SOCIAL STRUCTURE OF CONTRIBUTIONS TO WIKIPEDIA

Andreea Daniela Gorbatai  
Harvard University  
Wyss Hall 203D  
Harvard Business School  
Boston, MA 02163  
agorbatai@hbs.edu

Mikołaj Jan Piskorski  
Harvard University  
Morgan Hall 243  
Harvard Business School  
Boston, MA 02163  
Tel. (617) 495-6099  
mpiskorski@hbs.edu

* John Sheridan helped us assemble the dataset. Division of Research at Harvard Business School provided financial support. All errors are ours.
SOCIAL STRUCTURE OF CONTRIBUTIONS TO WIKIPEDIA

ABSTRACT

On-line platforms for socially generated content, such as Wikipedia, allow everyone to contribute and remove the contributions of others. This presents a trade-off: on the one hand, content improves quickly and mistakes are eliminated; on the other hand, deletion of content deprives its authors of the intrinsic pleasure of contributing and seeing their work in print, which may lead them to stop participating. We suggest that this trade-off is attenuated by the social structure around contributors. We hypothesize that actors involved in high density and high encapsulation structures will be least likely to abandon contributing. Analysis of the pattern of contributions to Wikipedia supports this assertion. The results underscore the importance of tight social structure for the success of collaborative content projects.

Keywords: social networks, density, encapsulation, collaborative content creation, benefits, secondary rewards, Wikipedia.
Introduction

Over the past decade we witnessed a proliferation of online collaborative projects. Starting with software design projects, online collaborations expanded into online content creation. Among these projects, the online encyclopedia Wikipedia stands out for its ambitious scope. Created in January 2001, Wikipedia is based on an intuitive editing and publishing platform, which makes it easy for the casual reader to contribute to articles. Its growth is indeed astounding – by the end of 2007 Wikipedia had over 2.15 million articles in English and 9.2 million articles in 253 languages\(^1\). By all reckonings, the site surpassed any known printed compendiums of human knowledge both in number of total characters and range of topics. The project is also vastly collaborative – globally there are 6.1 million registered contributors with over 800,000 participating to the English edition. The usage numbers are equally impressive. It is currently used by 8% of the Internet population, attracting .6% of all page views, which makes it the 9\(^{th}\) most visited website in the world\(^2\).

The success of Wikipedia is by no means obvious. In fact, there are at least four different factors that lead us to expect that the project would not be so successful. First, Wikipedia editors put a tremendous amount of effort into contributing to the various articles, but obtain no direct or indirect monetary benefits. Second, Wikipedia expects that editors contribute jointly to the article without a clear attribution of authorship of its different fragments. This suggests that they are either motivated by the intrinsic benefits of sharing the knowledge they deeply care about, or by the social benefits accruing from participation in such a project.

Third, Wikipedia essentially provides a pure public good – which implies that an individual's consumption leads to no subtractions from anyone else’s consumption of that good and that no one

---


can be excluded from accessing it (Olson 1971). These attributes lead to free-riding concerns whereby potential contributors wait for others to undertake the work and then benefit from it without incurring the costs of contribution. Finally, there is the problem of being edited out. In Wikipedia, one’s work is literally fused into the project and can be edited out in seconds, with a minimal level of effort, by any other user, registered or not, through mechanisms beyond the control of the original contributor. In other words, contributors voluntarily relinquish rights to their own work product to the community, exposing themselves not only to the risk of forsaking expected benefits from the product, but also having their own contribution obliterated.

This paper identifies the attributes of egocentric social structure which play an important role in continued contributions. We hypothesize that recurrent contributions to editing Wikipedia articles are due to rewards resulting from participation in a structured, collaborative endeavor. In brief, we state that working on a set of topics together with a group of connected and committed others creates the necessary socio-structural conditions for the production of the public good (Coleman 1987; 1990). We test these ideas using data from Wikipedia.

The rest of the paper is organized as follows. First, we discuss the notions of network density and network encapsulation. Then, we apply them to examine contribution to public goods, such as Wikipedia entries, and briefly summarize Wikipedia structure and work processes our model relies upon. We subsequently test the relationship between the two network measures and likelihood of contribution.

Network Attributes: Density and Encapsulation

Though there are many factors that may diminish individual motivation to contribute, we argue that the individual location in the social structure has such a positive impact on recurrent editing that it can compensate for these shortcomings of the collaboration model. We employ two
independent egocentric network measures, density and encapsulation, representative of the social structure of collaborative editing.

The first measure, egocentric density, is one of the most widely used and simple network indices. It takes values between 0 (for a graph with no ties between any two friends of the ego) to 1 (for a complete graph, where all possible ties between friends are realized); its value indicates the fraction of possible ties present in the egocentric network (Wasserman and Faust 1994:164; 181-2). Density as group level measure has been used as a proxy for group cohesion (Blau 1977) and social capital (Coleman 1990), as a key factor in information and influence flows, especially in the presence of structural holes (Burt 1992), and as predictor of performance in creative settings (Uzzi 1997). Its influence depends on the outcome of interest (Lin 2001:10): denser networks are better for resource preservation (Lin 1990; 1992), whereas sparse and more extended networks are more beneficial for resource search and procurement (Burt 1992; Lin 1999).

The second measure we examine, encapsulation, has been less extensively used. Following Uehara (1990:529), we define encapsulation as “the extent to which members of an action set are more ‘intensely’ linked to one another than to others outside the set.” This measure can be seen as a “connectivity” pattern (Moody 2005) or as an egocentric cohesion index.

**Insert Figure 1 about here**

The two measures – density and encapsulation – capture orthogonal characteristics of egocentric networks. This is illustrated in Figure 1. Consider the egocentric network around C first. This network is dense because C has four ties, and each of them is connected to two of the others. However, it is not encapsulated, because the ties also have outside connections in addition to in-group ties. If C’s ties were his friends, we could say his friends are connected to each other (dense network) but they also have other friends that C is not friends with (low encapsulation). Now consider the egocentric network around B: it is relatively sparse but encapsulated, because
individuals in her network do not have outside connections. In short, B’s friends aren’t friends of each others’ but neither do they have friends outside this circle. Cases A and D are examples of low encapsulation and density, and respectively high density and encapsulation. Imagining ties as friendship relations, A has friends who are not friends with each other, and have many outside friends, whereas D’s friends form a tight group because they are friends with each other and do not have other friends outside this group.

Wikipedia data does not contain information of friendship ties between editors; instead, the information that we have comes from affiliation matrices between editors and articles they contribute to. Wikipedia is a world-wide online project; editors are unlikely to communicate with each other offline, and have few reasons to engage in communications related to knowledge contributions beyond their sphere of interest. Since their interests are reflected in the edits that they make, the ties among editors can be defined through common participation on editing a topic. However, many articles have a large number of contributors who participated at different points in time, sometimes as remotely as five years apart: such extreme cases do not warrant to be labeled as “ties”. We therefore limit our definition to a maximum of one week lag between edits: given observed Wikipedia behavior such interval represents a plausible time length for interested contributors to be aware of each other’s work and express interest in a contribution.

Density, Encapsulation and Provision of Public Goods

The choice of density and encapsulation as measures of social networks attributes anticipates our framing the analysis of Wikipedia editing as public good provision. Provision of public goods is notably problematic because those who provide such goods are often unable to appropriate the full value of their contributions; this failure to fully benefit leads to under-provision and free riding. Two of the solutions proposed for alleviating free riding concerns in public good provision focus on means of internalizing externalities through an organization empowered to allocate the production
costs among those experiencing the benefits (Olson 1965), and, respectively, on contributions based on a “thoughtful conviction” in the value of the good for its own sake regardless of production costs (O’Mahony 2006:245).

However, Coleman (1987; 1990) suggested that the free riding could also be diminished through effective group norms, where the existence of a norm is defined as the state “in which a socially defined right to control an actor’s action is held not by the actor but by others” (Coleman 1990:266). He further distinguished between the demand and the realization of effective social norms. The interest and demand for a norm arise as consequence of an action that has externalities for a set of others, but no markets for control exist, and no single actor can engage in an exchange for rights of control (Coleman 1990:251). This demand for effective norms, coming from the need to encourage (or limit) the externality-producing action of another (Coleman 1990:249), will be met (realized) in situations “under which beneficiaries of a norm, acting rationally, either will be able to share appropriately the costs of sanctioning the target actors or will be able to generate” (Coleman 1990:273) sufficiently large sanctions to induce some beneficiaries to act.

Coleman argued that free riding is diminished and individuals help provide public goods under socio-structural conditions that facilitate demand and realization of effective norms of social behavior (Coleman 1990). More specifically, the social structure of a group plays an important role in reducing two types of free riding: first order free riding (manifested as suboptimal contribution to production of a good) and second order free riding - the reliance on other individuals to apply sanctions while benefiting from their enforcement (Coleman 1990:270).

In Coleman’s solution, secondary rewards, defined as others’ action of rewarding participants for contributing to a good which is commonly valued, are the mechanism through which social structure affects subsequent behavior. Positive and negative sanctions strongly linked to working toward the common goal therefore produce secondary, additional benefits to individuals as a reward
for helping others satisfy their interests. This way, “the existence of a[n efficient] norm with sanctions does what formal organization does in the presence of externalities: it internalizes them” (Coleman 1990:277).

These ideas can be easily applied to Wikipedia: given the nature of editing on the encyclopedia, each editor has an incentive to reward the others for working toward the common outcome of improving the quality of a certain knowledge topic. Under the framework proposed above, such secondary rewards are likely to occur on Wikipedia because they are easy to bestow (a word of praise on the editor's user page, or a “Thank You!” note) while conceivably being valued as signs of social support and acting as incentives for the editor to further contribute.

Two features of social structures are particularly important for granting and amplifying sanctions meant to reduce free riding occurrence: the existence of social relations between the actor and the beneficiaries of externalities from his actions, and the existence of social relations among those benefiting from actions. Such conditions ensure the existence of effective community norms.

In elaborating on this connection between social structure and the provision of public goods, we employ a social exchange lens by looking at individual contributions and secondary rewards as social exchanges giving rise to obligations among participants. Empirical work on group exchanges mirrors social networks theory predicting that group structure has implication for actor behavior and social exchanges (Mitchell 1969: 18, 27; Uehara 1990: 528-529).

We propose that Wikipedia editing can be interpreted as exchange embedded in the social network of co-editors. Wikipedia editors expose themselves to a double risk: not having their contributions appreciated and having their contributions permanently erased. Therefore, on topics where more than one individual participates, there are two main ways of manifesting support towards fellow co-editors: first, express gratitude for their contributions (that is, provision of secondary rewards - an action which alleviates free riding concerns), and second - defend their work
against random, out-group editors, which we define as all individuals who are not recognized as regular co-editors on the set of topics the in-group collaborates on.

An average Wikipedia user with a moderate interest in (reading) one or several articles may encounter an article or section she had additional information about or that she can easily improve. We note that an occasional editor’s interest in such pages is ordinarily extrinsic to Wikipedia (given by her prior knowledge and interests). However, upon contributing, such individuals may receive feedback on their contributions, usually from regular contributors who are either monitoring the page for changes or currently working on the article she participated on. Such feedback may come the form of a positive, laudatory note posted on the article page or personal user page if the editor took a few seconds to register, or as negative feedback posted on the article (most often expressed through altering or erasing her work).

It is in this context that we claim that the mechanisms responsible for recurring contributions to public goods become activated: if her participation occurs in a context of high density, where other editors are frequently collaborate on writing articles related to that particular topic, repeat edits earn her if not the praise at least the respect of other editors, and a potentially higher level of support. “Editing with friends” lowers the likelihood that her work will be edited out, thus increasing the total benefits she derives from seeing her work become part of the final product, and diminishing the risks she takes when contributing. The rewards derived from this relationship with co-editors will increase her overall benefits from the project, increasing her likelihood of repeated contribution. In a similar manner, collaborating in groups characterized by high network encapsulation makes it more likely that one’s presence and contributions are protected from random outsiders unfamiliar with the history of those particular topics the group is involved in, because co-editors are more focused on rewarding other group members and protecting their own, and less distracted by outside obligations or conflicts on interest.
Conversely, editing on topics characterized by sparse, erratic editing diminishes the rewards one receives from collaboration in at least two ways: first, editing alone, without sanctions from other editors, diminishes the likelihood that one’s work will survive contestation; second, lone editors do not receive the social rewards and positive sanctions which internalize the externalities produced by her contributions and increase the benefits she derives from her work. Similarly, participants in a low encapsulation group might receive rewards from co-editors, but it is possible that these editors are more focused on participation with other groups or on other topics, or might even use “outsiders” that the editor is not collaborating with as allies in arguments against individuals from the main group.

Given the nature of Wikipedia editing and interaction, which occurs mostly through serial exchanges on the wiki platform, the obligations arising from co-editing exchanges span over longer periods of time and types of support: one individual might need assistance to preserve a contested contribution in an article, whereas those helping her might benefit from being trusted with editing a certain section, or from receiving praise or other expressions of gratitude for their help. More importantly, obligations will be diffuse in the sense that members accept duties towards “others from whom one cannot directly expect the fulfillment of one’s ‘rights’” (Uehara 1990:526). The existence of in-group ties ensures group communication and translates in a network of exchanges; the more ties there are among group members the higher the potential for exchange, so group density is beneficial for Wikipedia contributions. Therefore egocentric density is one way to quantify the relationship between network structure and social exchange patterns (Uehara 1990:528), exchange which here represents repeated contribution to provision of a public good:

**Hypothesis 1:** An increase in egocentric density around an editor increases the likelihood of that editor’s contribution.
Nevertheless, high density is not a sufficient condition for the emergence and sustainability of generalized exchange and norm effectiveness. Encapsulation is also important: higher intensity of linkage with others results in higher social support and more effective norm enforcement thus inducing people to contribute despite the cost to edit. In other words, a registered editor collaborating to an article will be affected not only by the density of her egocentric network of collaborators, but also by the extent to which her network is encapsulated. As discussed above and as predicted by Coleman’s model, density clearly plays an important role because it affects communication flows between those benefiting directly from her contribution to the article(s). However, if, regardless of network density, individuals participating on the same article(s) have other ties that she is not connected with, they conceivably have additional obligations, and interests, and may receive or bestow other benefits unrelated to her goals. Hence encapsulation represents a good measure for the extent to which an editor’s goals overlap with those of other individuals in the network, which in turn affects the likelihood that she will receive secondary rewards.

As an interviewee suggested, “there are articles with editors who act as rather zealous caretakers … to protect their time investment [or an] article’s point-of-view slant (or…protecting against attempts at inserting slants…by outsiders)”. If peer editors are not strongly invested in articles the ego is participating on and have many outside ties (low encapsulation), they are more likely to forsake “caretaking” and bestowing secondary rewards on others; outsiders might intrude and damage the article. Such state of facts would be thrice disappointing to the ego: first, because her contributions are not acknowledged and rewarded, second, because her own contributions will not be protected from deletion, and lastly, because the benefits she expected to derive from improving that topic will not materialize.

Since all of the above disappointments derived from low encapsulation can act to lower incentives to contribute, we expect that:
Hypothesis 2: An increase in egocentric encapsulation around an editor increases the likelihood of that editor’s contribution.

At this point in our study, understanding the correspondence between the theoretical model of collaborative contributions and the empirical findings based on Wikipedia data requires a brief incursion into the structure and processes of Wikipedia.

Structure of Project and Article Space

Every Wikipedia article consists of a set of five interrelated pages: article (and edit article) page, discussion (and edit discussion) page, and a history page. The Article presents the contents of the most recent version of that respective entry. To make changes to this entry, users click on an edit this page button, which presents them with an editable version of the contents of the article page. This feature allows any reader willing to contribute to quickly edit the content. While more experienced editors might follow up with some clean-up work, edits incidental to the reading process represent important knowledge contributions which would not be possible were the interface more complex.

The media-wiki software provides a complete history of any given article: by reading the History tab attached to each article, a reader can find summary information about every edit to the article since its creation. Most importantly, any user can see difference between any two versions and undo any of them; this feature has been increasingly used by editors in anti-vandalism defense.

The article page is a collection of structured content, focused around text, about the topic stated in the title of the article. Articles contain text, media (pictures, graphics or sounds), references supporting the presented information or providing further detail, and a section linking the article to knowledge categories the article belongs to. Occasionally the title can be followed by a status bar.

---

3 This summary information contains the time and date of edit, user name/IP address, as well as a link to the former’s talk page, to a list of the registered user’s contributions, the size of the edit, and a brief note regarding the nature of the edit provided by the editor or generated automatically by a Wikipedia script.
indicating, for example, whether the article is incomplete, suspected of bias, featured article (of encyclopedic quality), or protected from editing as reaction to recent vandalism.

In addition to the main encyclopedic space, Wikipedia contains other spaces assigned to special pages. The Wikipedia space contains meta-pages of interest to the project, such as community rules, manuals of editing and writing style, pages presenting a collection of articles organized by Projects (e.g., History of France, Physics), historical pages relating to Wikipedia debates, and statistics about Wikipedia. This space is tightly linked to the Help, Special (pages), and Template spaces, which contain tools facilitating user involvement, administration and qualitative improvement of the project. User pages are personal spaces for registered editors’ self-expression, containing free-format information about the interests, (real-life) identity, or activities that the owner wants to share with the community. Lastly, Talk: pages are never independent of other pages (i.e., they are attached to rule pages, templates, user pages, or articles) and offer space for discussing and debating the content of the primary page, asking for information, or leaving messages for users of the primary pages.

We follow with second element necessary in connecting theory and empirical tests: the description of the normative framework and the work process for Wikipedia contributions.

Wikipedia Rules and Work Process

The first formally recorded Wikipedia rule, “Neutral Point of View” emerged during early 2001, and was later acknowledged as the foundational principle of the Wikipedia community. More policies and guidelines followed soon after; these policies are currently summarized as five central principles of Wikipedia. Central to any encyclopedia, the first two principles clarify the scope of content as a repository of unbiased, accurate and verifiable (hence, scientific) synthesis of expert

---

knowledge. The last three principles are uncharacteristic of traditional encyclopedias, and interest us because they set the code of conduct and social rules governing the behavior of contributors to the project. Thus, fostering an environment of open and enthusiastic contribution and benevolence are upheld against making ‘perfect contributions’. Additionally, one must recognize equality of (editing) rights and not expect contributions to survive the “boldness in editing” of fellow editors. We argue that it is this latter rule of open editing (which includes and even encourages editing-out other contributors’ work) that accounts for most of the observed everyday dynamics of editing Wikipedia, because it generates a need for social support in editing, support which is manifest not only as secondary reward for contribution, but also as protective action towards other co-editors contributions to a common goal.

The five main principles discussed above, together with the all-empowering *Ignore all rules* (if they hinder improvement) have arguably created a need for a large project infrastructure of volunteers and automatic software tools (robots, or, more informally, *bots*) dedicated to preserving and improving the quality of Wikipedia articles against the work of vandals and inexperienced or unskilled users. Bots were implemented as an addition to volunteer clean-up work performed by editors in order quickly perform many of the monotonous and repetitive tasks of maintenance and improvement of Wikipedia content. They are modules of software designed and managed by editors following a formal approval process. Examples of *bot* tasks include automatically signing unsigned edits, spell-checking text or capitalizing certain words, formatting dates or merely placing reminder tags for pending tasks (such as “cleanup required” or “reference needed”).

**Data and Measures**

This section addresses the data and the main assumptions underlying our models. The initial data has been obtained from the Wikimedia Foundation, parent of the Wikipedia project. The foundation provides end-users with direct, free-of-charge access to the entire content of its site,
among which complete, archived versions of the entire encyclopedia. We downloaded the complete article history of all English Wikipedia pages existing as of November 4, 2006 in the form of wiki-text source and metadata embedded in XML. The compressed UNIX file was 5GB large and contained a total of 54,020,662 edits. This dataset contains a complete history of Wikipedia edits between January 2001 and November 2006.

The data analysis proceeded as follows: data was initially imported into a MySQL database using a custom-designed parser to extract variables of interest for the analysis. Because we defined editor interaction as a dialogue reflected on the physical body of their work, and recorded as such in the history of the evolution of each article, we restricted the analysis to article text. Since the project focuses on substantive contributions to knowledge, a decision was made to exclude pages such as lists, links to pages consolidating content on a topic (redirects), and clarifying pages (disambiguations); these pages play a role in structuring the encyclopedia similar to a book’s table of contents or index, but do not contain actual knowledge in the sense that we are interested in. Finally, a slice of the initial database representing all Wikipedia edits between January 1st and June 30th, 2004 (1,679,399 edits) was extracted for analysis purposes.

The next step consisted in writing a Matlab algorithm to eliminate abusive edits (vandalism acts) and the edits restoring the vandalized article to the previous state (reverts). These actions are not of interest for us, because they do not reflect “normal” collaborating actions. Reverts are human editor or bot actions which do not bring substantive contributions to the article, but are part of maintenance activities. Vandalisms and reverts are substantively different from editing disagreements or conflicts; reverts are semi-automated actions that editors can perform by selecting the revert option on the article edit page. As per Wikipedia etiquette, reverting the work of another editor who

---

has contributed to an article in good faith is considered rude and offensive behavior; one is expected to respect others’ work, improve on it and communicate politely. Therefore, reverts are to a high degree reserved for intentionally harmful editing actions, such as erasing the article, or replacing article text with obscenities, so both they and their removal are irrelevant from the point of view of contributions to knowledge.

Lastly, we dropped the vandalism and revert edits (118,193 edits), as well as anonymous edits (IP edits) and automatic (bot) edits (which eliminated 340,300 and, respectively, 7,758 of the remaining edits), and proceeded to compute the network measures and control variables needed. This final algorithm ran on 1,213,148 Wikipedia edits of 14,221 registered users on 239,161 articles, and generated a discrete event history dataset based on six hour intervals (728 total intervals).

Social Network Measures

The two social network measures of interest, density and encapsulation, were constructed using the discrete-time editing records. In order to make best use of the longitudinal structure of the data, network measures were computed applying a sliding window filter of one-week length to the event history of article edits: a 28-interval long window moves along the time axis in one-interval increments. Thus, the network at time t consists of all ties among network members in the interval \([t-28; t-1]\). This technique has been previously used in social network analysis and it is of high relevance in analyzing effects of network dynamics and of past social structure on current behavior (Moody, McFarland, and Bender-deMoll 2005).

The egocentric density measure for each time interval was computed by taking the ratio of realized ties to possible ties in one’s network at time t. This measure reflects the extent to which an editor’s ties are connected to each other by participating to the same articles (though not necessarily always on the same articles which ties them to the editor). Conversely, the social network encapsulation measure reflects the extent to which an editor’s friends are friends only with the
editor’s friends. It is computed as an average of the ratio of friends that another editor in the egocentric network has in common with the editor, to the ratio of her total friends. As shown Figure 1 for B and D, the higher the encapsulation, the more disconnected the group formed around the ego is from the rest of the network. In the last step before data analysis, we eliminate records for which the encapsulation measure is undefined.

**Additional Covariates**

We also computed several control variables relevant to modeling the relationship between past social network characteristics and current behavior. These variables include a measure of past edits (previous editing intervals) and number of time intervals since the last edit. Time effects have been taken into account using three daily interval controls (with first interval as baseline), day of the week controls, and a linear time trend.

**Modeling Approach and Results**

We examine the likelihood of a registered editor’s contributing knowledge to Wikipedia during a given interval using logistic modeling with random and fixed effects. Let $e_{it}$ be a dummy variable indicating if editor $i$ performed at least one edit during interval $t$. We estimate the following model:

$$
\log\left(\frac{p_{it}}{1-p_{it}}\right) = \beta_1*density_i + \beta_2*encapsulation_i + B* covariates + \Gamma* timecontrols + \alpha_i + \epsilon_{it},
$$

where $p$ is the probability that an editor contributes to Wikipedia during time $t$ ($e_{it}=1$), and $\alpha_i$ is the estimated unobserved time-invariant individual characteristics in the fixed effects model, and a random, subject-specific intercept accounting for unobserved individual heterogeneity in the random effects estimates.

**Insert Table 1 about here**
The two main specifications chosen compute the effect of density and encapsulation on the likelihood of editing using fixed and random effects models and controlling for number of intervals since the previous edit, time effects (time of the day and day of the week when the editor is at risk of editing), and a time trend (in models 1 and 5), or, respectively, a control for the number of previous intervals when the editor contributed (in models 2 and 6). Since the number of previous contribution intervals is monotonously increasing, it is highly correlated with the time variable, so the two have been included in separate specifications due to colinearity concerns. The results indicate consistent support for the two hypotheses, with a higher positive coefficient on encapsulation (Hypothesis 2) than the coefficient on density (Hypothesis 1). Additional models 3 and 4 tested separately for the effect of each lagged social network attribute on the likelihood of contributing and produced similar coefficients in specifications, confirming that sign and magnitude of the two measures (Table 1). More sensitivity tests, including variations in the specification of time intervals, and controls for egocentric network size provide support for the two hypotheses; these results are available upon request.

**Conclusion**

This study provides evidence for the fact that social networks matter for public good provision. Participating with familiar others, in a context where collaborators know each other and are committed to provision of a public good increases the likelihood of contributing.

Further research should explore the rich archival data available on wiki platforms and analyze community emergence, individual and cohort trends in participation, and determinants of more complex editing patterns beyond the act of editing (e.g., edit sizes, choice of topics). Analysis of other-language Wikipedias would allow separating temporal (project age) and cultural effects (language-specific) from those determined by the structure of interaction on online collaborative platforms.
References.


