

# Motivational determinants of participation trajectories in Wikipedia

Martina Balestra<sup>1</sup>, Ofer Arazy<sup>2</sup>, Coye Cheshire<sup>3</sup>, Oded Nov<sup>1</sup>

<sup>1</sup>New York University; <sup>2</sup>University of Haifa; <sup>3</sup>University of California, Berkeley  
{mb5758, onov}@nyu.edu; ofer.arazy@gmail.com; coye@berkeley.edu

## Abstract

Many digitally mediated peer-production systems allow participants to define their own activities. The challenge in such systems, however, lies in retaining members beyond the first few interactions. To address this problem we must understand who these users are and why they begin to contribute. Importantly, there is scant empirical evidence on how motivations are associated with different trajectories of participation for new participants. Our study addresses this gap by combining a survey of new Wikipedia editors' motivations with an exploratory analysis of the editors' activity logs. Using clustering techniques to identify prototypical activity profiles from log data, we observe what motivations are associated with which prototypical activities. We find that new editors' motivations are predictive of their future activity. In particular our results indicate that reputation, social, enjoyment, and obligation motives differ among editor activity clusters.

## Introduction

Technology-mediated social participation (TMSP) systems like Wikipedia and open-source software projects are designed to harness the knowledge of distributed participants in the community and are therefore largely open to anyone who wants to participate (Kogut and Metiu 2001). Traditional incentive and compensation structures are frequently missing because the public-good quality of many of these systems favors the voluntary contributions and engagement of participants (Kogut and Metiu 2001, Oreg and Nov 2008). Participants have the ability to act with a high degree of autonomy: contributors to peer-production systems are frequently able to choose the project or task that they want to work on, and an approach to the task in line with their interests (Hippel and Krogh 2003). At the same time, these systems create organizational structural constraints and affordances by design (what individuals can do, where or how it can be done, and with whom), and therefore channel the potential for users' roles, responsibilities, and actions in ways that are not usually possible in the offline world. Wikipedia provides an especially useful arena for examining how and why individuals choose to behave within these affordances and constraints because its design and operation are observable and persistent.

In Wikipedia, "namespaces" constrain editors to a finite range of content production and interaction types. There currently exist 35 namespaces ("Wikipedia:Namespace" 2015) comprising designated spaces where the content of articles is produced, where editors can coordinate, interact and communicate, and where editors can accomplish administrative tasks. Within this framework, contributors tend to engage in consistent behavior patterns that constitute emergent, informal roles on the platform (Callero 1994). The extent to which users' emergent roles are attributable to the individuals' personal characteristics and motivations, however, is often unclear. We seek to understand this relationship better by focusing on the impact of motivation on behavior, particularly for newcomers. Retaining participants beyond the first several visits to the platform is a major challenge (Pei-Yu and Hitt 2002). Thus, a novel aspect of our work lies in characterizing *new* participants in peer-production platforms: we identify the emergent roles of new Wikipedia editors and study the relationship between these roles and editors' initial motivations for contributing to Wikipedia.

Prior research investigates personal characteristics that have an impact on individuals' behavior in online communities; however, these studies are focused on existing, veteran contributors. This has two main limitations; first, it undervalues early experiences. Initial interactions are formative (Trotter and Roberts 2006), and individuals' perceptions and motivations are liable to develop and change in this period. The second limitation is a methodological concern: those who respond to the survey are already self-selected - they keep contributing and may therefore not be representative of the much larger number of contributors who contribute and never come back.

To address these concerns, and to identify the relationship between motivations and consecutive activity trajectories, participants in our study were surveyed at the earliest point of engagement to identify what motivations are more likely to lead the participants toward certain activity patterns. Understanding the motivations and behaviors of new users can have important implications for informing the development of appropriate incentives and interfaces to motivate participants to return to the platform. We therefore investigate the following research questions: What types of prototypical activities do new Wikipedia editors tend to engage in during their first several months of activity? What is the relationship between new Wikipedia edi-

tors' motivations to become Wikipedia editors, and the activities they tend to engage in?

## Methods

### Sample

The majority of Wikipedia editors are not very active. To account for this, we examined the activity logs of thousands of new editors in their first two weeks of participation to determine the distribution of editors across the 50<sup>th</sup>, 75<sup>th</sup>, 90<sup>th</sup>, 95<sup>th</sup> and 99<sup>th</sup> percentiles of edit behavior: 0-1 edits, 2-4 edits, 5-8 edits, 9-14 edits, and 15+ edits, respectively. We then used a stratified sampling technique to recruit approximately equal numbers of editors from each of these five strata to ensure that there was sufficient data from each. In the analysis, the data were then adjusted using weights (the quotient of the population proportion divided by the sample proportion) to account for this sampling technique.

The first two weeks of editing behavior were observed for all newly created accounts during one of two month-long recruitment intervals: May-June 2014, and September-October 2014. Two intervals were used to account for periodicity effects. 100 potential participants with active email addresses were randomly identified from each sampling strata, and emails containing links to the study were sent to these accounts after their first two weeks. Participants who reported having other Wikipedia accounts were removed from the sample, leaving 154 viable participants (146.34 participants after weighting.)

### Data

*Motivations* - To measure individuals' motivations for editing Wikipedia, an online questionnaire was developed and distributed to all potential participants through the recruitment email. Participants received a \$10 online gift card for completing the survey. Motivation was measured using constructs drawn from two primary sources and measured using Likert scales: reputation, enjoyment, and obligation motives were derived from (Lakhani and Wolf 2003), and identification, social, and norm-oriented motives were derived from (Schroer and Hertel 2009).

*Activity Profiles* - We interpreted editors' activity profiles as the distribution of edits across the Wikipedia namespaces because they embody its designed structure. Each namespace is distinguished based on the type of content or interaction found on a page. Survey respondents' log data from all edits made across the six main namespaces: Main Article, Talk, User Page, User Talk, Wikipedia, and Wikipedia Talk, following from (Antin, Cheshire, and Nov 2012), were collected using the Wikipedia API. For descriptions of the namespaces see ("Wikipedia:Namespace" 2015).

Edit data from each namespace were collected for the first 209 days of activity on Wikipedia (i.e. the shortest interval of time between the creation of an account and the

date of data collection) for all editors in the sample. We used both count and proportion data in our analyses to create a comprehensive picture of activity, and because they can offer complementary views: analysis of the number of edits tells us how much editing work is done in each namespace and privileges the dichotomy of high- and low-producing editors, whereas analysis of the proportion of edits made in each namespace helps us to infer the relationship between activity types editors engage in and gives us a more detailed view of how individuals who make an intermediate number of edits behave.

### Data Analysis

#### Identifying Prototypical Activity Profiles

In order to identify emergent roles we used a clustering algorithm to group editors' activity in each namespace, interpreting each cluster's centroid as a prototypical activity profile. For this exploratory analysis we used the  $k$ -means clustering technique that partitions patterned behavior into  $k$  clusters using a Euclidean distance measure where each observation belongs to the cluster with the nearest centroid (Jain, Murty, and Flynn 1999).  $k$  therefore reflects the number of prototypical activity profiles discerned from the log data, characterized by the amount of activity in each namespace. Count and proportion data were clustered independently, and count data were log transformed before analysis to account for a substantial skew (Levine and Dunlap 1982). We iteratively tested  $k$ -means for  $k$  clusters, where  $k \in [3, 5]$ . A smaller number of cluster may be too simplistic, whereas more than 5 clusters would be difficult to interpret theoretically with such a small sample.

Within this range, we examined the ratio of the minimum cluster separation to the maximum cluster diameter (i.e. the Dunn Index) in addition to the correlation between clusters (i.e. the Normalized Gamma correlation) for each configuration with  $k$  clusters (Halkidi, Batistakis, and Vazirgiannis 2001). We also limited the possible cluster configurations to those with a minimum of 5% of the sample in any cluster because clusters with fewer members are not meaningful. Within these parameters, we found  $k=3$  to be the optimal cluster configuration for the  $k$ -means analysis using both the count of edits and the proportion of edits in each namespace.

#### Motivations and Their Relationship to Activity Profiles

Participants' membership in the activity profiles were observed and used in the analysis of motivation. Each member's results were weighted according to the sampling strata to reflect the true distribution of membership across Wikipedia editors. We used analysis of variance (ANOVA) and Tukey HSD post-hoc tests to compare measure means across individual clusters of behavior.

## Results

Participants made between 1-201 edits in the first 209 days of editing Wikipedia (M=19.12 edits, SD=29.67) a substantial majority of which were in the main article namespace.

### Activity Profiles Based On Count Data

We gave each cluster a descriptive title based on their primary contributions: Low-Volume Main Article Editors (LVMA), High-Volume Main Article Editors (HVMA), and Comprehensive Editors (CE) (Figure 1). LVMA Editors make a small number of edits in the main article space, and an even smaller number of edits in the user space. HVMA Editors make a high number of edits in the main article space, and a low number of edits in other namespaces. CE Editors make a high number of edits in both the main article and user spaces, and a fairly high number of edits in the remaining namespaces. LVMA Editors constitute 35.71% (55/154 participants) of the sample (75.13% of the weighted population). 56.49% (87/154 participants) are HVMA Editors (23.09% of the weighted population). Finally, only 7.8% (12/154 participants) of the unweighted sample are CE Editors (1.78% of the weighted population).

### Activity Profiles Based On Proportion Data

Prototypical activity profiles based on the proportion of edits in each namespace are given the following descriptive names: Main Article Editors (MA), Generalist Editors (GE), and User Page Editors (UP) (Figure 2). MA Editors make almost all their edits in the main namespace. UP editors make the vast majority of their edits on the user page. GE Editors made relatively equal contribution to all six namespaces. 110/154 participants (71.4% of the weighted population) in this study were MA Editors (84.6%). 21/154 participants (13.6%) are GE Editors (7.2%), and 23/154 participants (14.93%) are UP Editors (8.2% of the weighted population).

### Motivation

Table 1 contains the results of the ANOVA comparing motivation measures across activity profiles derived from weighted count data, and Table 2 shows these results for the activity profiles based on weighted proportion data.

When activity profiles are based on edit count, reputation and social motives differ significantly across activity profile ( $F(2,142)=8.76$ ,  $p=0.0003$ , and  $F(2,139)=6.00$ ,  $p=0.018$ , respectively). Specifically, HVMA Editors are more motivated by reputation and social motives than the LVMA Editors (significant at the  $p<0.00$  level and  $p<0.05$ , respectively).

When clusters are based on the proportion of edits made in each namespace, both enjoyment and obligation motives differ significantly across activity profile ( $F(2,141)=5.05$ ,  $p=0.0024$ , and  $F(2,141)=26.09$ ,  $p=0.000$ , respectively). GE editors are significantly more motivated by enjoyment than either MA Editors ( $p<0.01$ ) or UP Editors ( $p<0.01$ ). UP

Editors were significantly less motivated by obligation than either MA Editors ( $p<0.00$ ) or GE ( $p<0.05$ ).

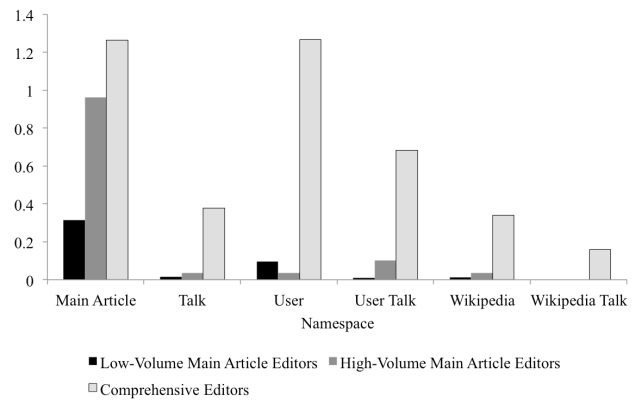


Figure 1. Clusters based on the log-transformed count of edits.

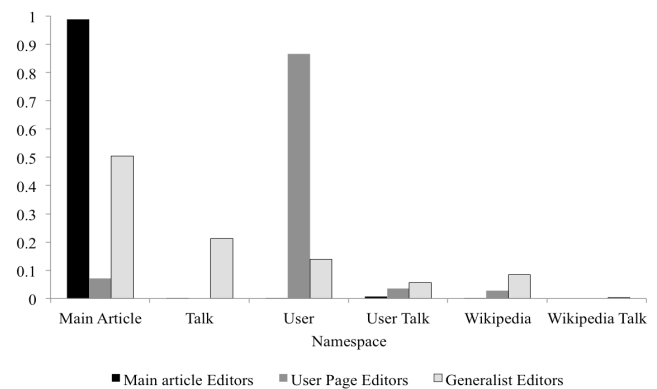


Figure 2. Clusters based on the proportion of edits.

	LVMA M (SD)	HVMA M (SD)	CE M (SD)	p
Reputation	2.32 (1.25)	3.41 (1.38)	2.23 (1.86)	0.0003
Identification	2.85 (0.77)	2.62 (1.08)	2.40 (0.91)	n.s.
Enjoyment	3.78 (0.87)	3.78 (1.143)	3.89 (1.10)	n.s.
Social	1.83 (1.13)	2.53 (1.44)	1.78 (1.03)	0.018
Obligation	4.04 (0.93)	3.99 (0.98)	3.75 (1.38)	n.s.
Norm Oriented	3.81 (0.82)	4.03 (0.61)	3.61 (0.74)	n.s.

Table 1: Results based on the number of edits (weighted).

	MA M (SD)	GE M (SD)	UP M (SD)	p
Reputation	2.60 (1.40)	2.60 (0.93)	2.22 (1.35)	n.s.
Identification	2.75 (0.86)	2.86 (0.96)	3.14 (0.57)	n.s.
Enjoyment	3.71 (0.94)	4.69 (0.58)	3.63 (0.74)	0.002
Social	2.00 (1.26)	2.59 (0.90)	1.46 (1.03)	n.s.
Obligation	4.32 (0.77)	3.73 (0.79)	2.73 (1.57)	0.000
Norm oriented	3.79 (0.79)	4.28 (0.84)	4.05 (0.45)	n.s.

Table 2: Results based on the proportion of edits (weighted).

## Discussion

Peer production systems often don't retain participants beyond their first few interactions. To address this problem, we need to understand the participation trajectories of newcomers, as well as the factors that drive them to participate. To fill this gap we combined a survey of new Wikipedia editors' motivations with an analysis of their activity logs, and examined their relationships using clustering techniques to identify prototypical activity profiles.

Our analysis highlights a distinction between High-Volume Main Article Editors and Comprehensive Editors, who are characterized by behaviors that we tend to associate with mature editors, and not newcomers. Legitimate Peripheral Participation (Lave and Wenger 1991) suggests that newcomers increase their participation through growing and diversifying their contributions over time. We find, however, that there also exists this class of editors that is drawn immediately to core participation. Others have observed this distinction in high-volume editors as well (Welser et al. 2011, Arazy et al 2015), yet none have looked at what drives newcomers to these different paths to participation.

We find that High-Volume Main Article Editors value both reputation and the social aspects of editing more than Low-Volume Main Article Editors. An explanation may be that individuals tend to contribute more when it can enhance their reputation, and that social individuals contribute more than less-social individuals in team settings (Wasko and Faraj 2005). Prior studies help to explain the contrast between High- and Low-Volume Main Article Editors, but they do not explain the behavior of Comprehensive Editors. Comprehensive and Generalist Editors may instead have prior experience in other TMSP systems such that their contributions are richer and more diverse when they become Wikipedia editors (Taylor and Todd 1995), or different socialization experiences early on (Choi et al. 2010).

Generalist Editors were more motivated by enjoyment than either Main Article Editors or User Page Editors. Other studies support these results (Lin 2007), particularly when knowledge sharing is viewed more comprehensively than simply producing content. Tacit knowledge sharing is more important to individuals who are more intrinsically motivated than those who are extrinsically motivated (Osterloh and Frey 2000).

We also find that User Page Editors rate obligation motives significantly lower than both Main Article and Generalist Editors. This can be interpreted in light of Ardichvili, Page, and Wentling's (2003) work: Main Article and Generalist Editors may be more likely to view their knowledge as a public good, and are therefore less motivated by self-interest. By this reasoning, User Page Editors are *more* motivated by self-interest, and therefore less motivated by a sense of obligation to the community at large.

## Acknowledgement

This work was partially supported by NSF grant ACI-1322218.

## References

- Antin, J., Cheshire, C., and Nov, O. 2012. *Technology Mediated Contributions: Editing Behaviors among New Wikipedians*. Proc CSCW.
- Arazy, O., Ortega, F., Nov, O., Yeo, L., and Balila, A. 2015. *Functional roles and career paths in Wikipedia*. Proc CSCW.
- Callero, P. 1994. *From Role-Playing to Role-Using: Understanding Role as Resource*. Social Psychology Quarterly 57 (3): 228–43.
- Choi, B., Alexander, K., Kraut, R., and Levine, J. 2010. *Socialization Tactics in Wikipedia and Their Effects*. Proc. CSCW.
- Halkidi, M., Batistakis, Y., and Vazirgiannis, M. 2001. *On Clustering Validation Techniques*. Journal of Intelligent Information Systems 17 (2-3): 107–45.
- von Hippel, E., and von Krogh, G. 2003. *Open Source Software and the 'Private-Collective' Innovation Model*. Organization Science 14 (2): 209–23.
- Jain, A., Murty, N., and Flynn, P. 1999. *Data Clustering: A Review*. ACM Computing Surveys 31 (3): 264–323.
- Kogut, B., and Metiu, A. 2001. *Open-Source Software Development and Distributed Innovation*. Oxford Review of Economic Policy 17 (2): 248–64.
- Lakhani, K., and Wolf, R. 2003. *Why Hackers Do What They Do*. SSRN Scholarly Paper ID 443040.
- Lave, J. and Wenger, E. 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press.
- Levine, D. and Dunlap, W. 1982. *Power of the F Test with Skewed Data*. Psychological Bulletin 92 (1): 272–80.
- Lin, H. 2007. *Effects of Extrinsic and Intrinsic Motivation on Employee Knowledge Sharing Intentions*. Journal of Information Science, 33, 135–149.
- Oreg, S., and Nov, O. 2008. *Exploring motivations for contributing to open source initiatives: The roles of contribution context and personal values*. Computers in human behavior, 24(5), 2055-2073.
- Osterloh, M. and Frey, B. 2000. *Motivation, Knowledge Transfer, and Organizational Forms*. Organization Science 11(5): 538–50.
- Panciera, K., Halfaker, A., and Terveen, L. 2009. *Wikipedians Are Born, Not Made*. In Proceedings of the ACM, 51–60.
- Pei-Yu, C. and Hitt, L. 2002. *Measuring Switching Costs and the Determinants of Customers Retention in Internet-Enabled Businesses*. Information Systems Research 13 (3): 255–74.
- Schroer, J. and Hertel, G. 2009. *Voluntary Engagement in an Open Web-Based Encyclopedia*. Media Psychology 12 (1): 96–120.
- Taylor, S. and Todd, P. 1995. *Assessing IT Usage: The Role of Prior Experience*. MIS Quarterly 19 (4): 561–70.
- Trotter, E. and Roberts, C. 2006. *Enhancing the Early Student Experience*. Higher Education Research & Development 25 (4): 371–86.
- Wasko, M. and Faraj, S. 2005. *Why Should I Share? Examining Social Capital and Knowledge Contribution in Electronic Networks of Practice*. MIS Quarterly 29 (1): 35–57.
- Welser, H., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gay, G., and Smith, M. 2011. *Finding Social Roles in Wikipedia*. Proceedings of the 2011 iConference.
- “Wikipedia:Namespace.” 2015. *Wikipedia, the Free Encyclopedia*. <https://en.wikipedia.org/w/index.php?title=Wikipedia:Namespace&oldid=692038707>.