contributions on all articles by one page would yield the spillover of 2 bytes (about two characters) to individual contributions to Wikipedia.

Another measure of peer effects, spillovers from an increase in the number of peers contributing to the articles, yields a positive effect on individual contributions of 0.06 per cent. The IV estimates tend to be lower in magnitude than the OLS. Apparently, peer performance indeed affects individual performance and translates to larger total contributions to Wikipedia. Individuals who have active peers seem to redistribute their effort to other arti-

Table 4: Peer effects at the editor level

	Log length of contribution (in bytes)					
	OLS	FE	2SLS	OLS	FE	2SLS
Log peer contrib. all articles (bytes)	0.092***	0.082***	0.114**			
	(0.016)	(0.018)	(0.048)			
# peers on all pages				0.089***	0.084***	0.066**
				(0.007)	(0.008)	(0.027)
Interest in the category (%)	0.026***	0.046***	0.046***	0.027***	0.046***	0.046***
	(0.002)	(0.006)	(0.006)	(0.002)	(0.006)	(0.006)
Editor experience (months)	-0.010***	-0.126	-0.128	-0.010***	-0.171	-0.160
	(0.001)	(0.106)	(0.111)	(0.001)	(0.105)	(0.111)
Edits on popular pages (%)	0.012***	0.000	0.000	0.011***	-0.000	-0.000
	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Peer interest in the category (%)	0.005***	0.005**	0.006**	0.003	0.003	0.003
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
Peer experience (months)	-0.002	-0.001	-0.002	-0.001	-0.001	-0.001
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Peer edits on popular pages (%)	0.013***	0.008***	0.008***	0.009***	0.006***	0.006***
	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Observations	8492	8492	8476	8492	8492	8476
Kleibergen-Paap Wald F statistic			674.65			127.34

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

NOTES: The table shows the results of the reduced form regressions to estimate peer effects. Columns (1)-(3) show the results for the peer average contribution and Columns (4-6) for the number of peers. Specification (1) and (4) show OLS results; (2) and (5) show FE results. In Columns (3) and (6), I assume that peer effects are endogenous and estimate them in two steps. All regression coefficients are presented with heteroscedasticity robust standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The unit of observations is a monthly contribution to Wikipedia by an editor. All month and year dummies are included.

cles that need further improvement rather than keeping to improve the quality of the articles to which they contributed before.

Other factors that matter for the length of contributions or the number of revisions are the preferences and interests of individuals. The results reveal the importance of an individual interest in a specific topic. Firstly, the interest in a specific topic is positively associated with the size of contributions. A 1 per cent increase in the interest in a concrete topic measured by the share of contributions to the category, leads to a 0.05 per cent increase in the size of a contribution per article or overall within a given period of time. This means that an editor, whose interest in the topic is higher by one standard deviation (13 per cent) would contribute 0.65 per cent more to Wikipedia.

As the data allow me to distinguish between the bytes, added to and deleted from articles, I can further decompose the average peer activity at the editor level in order to examine whether adding or deleting of information has a higher impact on individual contributions, or is descried more by an individual. The results in table 5 suggest that both activities are recognized by individuals but peer adding of information has a higher importance for individuals.

6 Robustness checks

In order to examine to which extent my results hold in the alternative specifications, I perform several robustness checks of the main result for peer effects at the editor level. Firstly, I examine how the choice of measures affects the presence and the magnitude of observed peer effects. Secondly, I address one of most severe potential problems for the identification of peer effects in the network, which is selection of editors into the network due to assortative mixing, or their inherent similarities.

Table 5: Peer effects at the editor level with decomposed peer activity

	Log length of contribution (in bytes)			
	FE	2SLS	FE	2SLS
Log peer added bytes	0.083***	0.124**		
	(0.019)	(0.052)		
Log peer deleted bytes			0.049***	0.071**
			(0.010)	(0.030)
Interest in the category (%)	0.046***	0.046***	0.046***	0.046***
	(0.006)	(0.006)	(0.006)	(0.006)
Editor experience (months)	-0.127	-0.129	-0.132	-0.138
	(0.106)	(0.111)	(0.106)	(0.111)
Edits on popular pages (%)	0.000	0.000	0.000	0.000
	(0.002)	(0.002)	(0.002)	(0.002)
Peer interest in the category (%)	0.005**	0.007**	0.005**	0.006**
	(0.002)	(0.003)	(0.002)	(0.003)
Peer experience (months)	-0.001	-0.002	-0.001	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)
Peer edits on popular pages (%)	0.009***	0.009***	0.008***	0.008***
	(0.002)	(0.002)	(0.002)	(0.002)
Observations	8492	8476	8492	8476
Kleibergen-Paap Wald F statistic		636.94		516.65

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

NOTES: The table shows the results of the reduced form regressions to estimate peer effects. Columns (1)-(2) show the results for the peer average added bytes of contribution and Columns (3-4) for the deleted bytes. Specification (1) and (3) show FE results. In Columns (2) and (4), I assume that peer contributions are endogenous and estimate them in two steps. All regression coefficients are presented with heteroscedasticity robust standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The unit of observations is a monthly contribution to Wikipedia by an editor. All month and year dummies are included.

6.1 Measurement

The results obtained in the previous section could be questioned from the perspective of their reliance on the chosen measures of individual and peer activity. To address this concern, I make a robustness check, in which I apply the baseline model, however, changing the measures of individual and peer activity on Wikipedia. Instead of total monthly amount of bytes contributed (added or deleted), I use the number of revisions both for individual and peer monthly contributions in the logarithmic form. I replicate the baseline estimation for the editor level in table 6. The first stage is shown in the Appendix (table 11).

Similarly to the baseline results, the estimates for the number of revisions demonstrate that a one per cent increase in peer activity would account for 0.1 per cent increase in an individual activity and one additional peer would be related to 0.06 per cent more revisions. In absolute numbers, this would mean that 500 revisions made by median peers induce one additional revision by a median individual. While adding text by peers is visible directly to the individual (on the article), edits are not directly observed unless an individual specifically browses the edit history of the article. Moreover, many contributors save every word they edited and the platform Wikipedia tracks this as a separate edit. This might explain that far more revisions need to be done to contribute significantly to an article and, subsequently, to help an editor notice her peers' ongoing activity.

6.2 Network formation

The following analysis addresses a harsh problem of endogeneity in network formation that is inherent for many networks. In the case of the network of editors on Wikipedia, individuals observe the article text and have to perform additional clicks on edit history links to learn about the intensity of modification of this article and even more clicks to learn whether other individuals, who previously contributed to the article, are productive and appealing. The second way to learn about individuals is to contribute to the article and to wait for the feedback of editors particularly feeling engaged with this article. All this takes

Table 6: Robustness check: Peer effects at the editor level measured by the number of revisions

	Log length of contribution (the number of revisions)					
	OLS	FE	2SLS	OLS	FE	2SLS
Log peer # revisions of all articles	0.085***	0.069***	0.100***			
	(0.006)	(0.008)	(0.020)			
# peers on all pages				0.061***	0.057***	0.057***
				(0.004)	(0.004)	(0.009)
Interest in the category (%)	0.012***	0.031***	0.031***	0.013***	0.031***	0.031***
	(0.001)	(0.003)	(0.003)	(0.001)	(0.003)	(0.003)
Editor experience (months)	-0.004***	-0.036	-0.039	-0.004***	-0.063*	-0.055
	(0.001)	(0.038)	(0.041)	(0.001)	(0.037)	(0.034)
Edits on popular pages (%)	0.004***	-0.002**	-0.002**	0.003***	-0.002***	-0.002***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
Peer interest in the category (%)	0.003***	0.003***	0.004***	0.001	0.001**	0.001**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Peer experience (months)	0.000	0.000	0.000	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Peer edits on popular pages (%)	0.006***	0.003***	0.003***	0.003***	0.001	0.001
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Observations	8492	8492	8476	8492	8492	8476
Kleibergen-Paap Wald F statistic			858.82			139.49

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

NOTES: The table shows the results of the reduced form regressions to estimate peer effects. Columns (1)-(3) show the results for the peer average contribution and Columns (4-6) for the number of peers. Specification (1) and (4) show OLS results; (2) and (5) show FE results. In Columns (3) and (6), I assume that peer effects are endogenous and estimate them in two steps. All regression coefficients are presented with heteroscedasticity robust standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The unit of observations is a monthly contribution to Wikipedia by an editor. All month and year dummies are included.

some time. Assuming that the newcomers have not yet learned about interrelations and edit intensity of other already existing editors, I use only the newcomers to examine whether they experience peer effects.

I choose the subsample of editors during their first month after joining Wikipedia. As they joined Wikipedia just recently, they had little time to learn about potentially existing stable productive clusters of peers who already recognize each other (in case there are any). I perform the same specification as in equation 2 using the two-stage least squares approach with instrumental variables but now in the cross-section framework.

In the results (see table 7), I consider the activity of exclusively inexperienced editors of Wikipedia, i.e. those who contributed to Wikipedia during a month for the first time. As the data sample now has a cross-section shape I only obtain OLS and IV results. The IV results show larger magnitude. However, they should be treated with caution as F-statistic is relatively low for the number of peers (see column 4 in table 7). According to the endogeneity tests, the measures of peer activity and amount should again be treated as endogenous.[14] While the OLS results demonstrate that the issue of selection into the network of editors does not bias our main findings upwards, IV specifications demonstrate that if endogeneity exists the extent of selection into the network biases results downwards. This can be the case if some unobserved group shocks act in the opposite direction to the endogenous effects, which yields lower estimates if groups shocks (or correlated effects) are not ruled out by the IV estimation. Overall, any given pair of peers would share only a subset of all shocks to articles so that it is difficult to unambiguously predict whether the OLS estimator should be larger than the IV.

The results of this robustness check suggest that the effects from the main results are robust to self-selection into network. For the newcomer, the peer effects are still present and amount to 0.27 per cent for the average amount of peers' contributions (slightly higher than in the baseline model) and 0.36 per cent for the number of peers (also higher than in the

[14]The corresponding χ^2 statistics ranges from 3.924 to 7.102.

Table 7: Robustness check: Peer effects at the editor level only for inexperienced editors

	Log length of contribution (in bytes)			
	OLS	2SLS	OLS	2SLS
Log peer contrib. per article (bytes)	0.109***	0.273***		
	(0.041)	(0.085)		
# peers on all pages			0.072***	0.356***
			(0.022)	(0.137)
Interest in the category (%)	0.021***	0.021***	0.021***	0.025***
	(0.003)	(0.003)	(0.003)	(0.003)
Editor experience (months)	0.155	0.141	0.156	0.102
	(0.134)	(0.136)	(0.133)	(0.146)
Edits on popular pages (%)	0.001	-0.001	0.001	-0.004
	(0.003)	(0.003)	(0.003)	(0.003)
Peer interest in the category (%)	0.010**	0.016***	0.006	0.007
	(0.005)	(0.005)	(0.004)	(0.005)
Peer experience (months)	0.003	0.001	0.005	0.009
	(0.005)	(0.005)	(0.005)	(0.006)
Peer edits on popular pages (%)	0.009**	0.008**	0.006	-0.006
	(0.004)	(0.004)	(0.004)	(0.007)
Observations	900	900	900	900
Kleibergen-Paap Wald F statistic		147.55		16.37

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

NOTES: The table shows the results of the reduced form regressions to estimate peer effects. Columns (1)-(2) show the results for the peer average contribution and Columns (3-4) for the number of peers. Specification (1) and (3) show OLS results; In Columns (2) and (4), I estimate peer effects in two steps using instrumental variables. All regression coefficients are presented with heteroscedasticity robust standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The unit of observations is a monthly contribution to Wikipedia by an editor. All month and year dummies are included.

baseline model). However, the magnitudes of the main results still provide a quantitatively trustworthy indication of the potential of peer effects in Wikipedia.

7 Concluding remarks

The existence and the size of potential peer effects in online communities has been examined by few studies in the context of social networks and open-source software projects. Wikipedia is an online platform for peer knowledge generation that shares some similarities with the other kinds of platforms but also has very distinct features. This study is (to the best of my knowledge) the first to analyze the existence of peer effects in content generation due to contributor interactions on Wikipedia. Moreover, my study addresses the importance of coordination on article talk pages for creating social ties between contributors and, as a consequence, the emergence of multiplicative effects in online content generation.

The results show that, while controlling for observable editor and peer characteristics, an increase in the monthly average peer contribution by 1 per cent increases the amount of individual monthly contributions to Wikipedia (among individuals who contribute to Wikipedia every month) by about 0.1 per cent. Similarly, spillovers coming from the number of peers yield a positive effect of 0.06 per cent per overall monthly contributions to Wikipedia. These effects are observed among active individuals that contribute to Wikipedia and engage into discussions on the article talk pages. Lastly, the other characteristic that matters for the amount of individual contributions is an individual's interest, or an expertise, in a special category of articles.

My findings show that, even in the absence of explicit social ties between individuals, peer effects are present. However, their magnitude remains modest. The existence of positive peer effects suggests that communication between active community members could be beneficial as it enhances privately contributed knowledge. To flourish, communications should supported by technological mechanisms built in the platforms.

References

Aaltonen, Aleksi and Stephan Seiler, "Quantifying Spillovers in Open Source Content Production: Evidence from Wikipedia," Technical Report, Centre for Economic Performance, LSE 2014.

Algan, Yann, Yochai Benkler, Mayo Fuster Morell, and Jérôme Hergueux, "Cooperation in a Peer Production Economy Experimental Evidence from Wikipedia," in "Workshop on Information Systems and Economics, Milan, Italy" 2013, pp. 1–31.

Andreoni, James, "Privately provided public goods in a large economy: the limits of altruism," *Journal of Public Economics*, 1988, *35* (1), 57–73.

_ , "Giving gifts to groups: How altruism depends on the number of recipients," *Journal of Public Economics*, 2007, *91* (9), 1731–1749.

Aral, Sinan, Lev Muchnik, and Arun Sundararajan, "Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks," *Proceedings of the National Academy of Sciences*, 2009, *106* (51), 21544–21549.

Aschhoff, Birgit and Christoph Grimpe, "Contemporaneous peer effects, career age and the industry involvement of academics in biotechnology," *Research Policy*, 2014, *43* (2), 367–381.

Bapna, Ravi and Akhmed Umyarov, "Do Your Online Friends Make You Pay? A Randomized Field Experiment in an Online Music Social Network," 2012.

Benkler, Yochai, "Coase's Penguin, or, Linux and" The Nature of the Firm"," *Yale Law Journal*, 2002, pp. 369–446.

Bercovitz, Janet and Maryann Feldman, "Academic entrepreneurs: Organizational change at the individual level," *Organization Science*, 2008, *19* (1), 69–89.

Bilodeau, Marc and Al Slivinski, "Toilet cleaning and department chairing: Volunteering a public service," *Journal of Public Economics*, 1996, *59* (2), 299–308.

Bloch, Francis and Matthew O Jackson, "Definitions of equilibrium in network formation games," *International Journal of Game Theory*, 2006, *34* (3), 305–318.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin, "Identification of peer effects through social networks," *Journal of Econometrics*, 2009, *150* (1), 41–55.

Burke, Moira, Robert Kraut, and Elisabeth Joyce, "Membership claims and requests: Conversation-level newcomer socialization strategies in online groups," *Small Group Research*, 2010, *41* (1), 4–40.

Contreras, Salvador, Frank Badua, and Mitchell Adrian, "Peer Effects on Undergraduate Business Student Performance," *International Review of Economic Education*, 2012, *11* (1), 57–66.

Faulkner, Ryan, Steven Walling, and Maryana Pinchuk, "Etiquette in Wikipedia: Weening New Editors into Productive Ones," in "in" WikiSym 2012.

Fowler, James H and Nicholas A Christakis, "Estimating peer effects on health in social networks: A response to Cohen-Cole and Fletcher; Trogdon, Nonnemaker, Pais," *Journal of Health Economics*, 2008, *27* (5), 1400.

Giorgi, Giacomo De, Michele Pellizzari, and Silvia Redaelli, "Identification of social interactions through partially overlapping peer groups," *American Economic Journal: Applied Economics*, 2010, pp. 241–275.

Gorbatai, Andreea, "Aligning collective production with demand: Evidence from Wikipedia," *Available at SSRN 1949327*, 2011.

Gorbatai, Andreea Daniela and M Piskorski, "Social Structure of Contributions to Wikipedia," Technical Report, Working Paper 2012.

Hahn, Jungpil, Jae Yun Moon, and Chen Zhang, "Emergence of new project teams from open source software developer networks: Impact of prior collaboration ties," *Information Systems Research*, 2008, *19* (3), 369–391.

Hanushek, Eric A, John F Kain, Jacob M Markman, and Steven G Rivkin, "Does peer ability affect student achievement?," *Journal of Applied Econometrics*, 2003, *18* (5), 527–544.

Harper, Yan Chen F Maxwell, Joseph Konstan, and Sherry Xin Li, "Social comparisons and contributions to online communities: A field experiment on movielens," *The American Economic Review*, 2010, pp. 1358–1398.

Jackson, Matthew O and Asher Wolinsky, "A strategic model of social and economic networks," *Journal of economic theory*, 1996, *71* (1), 44–74.

Kacperczyk, Aleksandra J, "Social influence and entrepreneurship: The effect of university peers on entrepreneurial entry," *Organization Science*, 2013, *24* (3), 664–683.

Keegan, Brian, Darren Gergle, and Noshir Contractor, "Staying in the loop: structure and dynamics of Wikipedia's breaking news collaborations," in "Proceedings of the Eighth Annual International Symposium on Wikis and Open Collaboration" ACM 2012, p. 1.

Kittur, Aniket and Robert E Kraut, "Beyond Wikipedia: coordination and conflict in online production groups," in "Proceedings of the 2010 ACM conference on Computer supported cooperative work" ACM 2010, pp. 215–224.

Kleibergen, Frank and Richard Paap, "Generalized reduced rank tests using the singular value decomposition," *Journal of Econometrics*, 2006, *133* (1), 97–126.

Kummer, Michael E, "Spillovers in networks of user generated content: Evidence from 23 natural experiments on Wikipedia," Technical Report, ZEW Discussion Papers 2013.

Lacetera, Nicola and Mario Macis, "Social image concerns and prosocial behavior: Field evidence from a nonlinear incentive scheme," *Journal of Economic Behavior & Organization*, 2010, *76* (2), 225–237.

Laniado, David and Riccardo Tasso, "Co-authorship 2.0: Patterns of collaboration in Wikipedia," in "Proceedings of the 22nd ACM conference on Hypertext and hypermedia" ACM 2011, pp. 201–210.

Lerner, Josh and Jean Tirole, "Some simple economics of open source," *The Journal of Industrial Economics*, 2002, *50* (2), 197–234.

Manski, Charles F, "Identification of endogenous social effects: The reflection problem," *The Review of Economic Studies*, 1993, *60* (3), 531–542.

Moon, Jae Yun and Lee S Sproull, "The role of feedback in managing the Internet-based volunteer work force," *Information Systems Research*, 2008, *19* (4), 494–515.

Nov, Oded, "What motivates wikipedians?," *Communications of the ACM*, 2007, *50* (11), 60–64.

Osterloh, Margit and Sandra Rota, "Open source software developmentJust another case of collective invention?," *Research Policy*, 2007, *36* (2), 157–171.

Panciera, Katherine, Aaron Halfaker, and Loren Terveen, "Wikipedians are born, not made: a study of power editors on Wikipedia," in "Proceedings of the ACM 2009 international conference on Supporting group work" ACM 2009, pp. 51–60.

Sassenberg, Kai, "Common bond and common identity groups on the Internet: Attachment and normative behavior in on-topic and off-topic chats.," *Group Dynamics: Theory, Research, and Practice*, 2002, *6* (1), 27.

Shriver, Scott K, Harikesh S Nair, and Reto Hofstetter, "Social Ties and User-Generated Content: Evidence from an Online Social Network," *Management Science*, 2013.

Zhang, Xiaoquan and Feng Zhu, "Group size and incentives to contribute: A natural experiment at Chinese Wikipedia," *American Economic Review*, 2010, pp. 07–22.

8 Appendix

Table 8: First stage equations for log-peer contributions (bytes) at the editor-article level

	(1)	(2)
Av. # indirect peers on all pages	0.027***	0.044***
	(0.004)	(0.012)
Interest in the category (%)	-0.004	-0.008
	(0.004)	(0.007)
Editor experience (months)	0.009	0.272**
	(0.057)	(0.126)
Edits on popular pages (%)	0.001	0.004*
	(0.001)	(0.002)
# editors per page	0.003***	0.155***
	(0.001)	(0.008)
Page size (bytes)	0.000	-0.000
	(0.000)	(0.000)
Page is popular	0.013	-0.338
	(0.104)	(0.207)
Peer interest in the category (%)	-0.004**	-0.009***
	(0.002)	(0.002)
Peer experience (months)	-0.001	-0.002
	(0.001)	(0.002)
Peer edits on popular pages (%)	0.015***	0.006***
	(0.001)	(0.002)
Observations	4398	4398

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: First stage equations for log-peer contributions (bytes) at the editor level

	(1)	(2)
Av. # indirect peers on all pages	0.239***	0.440***
	(0.010)	(0.028)
Av. # indirect peers on all pages (sq.)	-0.006***	-0.012***
	(0.001)	(0.001)
Interest in the category (%)	0.002	0.001
	(0.004)	(0.009)
Editor experience (months)	-0.008	0.468***
	(0.069)	(0.175)
Edits on popular pages (%)	0.002	0.008**
	(0.001)	(0.004)
Peer interest in the category (%)	-0.037***	-0.013***
	(0.001)	(0.003)
Peer experience (months)	0.008***	0.004
	(0.001)	(0.003)
Peer edits on popular pages (%)	-0.005***	0.022***
	(0.001)	(0.002)
Observations	8476	8476

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: First stage equations for log-peer contributions (bytes) at the editor level

	(1)	(2)
Av. # indirect peers on all pages	0.222***	0.390***
	(0.010)	(0.017)
Av. # indirect peers on all pages (sq.)	-0.005***	-0.010***
	(0.001)	(0.001)
Interest in the category (%)	0.003	0.003
	(0.004)	(0.008)
Editor experience (months)	0.005	0.125
	(0.067)	(0.128)
Edits on popular pages (%)	0.001	0.003
	(0.001)	(0.002)
Peer interest in the category (%)	-0.037***	-0.060***
	(0.001)	(0.003)
Peer experience (months)	0.008***	0.013***
	(0.001)	(0.002)
Peer edits on popular pages (%)	-0.007***	-0.003
	(0.001)	(0.003)
Observations	8476	8476

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: First stage equations for log-peer contributions (the number of revisions) at the editor level

	(1)	(2)
Av. # indirect peers on all pages	0.224***	0.450***
	(0.009)	(0.027)
Av. # indirect peers on all pages (sq.)	-0.005***	-0.011***
	(0.001)	(0.001)
Interest in the category (%)	0.000	-0.000
	(0.003)	(0.009)
Editor experience (months)	0.034	0.733***
	(0.057)	(0.157)
Edits on popular pages (%)	0.001	0.010***
	(0.001)	(0.004)
Peer interest in the category (%)	-0.034***	-0.015***
	(0.001)	(0.003)
Peer experience (months)	0.008***	0.005
	(0.001)	(0.003)
Peer edits on popular pages (%)	-0.013***	0.024***
	(0.001)	(0.002)
Observations	8476	8476

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Robustness check: First stage equations for log-peer contributions (bytes) at the editor level only for inexperienced editors

	(1)	(2)
Av. # indirect peers on all pages	0.288***	0.136***
	(0.023)	(0.034)
Av. # indirect peers on all pages (sq.)	-0.008***	
	(0.001)	
Interest in the category (%)	-0.002	-0.014***
	(0.002)	(0.003)
Editor experience (months)	0.083	0.139
	(0.098)	(0.199)
Edits on popular pages (%)	0.006***	0.014***
	(0.002)	(0.003)
Peer interest in the category (%)	-0.039***	-0.005
	(0.003)	(0.004)
Peer experience (months)	0.012***	-0.012**
	(0.003)	(0.005)
Peer edits on popular pages (%)	-0.008**	0.034***
	(0.004)	(0.005)
Observations	900	900

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$