

Turbulent Stability of Emergent Roles: The Dualistic Nature of Self-Organizing Knowledge Co-Production

Ofer Arazy*, Johannes Daxenberger†, Hila Lifshitz-Assaf‡, Oded Nov§ and Iryna Gurevych†

*Department of Information Systems, Faculty of Social Sciences, University of Haifa, Israel

Email: oarazy@is.haifa.ac.il

†Ubiquitous Knowledge Processing Lab, Dep. of Computer Science, Technische Universität Darmstadt, Germany

URL: www.ukp.tu-darmstadt.de

‡Stern School of Business, New York University, USA

Email: h@nyu.edu

§Dep. of Technology Management & Innovation, Tandon School of Engineering, New York University, USA

Email: onov@nyu.edu

Abstract—Increasingly, new forms of organizing for knowledge production are built around self-organizing co-production community models with ambiguous role definitions. Current theories struggle to explain how high-quality knowledge is developed in these settings and how participants self-organize in the absence of role definitions, traditional organizational controls, or formal coordination mechanisms. In this article, we engage the puzzle by investigating the temporal dynamics underlying emergent roles on individual and organizational levels. Comprised of a multi-level large-scale empirical study of Wikipedia stretching over a decade, our study investigates emergent roles in terms of prototypical activity patterns that organically emerge from individuals’ knowledge production actions. Employing a stratified sample of a thousand Wikipedia articles, we tracked two hundred thousand distinct participants and seven hundred thousand co-production activities, and recorded each activity’s type. We found that participants’ role taking behavior is turbulent across roles, with substantial flow in and out of co-production work. Our findings at the organizational level, however, show that work is organized around a highly stable set of emergent roles, despite the absence of traditional stabilizing mechanisms such as pre-defined work procedures or role expectations. This dualism in emergent work is conceptualized as “Turbulent Stability”. We attribute the stabilizing factor to the artifact-centric production process and present evidence to illustrate the mutual adjustment of role taking according to the artifact’s needs and stage. We discuss the importance of the affordances of Wikipedia in enabling such tacit coordination. This study advances our theoretical understanding of the nature of emergent roles and self-organizing knowledge co-production. We discuss the implications for custodians of online communities, as well as for managers of firms engaging in self-organized knowledge collaboration.

I. INTRODUCTION

Recent years have seen the rise of new forms of organizing for knowledge production, with a predominant one being open online co-production communities such as Wikipedia and open-source software (Benkler 2006, Von Krogh and Von Hippel 2006). The distinct principles of these new forms are leading to a re-investigation of traditional assumptions in organizational theory and a development of new theoretical

understandings and constructs (Lakhani et al. 2013, Schreyögg and Sydow 2010, Zammuto et al. 2007, Lifshitz-Assaf 2016). One of the key guiding principles of these open co-production knowledge communities is self-organizing, where participants themselves select how, and when to work, and what to work on (Lakhani and Panetta 2007, Benkler 2006, Von Krogh and Von Hippel 2006, O’Mahony and Lakhani 2011, Oreg and Nov 2008). This self-organizing is incommensurate with the logic of traditional organizations, based on a Chandlerian logic (Chandler 1962) that emphasizes authority, centralized hierarchy, and control. This has led to an inquiry of how emergent (or “informal”) knowledge work is developed in such new forms that, without utilizing clear role definitions, traditional organizational control, or coordination mechanisms, nevertheless result in a cumulative and high quality knowledge-based product (Faraj et al. 2011, Kane et al. 2014, Okhuysen and Bechky 2009, Arazy et al. 2011, Ransbotham and Kane 2011, Arazy and Nov 2010).

In order to build a comprehensive conceptualization of the work process, such an inquiry requires a focus on the emergent roles that individuals enact based on the work itself (Orlikowski 2000) and the tasks that are performed as they emerge, in order to build a comprehensive conceptualization of the work process (Bechky 2006). This perspective on roles, that focuses on individuals’ role behavior in relation to their work, is similar to the interactionist view of roles (Goffman 1961, Turner 1986). This view is in contrast to the traditional structural perspective of roles that understands roles as based on social expectation, norms, and status positions (Katz and Kahn 1978). Recently, there has been a call to focus on the nature of these emergent roles and how knowledge co-production work organically develops over time (Faraj et al. 2011, Kane et al. 2014, Majchrzak et al. 2013). The study of emergent roles in traditional organizations is well established, and scholars often refer to these as part of the “informal” aspects of work in organizations (Okhuysen and Bechky 2009). However, new forms of organizing for knowledge

production, such as open online co-production communities, introduce a novel area for theoretical exploration, as they rely heavily and almost exclusively on these emergent roles (Zammuto et al. 2007). Thus, the emergent becomes the center of knowledge production processes, rather than a component that complements formal roles.

The temporal view of the relationship between emergent roles and work has long been argued to be an important and missing perspective (Orlikowski and Yates 2002, Langley et al. 2013, Hernes 2014). In self-organizing knowledge co-production, this perspective is particularly relevant, as the high level of fluidity in participation results in multiple tensions in the creation of a cumulative body of knowledge (Faraj et al. 2011, Kane et al. 2014). Therefore, in this study, we unpack the black box of the emergent roles in knowledge co-production work over time. We suggest that addressing this research objective warrants a multi-level perspective combining the individual and the organizational levels. Theoretical advances in this area have been made in recent years, calling for a more nuanced conceptualization of emergent roles and work at both the organizational (Schreyögg and Sydow 2010, Langley et al. 2013, Farjoun 2010) and individual level (Faraj et al. 2011, Preece and Shneiderman 2009). However, empirical studies have yet to follow through. This investigation is particularly warranted given conflicting views in the literature on the extent to which and how individuals change their emergent roles and behavior in online co-productions communities over time (Panciera et al. 2009, Kane et al. 2014).

To address these gaps, we investigated new forms of organizing to learn about knowledge production with minimal formal role definition and structuring of the knowledge production activities. Moreover, we sought an organization that had developed a large number of sustained co-production efforts over extended periods and multiple knowledge-based products, thereby allowing us to study participants' role-taking behavior across such products. We therefore selected Wikipedia as the setting for our investigation. Wikipedia is one of the most notable examples of peer-production (Benkler 2006). Roles in Wikipedia are largely informal and emergent, and are organized around practices (Faraj et al. 2011, Gleave et al. 2009). Thus, the richness of its participant behavior data, both in depth and breadth, makes it particularly suitable for conducting a multi-level investigation of emergent roles in online co-production communities.

Our empirical investigation focuses on 1,000 representative Wikipedia articles from various topical domains and of varying maturity levels (in terms of the number of revisions they have gone through), and analyzes the editing activities in the co-authoring process of these articles. Our data collection and analysis procedure combined manual annotation processes (over 30,000 editing activities), machine learning algorithms (scaling up and automating the annotation process to 700,000 activities), and complex software scripts (to track the behavior of over 200,000 distinct participants over a period of eleven years). We recorded contributors' detailed profiles of wiki edit work and used statistical methods to identify behavioral regularities, or more specifically, prototypical activity patterns, as a proxy for emergent roles (Liu and Ram 2011, Welser et al.

2011). Seeking to understand the temporal dynamics by which the organizational and individual levels interact, we compared role dynamics between the "forming" period (from 2001 to the end of 2006) and the "establishing" period (from January 2007 to 2012) in Wikipedia's evolutions (Halfaker et al. 2012).

We found that while there is turbulent and intense mobility at the individual level, where many participants often take and shed roles instantaneously, the global structure of work is highly stable over epochs in Wikipedia's life, despite fundamental changes in its governance mechanisms. We conceptualize this dualistic interplay between individual-level mobility and organizational-level stability as "Turbulent Stability". A qualitative analysis of contributors' comments when editing articles, complemented with a quantitative analysis of role distribution across stages of articles' development, suggest that in enacting a particular role, contributors respond to the immediate needs of the co-produced artifact. We suggest the artifact-centric production as the critical enabling mechanism for the stability in emergent role behaviors across epochs in Wikipedia's evolution. Finally, we uncover the nature of emergent roles within Wikipedia, revealing some conceptualized, yet not empirically recorded roles.

II. THEORETICAL PERSPECTIVES

In this section we provide the theoretical perspectives for this work, by reviewing relevant streams in the literature. We first review prior works on emergent roles in knowledge co-production communities; next, we turn our attention to the literature on the dualistic nature of self-organizing knowledge co-production; and finally we discuss the role of the artifact in facilitating co-production.

A. Emergent Roles in Knowledge Co-Production

In recent years, new forms of organizing for knowledge production have emerged, where one prominent form are open online co-production communities such as Wikipedia and open-source software development (Benkler 2006, Von Krogh and Von Hippel 2006). Benkler (2006) defines this commons-based peer-production as a "system of production, distribution, and consumption of information goods characterized by decentralized individual action carried out through widely distributed, nonmarket means that do not depend on market strategies, *autonomous, self-selected, decentralized action*" (Benkler 2006, emphasis added). The distinct principles of these new forms call for a re-investigation of traditional assumptions in organizational theory and a development of new theoretical models applicable to this setting (Baldwin and von Hippel 2011, Lakhani et al. 2013, Schreyögg and Sydow 2010, Zammuto et al. 2007, Lifshitz-Assaf 2016). Despite the absence of clear role definitions, or traditional organizational control and coordination mechanisms, the community-based model has shown to be very effective, yielding high quality knowledge products (Faraj et al. 2011, Kane et al. 2014, Okhuysen and Bechky 2009, Arazy et al. 2011).

To pursue this inquiry, we focused on the emergent roles of knowledge co-production and their temporal dynamics. A theoretical focus on roles, as Turner (1986, p. 360) suggests,

provides an “understanding of why different patterns of social organizations emerge, persist, change, and break down”. Emergent roles organically materialize as work activities are enacted and are characterized by the tasks performed. The investigation of emergent roles can shed light on roles in action (Orlikowski 2000) and increase our understanding of labor division in the work process (Bechky 2006). Cohen (2013) stresses the importance of studying how tasks are assembled, bundled, and amalgamated into a job or a role. This perspective resonates with the interactionist view of roles (Goffman 1961, Turner 1986) and stands in contrast to the traditional structural perspective of roles, which views them as based on social expectations, norms, and status positions (Katz and Kahn 1978). It is also important to note that the sociological and organizational literature have been investigating the existence of emergent roles, usually as associated with the “informal” aspects of work in organizations, vis-a-vis the formal aspects (see the review of Okhuysen and Bechky (2009)). However in self-organizing knowledge co-production communities, using the “informal vs. formal” dichotomy is less applicable, since the production aspects of these online communities are largely “informal” and rely heavily on emergent work.

So far, the majority of empirical studies of roles within online communities have paid particular attention to the more “formal” aspects of roles similar to those in traditional organizations. Prior studies in this area have investigated leadership roles (Butler et al. 2008), organizational roles that enable power, authority and status (Arazy et al. 2014, Forte et al. 2009, Stvilia et al. 2008), and promotion processes from one formal role to another (Burke and Kraut 2008, Arazy et al. 2015). However, recent conceptualizations of self-organized knowledge co-production call to shift the focus to emergent roles and to the ways in which they are enacted in the moment, on a transient basis (Faraj et al. 2011, Kane et al. 2014, Majchrzak et al. 2013). Faraj et al. (2011) theorize that in these generative organizations – characterized by fluid participants, boundaries, and norms, loose governance, and absence of deep social relationships – roles rapidly emerge and change. They describe knowledge collaboration as “the enactment of temporary sets of behaviors that are volitionally engaged in, self-defined, and inductively created for the purposes of the online community” (Faraj et al. 2011, p. 1231). Few empirical studies have followed this perspective and tried to characterize the emergent roles that are created in response to tensions (Faraj et al. 2011), such as those between knowledge, change, and retention (Kane et al. 2014). We build on these recent conceptualizations to define emergent roles based on the knowledge co-production work itself and the sets of activities being enacted, and operationalize emergent roles as prototypical activity patterns (Welser et al. 2011, Liu and Ram 2011).

B. The Dualistic Nature of Self-Organizing Knowledge Co-Production

In order to shed light on the broader puzzle of how emergent knowledge work is developed without clear role definitions, we need a comprehensive understanding of the

nature of these roles and their development over time. The temporal view of emergent work has long been argued to be an important and missing one (Orlikowski and Yates 2002, Langley et al. 2013, Hernes 2014). In particular, there is a need to investigate the interplay between change and stability in organizations (Langley et al. 2013, Farjoun 2010, Faraj et al. 2011). In the context of our investigation of roles, it is not clear how individuals’ mobility in-and-out of community co-production work affects the characteristics of emergent roles (i.e. the extent to which emergent role behaviors persist over time). Traditional structural role theory (Katz and Kahn 1978) attributes the stability of role behaviors in traditional organizations (despite turnover in role occupants) primarily to norms and expectations held by role partners. Online co-production communities, however, differ. Traditional mechanisms for sustaining stable roles behaviors are absent in online co-production communities and the question of emergent role stability in such new forms of organizing remains open.

A review of the literature on knowledge co-production reveals conflicting views regarding the question of mobility vs. stability of emergent roles. On one hand, Faraj et al. (2011, p. 1231) suggest that in these fluid online communities “role-making contributions do not appear to be part of a repeated pattern, but rather a reaction by a single participant to a perceived state of the community”. Based on this perspective, we may postulate that the high levels of change and fluidity in individuals’ role-taking-and-shedding will translate into unstable role behaviors. That is, over time it will yield a high level of mobility on the individual level of taking and shedding roles, yet a low level of stability on the organizational level of these roles. Moreover, organizational theory perspective will also strengthen the prediction of a low level of stability over time since online co-production communities’ organizations change dramatically as they evolve (Halfaker et al. 2012).

Yet an alternative view suggests that participants do not change their role behavior significantly over time. For instance, Panciera et al. (2009, p. 59) argue that “Wikipedians are consistent. Wikipedians tend to maintain a high and constant level of participation for the majority of their lifespan”. Other views suggest that changes in participants’ behavior follow a particular trajectory, such as increasing the breadth and depth of their activities (Preece and Shneiderman 2009). We may therefore predict that individuals will keep to the same sets of tasks and enact the same emergent roles over time. Such a systematic activity pattern is likely to result in stable (emergent) role definitions, wherein the nature of emergent roles would remain constant over extended periods.

Many organizational theorists have discussed changes that “sustain and, at the same time, potentially corrode stability” in organizations (Tsoukas and Chia 2002, p. 568). Yet empirically, such changes have been challenging to validate and conceptualize. Online co-production communities offer such an opportunity. These new forms enable an investigation that can both “zoom in and zoom out” (Gaskin et al. 2014) of the actual work in order to find patterns of both change and stability in the unfolding self-organizing co-production. Despite these opportunities, research to date has tended to focus on only a single level: either exploring individuals’ role

dynamics (Preece and Shneiderman 2009) or characterizing the roles that emerge through collective action (Liu and Ram 2011). We therefore take a multi-level perspective, aiming to capture the tension between change and stability at both levels. As Aaltonen and Kallinikos (2013) stress, for co-production communities such as Wikipedia, in order to understand collective action, we need “knowledge making and learning that transcends methodological individualism”.

C. Artifact Centric Co-Production

Traditional role theories suggest that stability and change in roles are based on either role definitions and expectations (the structural perspective, see Katz and Kahn (1978)) or social interactions and negotiation (the interactionalist perspective, see Goffman (1963) and Turner (1986)). However, in online co-production communities, scholars argue that roles emerge based on tensions that arise in the co-production of artifacts and require balancing (Faraj et al. 2011). This perspective highlights the role of the co-produced artifact as a central mechanism enabling tacit coordination in peer-production (Howison and Crowston 2014). Our study advances this perspective. A focus on the artifact as a key factor facilitating emergent work highlights the role of materiality in organizational change (Leonardi and Barley 2008) and is aligned with the call for combining the social and the material dimensions in studying organizations (Orlikowski and Scott 2008).

The organizational and sociological literatures have referred to transparent and accessible artifacts that serve as a common substrate of knowledge as “boundary infrastructure” (Bowker and Star 1999), arguing that this infrastructure facilitates shared work. Boundary infrastructures enable knowledge production between professionals with different epistemic cultures (Cetina 1999) and in multiple organizations and professional communities (Bechky 2006, Carlile 2002). For instance, Tuertscher et al. (2014, p. 1588) investigated knowledge work around the development of a complex technological system (ATLAS) and have suggested that “Undergirding ... was the boundary infrastructure comprising objects and representations such as simulations that enabled common ground among the geographically distributed participants hailing from different epistemic communities”. Kellogg et al. (2006) have also conceptualized the coordination in temporary organizations using the notion of boundary objects that serve as a “trading zone” (Galison 1997) facilitating the dynamic, and ongoing work accommodation among online advertising professionals, and between them and their clients.

For co-production communities, the boundary infrastructure is much more than a bridge connecting disparate individuals. Rather, it is a ‘sine pro quo’; the existence of an online community is based on and is shaped by the artifact and its affordances (Faraj and Azad 2012). Okhuysen and Bechky (2009) highlight the roles of objects and representations in creating a common understanding of the work process and in facilitating coordination. Other scholars refer to this coordination as “stigmergic”, borrowing notions from natural collective intelligence systems: “In these virtual settings traditional coordination mechanisms (hierarchical direction, mutual

adjustment in face to face meetings, etc.) face limitations and the artifact takes on a more important role” (Bolici et al. 2016). They propose that stigmergic coordination plays a central role in co-production communities of open source software development, where “actors are leaving traces of their actions in the code and they are reading and reflecting on the code written by others in order to take coordinated action”. Our study strengthens this line of research.

III. RESEARCH METHODOLOGY AND FINDINGS

In the sections above, we reviewed the literature on emergent roles and have highlighted some of the gaps in this literature, namely in terms of: (a) the nature of emergent roles and the extent to which they are stable over time; (b) role-making dynamics; and (c) the way in which multiple levels (individual/organizational) and dynamic patterns (mobility/stability) interact. Our objective in this study is to fill these gaps in the literature and advance our understanding of emergent work in online co-production communities. In what follows we discuss the method employed for addressing this research objective.

The setting for our study is the online encyclopedia co-production community Wikipedia. Wikipedia has been able to recruit thousands of volunteers to produce millions of encyclopedic entries in 287 languages, and develop extensive policies and mechanisms for governing its collaborative authoring process. Wikipedia operates many different projects, defined as the co-production of a particular knowledge-based product (i.e. authoring and editing of a particular encyclopedic article on a wiki page), where the project group is comprised of the set of volunteers that has contributed to this wiki article. Wikipedia’s success has attracted the attention of both organizational and information systems scholars (Arazy et al. 2011, Forte et al. 2009, Ransbotham and Kane 2011).

The availability of temporal data harvested from peer-production system logs could be employed in computational social science – the quantitative modeling of technology-mediated social participation systems – similar to how the capacity to collect and analyze massive amounts of data transformed the fields of biology and physics. Technology-mediated interactions in socio-technical systems, such as online peer-production communities, capture the sequential contributions to a common artifact. Thus, analyzing these temporal sequences can reveal key insights regarding groups’ collaboration patterns in their natural setting, and allow us to identify emergent roles. In addition, we have employed a qualitative manual annotation procedure to interpret system log data. We have found that a multi-method approach is advantageous for studying emergent roles in online communities (Gleave et al. 2009, Welser et al. 2011).

We employed a sample of Wikipedia knowledge-based products (i.e. articles), tracking all editors and edit activities in each article in the sample from the article’s creation until our cut-off date (January 4th, 2012). After categorizing each activity, we created an activity profile for each contributor, and then clustered contributors to identify prototypical activity profiles. Our goal is to investigate co-production of knowledge

artifacts and thus the focal object of our analysis is a Wikipedia article. Given the dependencies between activities (i.e. each contribution is a response to earlier contributions (Kane et al. 2014)), it is essential that the analysis of activities tracks complete co-production sequences around a particular article, rather than tracking a set of participants and their contributions across many articles. Our primary strategy for capturing the complex role dynamics is temporal bracketing: recording a series of “snapshots” of the process over time (Langley et al. 2013). We apply this temporal bracketing strategy in our various analyses: (a) comparing two periods in Wikipedia’s life (in analyzing the stability of emergent roles and individuals’ role-taking dynamics); and (b) comparing four stages of articles’ evolution (based on the number of revisions).

A. Sample

We employed a double-stratified sampling procedure, randomly selecting 1,000 articles from the January 2012 dump of the English Wikipedia. Our strata were based on: (a) the maturity of articles (in terms of the number of revisions), and (b) the articles’ topical domains. This is important given that collaboration patterns could differ across articles in different stages of their life cycle (Hallerstede 2013) and across topical domains (Arazy et al. 2011, Kittur et al. 2009). This sampling approach is in line with prior studies of Wikipedia (Arazy et al. 2011). Given the power law distributions in the number of articles’ revisions (Ortega et al. 2008), we used the following four maturity strata: (a) 1-10 revisions; (b) 11-100; (c) 101-1,000; and (d) more than 1,000 revisions, and refer to these stages as: *inception*, *creation*, *growth*, and *maturity*, respectively (Hallerstede 2013). The topical strata were based on Wikipedia’s categorization system, using the main topics scheme.¹ The 25 topical categories are: Agriculture; Arts; Business; Chronology; Concepts; Culture; Education; Environment; Geography; Health; History; Humanities; Humans; Language; Law; Life; Mathematics; Medicine; Nature; People; Politics; Science; Society; Sports; and Technology. With four maturity strata and 25 topical categories, we have 100 cells with ten randomly selected articles in each (i.e. 250 articles in each maturity stratum and 40 articles in each topical category). Altogether, our sample contained 721,806 activities (i.e. article revisions), authored by 222,119 contributors.

B. Categorizing Activities

In order to create activity profiles of contributors, we needed to first determine the categories of edit activities. The categorization of activities was based on a two-step approach: first, we manually annotated a data sample; second, by employing the manual annotation as a training set, we applied a machine learning algorithm to categorize all 721,806 revisions in our sample set of 1,000 articles. In contrast to prior studies that have focused on active contributors (Liu and Ram 2011), we included all contributors, even those with very few editing activities assuming that such activities were intentional (rather

than random). This inclusive approach enabled us to model vandals and other types of occasional contributors (note that contributors with only one activity make up more than half of all contributors in our sample). We tested the sensitivity of the clustering solution to this decision and found that it is relatively little (for details on this see Appendix).

The annotation of revisions was based on the taxonomy of wiki-work developed in prior works (Kriplean et al. 2008, Arazy et al. 2010), which was already employed as a basis for the large-scale manual annotation task in Antin et al. (2012). Kriplean et al.’s (2008) original taxonomy included ten edit categories, which were refined by Antin et al. (2012) after some pilot testing. We further refined this taxonomy through pilot testing until we generated a comprehensive list of twelve meaningful editorial work types that could be understood and identified by coders, as described in Table I. The unit of analysis for our annotation was at revision level, and each revision could contain multiple types of “editing work”; in other words, we allowed for multi-labeling. For example, a revision could be annotated as both *Delete Substantive Content* and *Add Hyperlinks*. The Appendix provides details on the process of manual annotation study.

Once the training set was created, we used a machine learning algorithm to classify all revisions in our 1,000 article set. Machine learning algorithms build a model based on labeled input and then make predictions; they are useful in tasks that do not lend themselves to the explicit programming of rule-based algorithms. A machine learning algorithm typically employs a set of features – in this case: features of the Wikipedia revision – for making the classification. Through an extensive set of experiments, Daxenberger and Gurevych (2013) have identified the most important features for this task, including features based on: meta data (information extracted from the revision comment, author name, time stamp or other flags); textual features; wiki markup; and language features. We built on this approach, with some modifications. In particular, our unit of analysis was the wiki revision, whereas in the prior work each revision is decomposed into several ‘edits’ (representing distinct local changes to the wiki page). In order to verify that the features are well-suited for our task, we tested their performance on the manually classified data set, using a Random k-Labelsets (RAKEL) classifier. Overall, the performance of this classifier is satisfying with a Micro-F1 score of 0.78. The classifier performs close to human agreement, as shown by the Macro-F1 score of 0.68 (as compared to human agreement of 0.73). The Appendix provides more detail on the automatic classification procedure.

After verifying that a our classifier performs well on test data, we employed it to classify all revisions in our 1,000 article set. This resulted in 689,514 revisions classified with a valid category, contributed by 222,119 distinct participants. Our analysis showed that the distribution of contributors’ activity follows a power law, whereby more than half of the contributors in our sample performed only a single activity, and the most active contributor has performed 3,815 activities across the 1,000 article set. Roughly 12% of all contributors in our sample have been active four times or more, and 10% of all contributors were active in more than one article.

¹The English Wikipedia main topic categorization scheme is developed by the community and is subject to frequent changes; see http://en.wikipedia.org/wiki/Category:Main_topic_classifications.

Category	Description
Move or Create New Article	An article is created or moved
Add Substantive New Content	New information is added, changing the meaning of the article
Delete Substantive Content	Existing information is removed, changing the meaning of the article
Fix Typos and Grammatical Errors	Grammatical, spelling and/or minor formatting errors are corrected
Rephrase Existing Text	Sentences are re-structured for clarity, not changing the article's meaning
Hyperlinks (to other Wikipedia pages)	A link target is changed; a link is added; an existing link is deleted
References (to external sources)	References to external sources are added, deleted or changed
Add or Change Wiki Markup	A text body containing wiki markup is added, deleted or changed
Reorganize Existing Text	One or more text bodies are moved; headings or categories are added or deleted, changing the articles' overall structure
Insert Vandalism	Malicious content is added, text is deleted without any obvious reason
Remove Vandalism	Damage done by a vandal is reverted
Miscellaneous	A change which does not fall under any of the other categories is performed

TABLE I: The 12 edit categories used to annotate the revisions in our data sample.

The contributors in our sample performed various types of activities, where often an activity was associated with several categories from our taxonomy. The most frequent category was the *Wiki Markup* (43% of activities), followed by the *Add Substantive Content* category (30%); and the least frequent were: *Hyperlinks* (2%), *Miscellaneous* (2%), and *New Article* (less than 1% of activities). Please see details in the Appendix.

C. Identifying Prototypical Activity Profiles

Each of the contributors in our sample was represented through a vector listing the number of activities he has made, as well as the activities' categories. We assumed that a contributor may enact different roles at different article co-authoring projects (Gleave et al. 2009), and created several activity profiles for each contributor, one for each article he contributed to.² In total, we created 325,417 activity vectors. For example, a contributor working on a particular article can perform 17 activities with category *Add Substantive New Content*; 13 *Delete Substantive Content* activities; and so on. Given our goal of modeling roles (rather than individual contributors), we normalized the activity profiles, dividing the count of revisions in each category by the overall number of activities made by the particular contributor on the article at hand. Thus, we eliminated distinctions between contributors with varying activity levels.

We then employed a clustering algorithm to group contributors' activity profiles, referring to each cluster's centroid as the prototypical activity profiles. These prototypical profiles are interpreted as emergent roles (Gleave et al. 2009, Liu and Ram 2011). The input to clustering are the contributors' activity profiles, one profile for each article $p_i \in P$ the contributor was active on. Let $e_{u_m, p_i}^1, e_{u_m, p_i}^2, \dots, e_{u_m, p_i}^{12}$ denote the number of each of our twelve edit categories performed by contributor u_m to the article p_i , where e_{u_m, p_i}^T denotes the total number of edits by contributor u_m to article p_i . Then, we defined the activity profile vector of contributor u_m to article p_i as

$$\overrightarrow{prof_{u_m, p_i}} = \left\langle \frac{e_{u_m, p_i}^1}{e_{u_m, p_i}^T}, \frac{e_{u_m, p_i}^2}{e_{u_m, p_i}^T}, \dots, \frac{e_{u_m, p_i}^{12}}{e_{u_m, p_i}^T} \right\rangle.$$

We employed the K-means clustering algorithm with Euclidean distance measure, which aims to partition a set of observations (in our case, a contributor's activity profile) into

k clusters; a cluster's centroid serves as a prototype of the cluster, and each observation belongs to the cluster with the nearest centroid (Jain et al. 1999). We iteratively tested K-means for k clusters, where $k \in [2, 10]$ (a larger number of clusters would be difficult to interpret intuitively). In order to determine the optimal number of clusters, for each value of k , we calculated the cluster *Compactness* and *Separation* metrics for the results of K-means clustering (Liu and Ram 2011, He et al. 2004). *Compactness* is based on the homogeneity of vectors in each cluster (smaller values indicate higher average compactness). *Separation* measures the overall dissimilarity between the clusters (smaller values indicate higher average separation). We combined the two metrics using the *Optimal Cluster Quality (OCQ)* measure (He et al. 2004), giving *Compactness* and *Separation* equal weight. Given that clustering results depend on the selection of initial random seeds, we instantiated the seeds using the K-means++ method (Arthur and Vassilvitskii 2007), and iteratively tested a range of values for the initial seed. The lowest *OCQ* score (indicating the best clustering quality) was obtained for $k = 7$. A plot with the *OCQ* values for different k can be found in the Appendix. Additionally, we qualitatively compared clustering solutions across values of k , by trying to interpret the vectors describing the cluster centroids. This manual analysis confirmed that the clustering solution with $k = 7$ produced centroids that could be interpreted intuitively as emergent roles.

Our findings illustrate the nature of emergent roles, as represented through the activity profiles of clusters' centroids (see Table II). Each cluster was given a representative title, as follows: *All-Round Contributors*; *Quick-and-Dirty Editors*; *Copy-Editors*; *Content Shapers*; *Layout Shapers*; *Watchdogs*; and *Vandals*. The *All-Round Contributors* cluster has the highest percentage of contributors, 41% of all contributors' profiles in our sample are assigned to this cluster. As shown by its centroid, contributors with this role are active in many edit categories, with a slight tendency towards adding content and wiki markup. The *Quick-and-Dirty Editors* cluster (11%) represents contributors with a relatively clear focus on adding new content. However, some of their contributions were labeled as vandalism. Differently from the *Vandalism* cluster which has a clear focus on vandalism activities (see discussion below), here, vandalism activities are coupled with the addition of new content. We assumed that unlike the

²Please see in the Appendix for the verification that assuming individual profiles per articles is sensible.

activities of vandals these are contributions made in good faith, which were often reverted because they were not done properly and did not comply with Wikipedia’s policies (e.g. Neutral Point of View, supporting claims by references, etc.). *Copy-Editors* show a clear tendency towards one activity category, namely fixing grammar and spelling errors. The two clusters representing “shaping” activity contain relatively few profiles: *Content Shapers* (4%) concentrate on activities associated with the (re)organization of content; whereas *Layout Shapers* (6%) focus almost entirely on adding markup to an article. The *Watchdogs* and *Vandals* clusters both have equal size (13% of profiles) and contain contributors with a clear focus on a single edit category, namely inserting or removing vandalism, respectively.

D. The Organization of Work: Stability Analysis of Prototypical Activities

The clustering procedure described above aggregates contributors’ activity profiles, and the resulting solution describes the organization of work in Wikipedia. Here, each cluster corresponds to a particular prototypical activity pattern which corresponds to an emergent role. When using such automatic clustering techniques, we need to ensure that space defined by contributors’ activity vectors naturally organizes into clusters, and thus we performed an analysis of clustering reproducibility. Clustering results may be of low quality in the sense that different clustering approaches may claim to summarize a given data set equally well, and we cannot tell which ones better reflect the intrinsic structure of the data (Bayá and Granitto 2013). The metrics described above (*Compactness*, *Separation* and *OCQ*) are useful in determining the best clustering solution for a K-means algorithm on a given solution space, but cannot generalize to compare clustering solutions across algorithms and different data spaces. Thus, to assess clustering quality, a much more general approach is required. Lange et al. (2004) devised a validation method for detecting the number of arbitrary shaped clusters. They trained a classifier that learned the structure that was found by a clustering algorithm using the “natural groups” produced by the clustering algorithm as labels of the input to the classifier. *Cluster reproducibility*³ measures the classification risk of the labels produced by the clustering.

Following Lange et al. (2004), we calculated the cluster reproducibility, $\bar{S}(A_k)$, where A is the clustering algorithm and k the number of clusters. In several rounds, we split the full data sample randomly in two halves X and Y . The average 0-1 loss between $A_k(Y)$ and a classifier prediction $\phi(Y)$ (ϕ is trained on $A_k(X)$) corresponds to the average dissimilarity of clustering solutions. After normalizing this value by the misclassification rate of a random labeling, we arrived at the cluster stability value, $\bar{S}(A_k)$. Smaller values of $\bar{S}(A_k)$ signify a lower misclassification risk and thus, a higher reproducibility of the clustering solution. Our analysis indicated that cluster

³Please note that Lange et al. (2004), whose approach we adopted, referred to clustering reproducibility as “stability”. In order not to confuse with our analysis of stability across time, we chose to refer here to “clustering reproducibility”.

reproducibility $\bar{S}(A_k)$, for values of $k \in [2, 10]$, reached a local minimum at $k = 7$, corroborating our earlier findings regarding the optimal number of clusters. The clustering reproducibility value $\bar{S}(A_7)$ was 0.31, and the average 0-1 loss between our clustering solution $A_7(Y)$ and a classifier prediction $\phi(Y)$ was 0.27⁴, indicating that the risk of irreproducible clusters in our solution is not high. For $k < 7$, reproducibility values are consistently worse compared to $\bar{S}(A_7)$. For values of $k > 7$, only $k = 9$ and $k = 10$ yield slightly lower values of $\bar{S}(A_k)$. The classifiers we tested for ϕ were SMO (Platt 1998) and C4.5 (Quinlan 1993).

E. The Organization of Work: Stability across Periods in Wikipedia’s Evolution

Wikipedia has gone through two major periods in its evolution: (I) “forming” (from 2001 to the end of 2006) and (II) “establishing” (from January 2007 to 2012) (Halfaker et al. 2012). The first period is characterized by the establishment of the technical infrastructure, the introduction of basic procedures and policies around the organization of work, and rapid growth in the size of the community. The second period is exemplified by the development of a bureaucratic structure (Butler et al. 2008), a greater emphasis on policies, norms, and procedures (Kittur and Kraut 2010), and the formation of a complex organizational structure (Arazy et al. 2014).

In order to test whether the clustering solution describing the organization of work in Wikipedia is stable over time, we split our data for the two periods (2001-2006 and 2007-2012), and for each period we created profiles of contributors’ activities. We applied the same clustering procedure described earlier. There were 96,757 contributors’ activity vectors in the “forming” period and 233,687 in the “establishing” period, where only 5,027 contributors remained active in a particular article across both periods (this translates into an outflow of 95% at the end of the first period, and an inflow of 98% entering the second period). Based on the Euclidian distance between centroids, we were able to align the two clustering solutions, mapping each centroid in one solution to the cluster centroid in the other.

Comparing the clustering solutions for the two periods shows that the prototypical activity profiles are stable across the two periods in Wikipedia’s evolution, as the two clustering solutions are highly similar and align well. The analysis of the distances between clusters’ centroids shows that the average distance is 0.12 (11% of the average centroid distance in the clustering solution on the entire data), indicating that the nature of emergent roles (defined in terms of centroids’ activity profile) changed very little between the two periods.⁵ Figure 1 visualizes the alignment between the clustering solutions for the two periods. We were surprised to find such a high stability in the characteristics of emergent roles, despite fundamental changes in Wikipedia’s governance mechanisms.

⁴Further parameters as listed in Lange et al. (2004): $r = s = 20$ (number of splits/iterations).

⁵The reported alignment of clustering solutions used the Euclidean distance measure. In order to verify the robustness of this alignment between the clustering solutions for the two periods, we confirmed the analysis with the help of the Manhattan distance metric. Using both metrics we arrived at the same alignment between clustering solutions.

Emergent Role \ Edit Category	All-round Contr.	Quick-and-Dirty Ed.	Copy-Editors	Content Shapers	Layout Shapers	Watchdogs	Vandals
Move or Create a New Article	0	0	0	0	0	0	0
Add Substantive New Content	24	77	0	9	0	0	2
Delete Substantive Content	7	5	0	1	0	1	1
Fix Typo(s)/Grammatical Errors	4	2	95	4	1	1	0
Rephrase Existing Text	9	2	2	1	0	0	2
Hyperlinks	3	0	0	0	1	0	0
References	7	0	0	1	1	0	0
Add or Change Wiki Markup	39	2	0	29	97	1	1
Reorganize Existing Text	1	0	0	53	0	0	0
Insert Vandalism	3	12	2	1	0	2	95
Remove Vandalism	2	1	0	1	0	94	0
Miscellaneous	2	0	0	1	0	0	0

TABLE II: Emergent roles as prototypical activity patterns, values are percentages (values above 5% in bold).

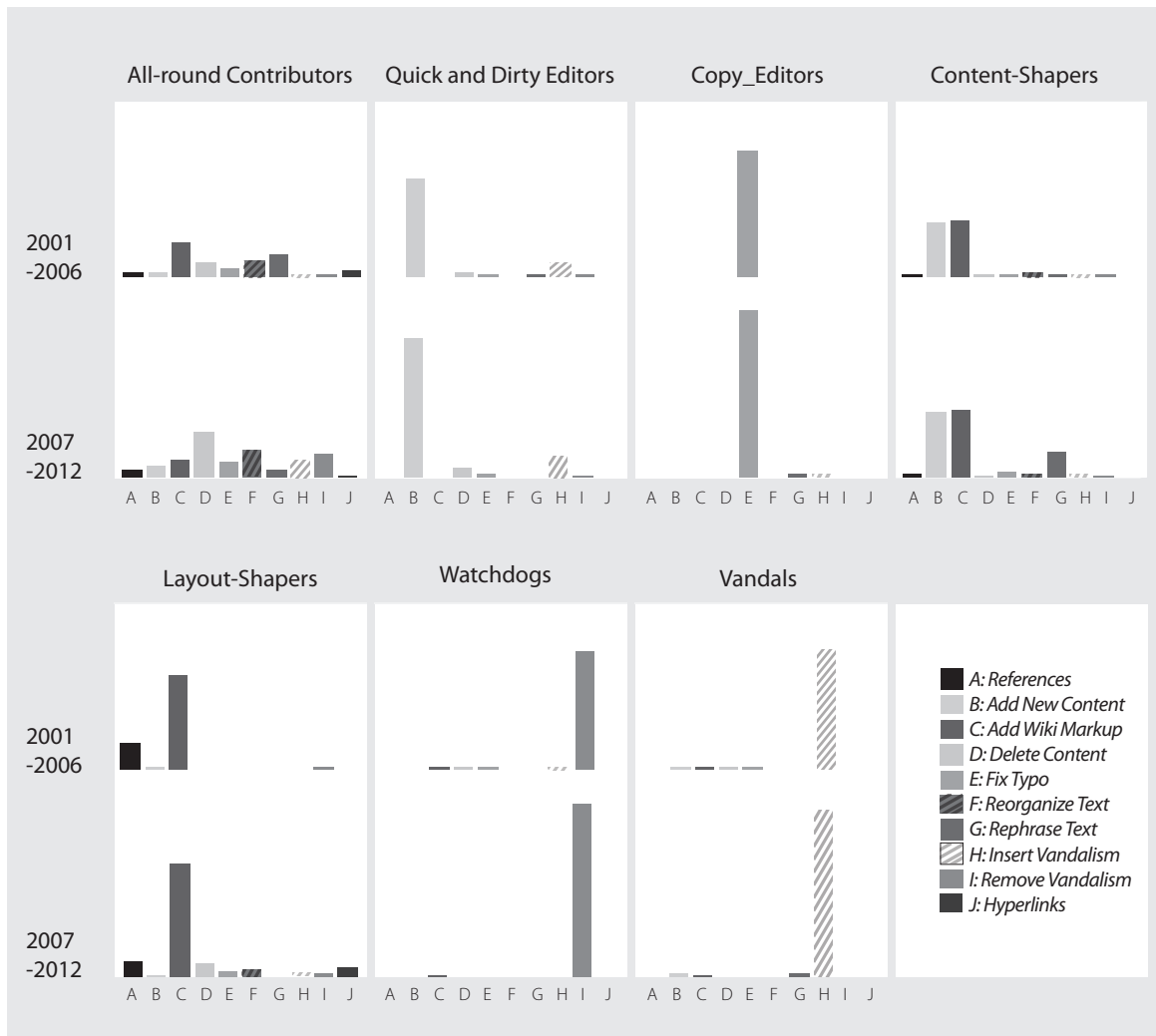


Fig. 1: The characteristics of emergent roles compared for two time periods.

F. Individual-Level Analysis: Contributors' Dynamics

We based our analysis of individuals' dynamics on the comparison between the two periods of Wikipedia's life (2001-2006 and 2007-2012). First, we found evidence for massive fluidity, illustrating the high level of in- and out-flows from the knowledge production process (Faraj et al. 2011). Our results show that 95% of those active in the "forming" period did not continue to the next period, and 98% of those active in the "establishing" period were newcomers. A closer look at the year-by-year attrition revealed that in the early years (2001-2006), 20%-25% continued their participation at the end of the year, and those numbers dropped to approximately 10% in later years (i.e. an outflow of 75%-80% in early years and roughly 90% yearly outflow after 2006). In terms of inflow: in the early years, 92%-94% of the users active in every calendar year were new (i.e. 6%-8% sustained their participation), and the inflow values dropped to roughly 90% in later years.

Building on our previous analyses, and after verifying that the nature of emergent roles is stable over time (i.e. clustering solutions for the two periods are almost identical), we were able to zoom-in on individual contributors and investigate how they transition between roles across the two time periods. We found that among the 5,027 contributors active within an article at both periods, more than 50% changed their role over time. A detailed analysis of role transitions reveals that contributors tend to move towards the *Layout Shapers* role (incoming: 956; leaving: 346), and to a lesser extent to the *Watchdogs* role (482/187). In contrast, contributors tend to leave the *All-Round Contributors* role (378/1236). These role transitions suggest that while the nature of roles is quite stable across time, contributors do change their role within the same article, often taking on more complex co-authoring roles (e.g. *Shapers* or *Watchdogs*). Table III presents the between-periods role transitions.

In sum, the series of analyses we performed sheds light on the nature of emergent roles, as well as on participants' role-taking dynamics. First, we described the nature of emergent roles and validated the stability and robustness of our results (Section III-C, III-D and Appendix). Our analysis of participants' role taking behaviors (Section III-F) shows massive in- and out-flows, indicating that on a year-by-year basis the vast majority of participants flow into and out of the co-production process. Moreover, those continuing their participation across the two periods of Wikipedia's life are likely to change the role they play in the production of a particular article. In the face of this high mobility – as well as the fundamental changes the Wikipedia organization has gone through between the "forming" and "establishing" periods – one could expect that the patterns of contributors' activities would also change across periods. Surprisingly, our results (Section III-E) indicate that the nature of emergent roles remained highly stable across the two time periods. We refer to this interplay between individual-level mobility and organizational-level stability (in terms of the nature of emergent roles) as "Turbulent Stability" in knowledge co-production, and perceive the introduction of this construct as an important theoretical contribution of this study. Please see Figure 2 for an illustration.

G. Investigating Artifact-Centric Coordination

A core question of this work is about how contributors self-organize around a stable set of emergent roles. Prior research points to formal control mechanisms, norms and policies as key coordinating mechanisms. However, our results regarding the stability of emergent roles across periods where Wikipedia's organization differed greatly suggest that an alternative coordination mechanism is possibly at play. Following the artifact-centric line of reasoning (Bolici et al. 2016), we sought to explore the role played by the artifact and its affordances in facilitating coordination. We performed two types of analyses towards this goal. First, we conducted a limited-scope qualitative analysis of the traces left by participants, in terms of the actual activities they performed and the comments they left, shedding light on the rationale behind their activities. Second, we performed a statistical analysis of articles' evolution, comparing role distribution across these stages in articles' development. Given that knowledge products call for different types of work at different stages of their evolution (Benkler 2006), we conjectured that if contributors indeed responded to the needs of the evolving article, they would enact different emergent roles at different stages.

Our qualitative analysis of a random selection of articles in our sample investigated the history of the article's co-production process. Wikipedia maintains a "History" page for each article, tracking its revisions. This page not only allows tracing the detailed activities performed by contributors, it also records their comments. Although these comments are intended to help others understand the nature of the changes made, they also shed light on the contributor's motivation and rationale for making the changes.⁶ Our findings provide evidence that contributors choose to enact a particular role as a response to the work required at a particular point in time. Below, we provide a few examples.

Our first example includes a set of consecutive edits made by the same contributor working on the "Benito Mussolini" article. The series of edits include a variety of actions: the insertion and removal of hyperlinks, copy-editing, and the addition and deletion of content, an activity profile typical of *All-Round Contributors*. The comments made by this contributor, as presented in Figure 3, demonstrate the responsive nature of actions, where the contributor chooses to fix flaws and refine the articulation.

Next, we provide indirect evidence for the action of a *Quick-and-Dirty Editor*. These editors are characterized by additions of content, often including mistakes (and thus occasionally tagged as vandalism). They are content-oriented and care less for Wikipedia's standards and norms (Arazy et al. 2011), thus, they are not likely to exert extra effort in adding comments. Nonetheless, a comment by a contributor correcting such quick-and-dirty edit in the "Bianca Jackson" article illustrates the nature of this emergent role and provides evidence for how editors are responding to the state of the artifact:

"Reverted good faith edits by 146.199.244.239 (talk): Lorraine not her stepmother, she was grown up before

⁶Note that these comments we also used by our automatic revision classification algorithm to help determining the revision category.

To \ From	All-round Contr.	Quick-and-Dirty Ed.	Copy-Editors	Content Shapers	Layout Shapers	Watchdogs	Vandals	Sums
From Period A								
All-round Contr.	350	40	93	387	476	231	9	1,586
Quick-and-Dirty Editors	17	25	5	48	15	11	9	130
Copy-Editors	47	8	86	67	123	41	1	373
Content Shapers	172	54	84	674	258	119	20	1,381
Layout Shapers	80	12	47	125	240	79	3	586
Watchdogs	57	2	20	31	76	721	1	908
Vandals	5	4	3	18	8	1	24	63
Sums	728	145	338	1,350	1,196	1,203	67	5,027

TABLE III: Role transitions for the set of 5,027 contributors active in both periods. Values on the diagonal (in bold) represent contributors maintaining the same emergent role.

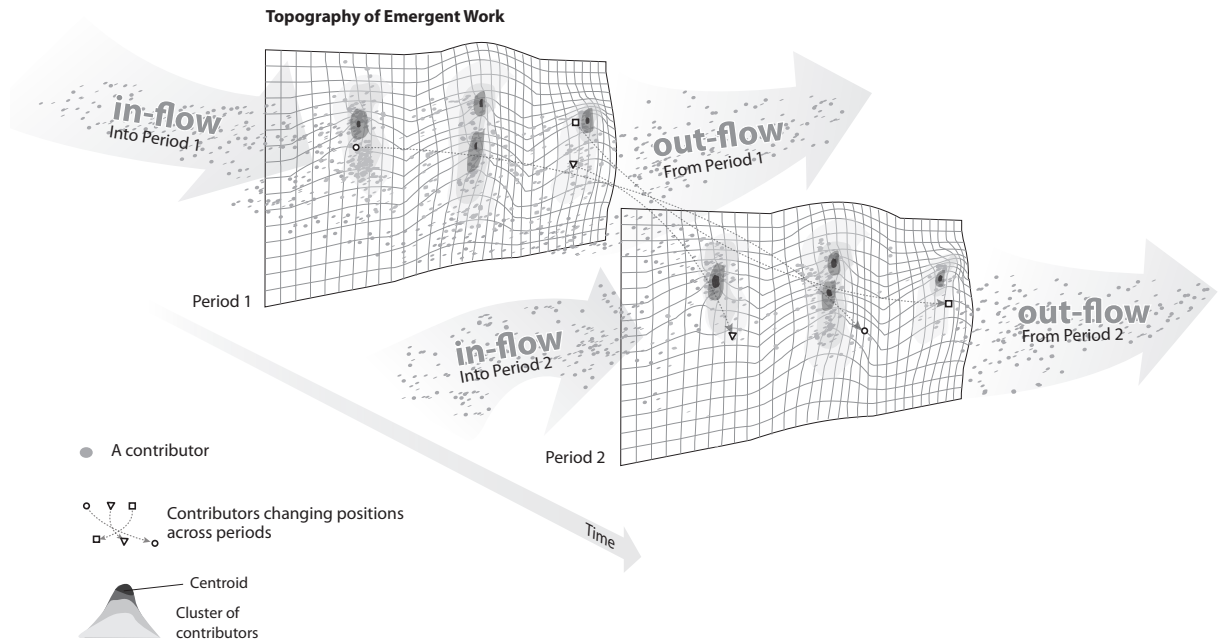


Fig. 2: Illustration of our concept of “Turbulent Stability” (based on fictional data): system-level stability in the face of individual level mobility. The grids reflect the activity spaces at two points in time. Darker regions (‘mountain tops’) represent cluster centroids corresponding to emergent roles.

06:42, 14 October 2012 Brothernight (talk | contribs) .. (143,589 bytes) (+37) .. (→Expulsion from the Italian Socialist Party: Added citation needed tag. -- ~~~~)

05:22, 14 October 2012 Brothernight (talk | contribs) .. (143,552 bytes) (+128) .. (→In popular culture: Corrected the spelling of the insult chiseler and explained why it is written that way. -- ~~~~)

05:12, 14 October 2012 Brothernight (talk | contribs) .. (143,424 bytes) (+300)

04:33, 14 October 2012 Brothernight (talk | contribs) m .. (143,124 bytes) (+61)

04:30, 14 October 2012 Brothernight (talk | contribs) .. (143,063 bytes) (-110) .. (Made several changes, one of which was a repeat. -- ~~~~)

03:54, 14 October 2012 Brothernight (talk | contribs) .. (143,173 bytes) (+66) .. (Added an appropriate link. -- ~~~~)

03:18, 14 October 2012 852derek852 (talk | contribs) ... (143,107 bytes) (+101) ... (Removed vandalism)

03:10, 14 October 2012 Brothernight (talk | contribs) m .. (143,006 bytes) (+7)

03:01, 14 October 2012 Brothernight (talk | contribs) .. (142,999 bytes) (+73) .. (Added a link to what really happened with Mussolini's futile attempt to deal with the Pontine Marshes. -- ~~~~)

02:11, 14 October 2012 Brothernight (talk | contribs) .. (142,926 bytes) (-84) .. (Made a heartfelt attempt to explain what Italian Fascism was all about. -- ~~~~)

01:17, 14 October 2012 Brothernight (talk | contribs) .. (143,010 bytes) (-17) .. (Removed the linked word "corporatism" as it is meaningless to most speakers of English in this particular context. -- ~~~~)

Fig. 3: A set of actions performed by a typical All-Round Contributor, interrupted by one different contributor.

she knew Lorraine existed."

Copy-Editors often make small changes at the word-level, correcting small errors. Below are examples of editors acting as *Copy-Editors* in the "Avro Canada CF-105 Arrow" and "Newcastle upon Tyne" articles:

"Delete double period"

"Took kilometres out of () and put miles in. Since it is located in the United Kingdom, kilometres should be the default measurement."

Content Shapers are mostly involved in improving the organization of content on the wiki page, moving sections around, removing duplications and often categorizing articles. The comments below illustrate the actions of two *Content Shapers* reorganizing the order of content on the "Kurdistan Workers' Party" article:

"moved into chrono[logical] order"

"rv. it was in chronological order (a reverse one, to be precise), before this edit"

Layout Shapers are mostly involved in changing wiki mark-up: formatting lists, tables, links, section headers, etc. Below is an example for an action of a typical *Layout Shaper* working on the "Benito Mussolini" article:

"... added list formatting, removed flash movie"

Finally, *Watchdogs* are involved in the correction of vandalism. Thus by definition their work is responsive in nature and involves close monitoring of the artifact. While vandals are not likely to leave comments describing their actions, *Watchdogs* often do leave comment, as exemplified by the comment to the "Bianca Jackson" article above, as well as by an editor working on the "Economy of Angola" article:

"Reverting possible vandalism by 2605:6000:8281:BA00:3C5A:1B48:EB81:29DC to version by 2001:4C50:21D:F400:B565:F3EB:735B:522. False positive? Report it..."

In addition to the evidence for artifact-centric coordination derived from our qualitative analysis, we sought a quantitative evidence for the relationship between an articles' state and a contributors' decision to enact particular roles. Our next analysis thus investigated the distribution of emergent roles at different stages in an article's evolution. We focused our analysis on the 250 articles from our sample that have passed through all stages in an article's life to reach maturity (more than 1,000 revisions). We portioned each article in our sub-sample into these stages. For example, an article with 1,253 revisions was partitioned into: 1-10 revisions, 11-100 revisions, 101-1,000 revisions, and 1,001-1,253 revisions. We then created contributor-article-partition activity vectors for each article partition, and associated each vector with a cluster centroid (i.e. emergent role). Next, for each article in a partition, we listed the percentage of contributors playing each role. We then aggregated the data for each of the partitions, averaging each role's percentages. Comparing the means for each role across the four partitions shows whether an emergent role changes its relative concentration between article's life stages.

Findings from our analysis reveal some significant differences between the four stages, as illustrated in Figure 4. To better assess those differences, we applied an ANOVA model

(Edwards 1979) and found that the differences between stages are statistically significant ($p < 0.001$) for all of the roles. Three roles – *All-Round Contributors*, *Content Shapers* and *Layout Shapers* – start at relatively high proportion and then generally decline with an articles' maturity, whereas three other roles – *Quick-and-Dirty Editors*, *Vandals* and *Watchdogs* – show a constant increase in proportion. A Least Significant Difference (LSD) post-hoc analysis (Williams and Abdi 2010) shows that these increase/decrease trends are to a large extent consistent, such that pair-wise differences for consecutive life cycle stages are statistically significant (for all roles across most stage transitions; $p < 0.001$). However, for most roles the relative proportion stabilize at the *growth* stage, such that the differences between *growth* and *maturity* stages were statistically insignificant (except for the roles of *All-Round Contributors* and *Watchdogs*, where the trend continues into the *maturity* stage; $p < 0.001$).⁷ The relative proportion of *All-Round Contributors* is highest across all life cycle stages, yet this proportion decreases as articles mature (starting with 60% at the *inception* stage and ending at 36% at the *maturity* stage). This suggests that articles at more mature stages required more specialized work. The shaping roles (*Content Shapers* and *Layout Shapers*) start at moderately high proportions (together 20% at the *inception* stage) and end low (together 8% at the *maturity* stage), indicating that delineating the article's structure is more important in the early stages of an article's life cycle. The *Quick-and-Dirty Editors* role takes the opposite trajectory: 7% at *inception* and growing constantly to 12% at *maturity*. This could possibly be explained by the fact that more mature articles attract a broader readership, and consequently invite one-time editors. The constant rise in the proportion of *Watchdogs* (1%-4%-12%-16%) follows closely the increase in the *Vandals* population, suggesting that *Watchdogs* react to the rising concern of vandalism. Together, these results – described in Figure 4 – provide indirect quantitative evidence for the role of the artifact in determining the types of emergent roles that contributors choose to enact.

IV. DISCUSSION

While scholars investigating online co-production communities are beginning to uncover the nature of emergent work, much is still unknown. At the organizational level, there have been attempts to characterize the bundles of tasks that make up emergent roles (Fisher et al. 2006, Gleave et al. 2009, Liu and Ram 2011, Welser et al. 2011), yet we lack an understanding about the extent to which these roles are robust and stable over time. At the individual level, not only do existing conceptualizations disagree on the extent to which emergent roles are fluid and transient, there has also been a scarcity of empirical investigations validating these conceptualizations (Faraj et al. 2011, Kane et al. 2014, Panciera et al. 2009). Scholars have suggested that new organizational forms combine both stable and dynamic elements (Schreyögg and Sydow 2010, Langley et al. 2013, Farjoun 2010) and have called for a multi-level analysis that would explore these tensions (Gaskin et al. 2014).

⁷With one exception for *Copy-Editors*, where the proportion remained constant (at 12%) through the *creation*, *growth*, and *maturity* stages.

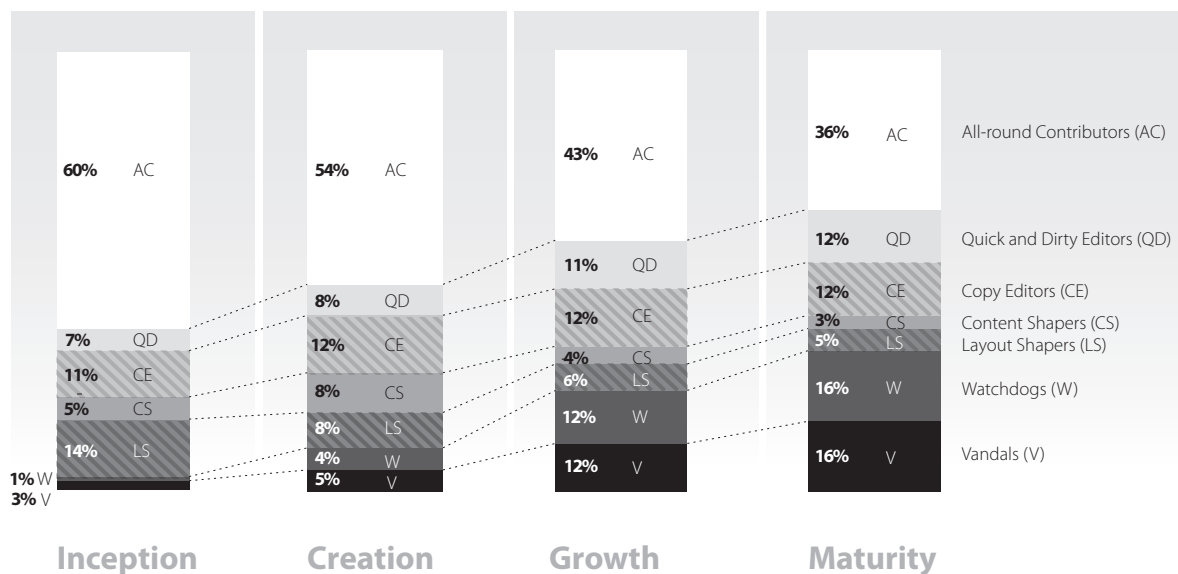


Fig. 4: The distribution of emergent roles across the four stages of articles' life cycle.

Yet empirical validations have been slow to follow. Our study has made inroads towards filling in these gaps. In the sections that follow, we argue our study's contributions to the various literatures we drew on.

A. Emergent Roles in Knowledge Co-Production

While role theorists allude to the notion of emergent roles, to date, the theoretical understanding of the temporal dynamics of emergent roles is far from comprehensive. Turner (1978, p. 1) describes roles that are “put on and taken off like clothing” without lasting effect on personality. Faraj et al. (2011, p. 1231) call for “the enactment of temporary sets of behaviors that are volitionally engaged in, self-defined, and inductively created for the purposes of the online community”. Others (Gleave et al. 2009, Welser et al. 2011) advocate for a broader understanding of “role ecologies”, stating that “we should aim for systems that can assess degree of role performance, and, ideally, to track assessment across time to monitor role change” (Welser et al. 2011, p. 128). Our empirical study represents a preliminary step toward this aim.

Extant conceptualizations of online production communities have pointed to two prototypical role behaviors: (a) those adding new content; and (b) those “shaping” existing content (Kane et al. 2014, Majchrzak et al. 2013, Yates et al. 2010). Such roles were operationalized using surveys of contributors' perceptions. However, these operationalizations are divorced from taxonomies of wiki-work (Antin et al. 2012, Kriplean et al. 2008) and fail to account for a variety of disparate activities (e.g. copyediting, adding hyperlinks). Our study's findings point to seven emergent roles as they are practiced in Wikipedia: *All-Around Contributors*, *Quick-and-Dirty Editors*, *Copyeditors*, *Content Shapers*, *Layout Shapers*, *Watchdogs*, and *Vandals*. While the majority of these roles map well to those identified in earlier studies that employed bottom-up approaches (Liu and Ram 2011), we were also able to identify previously unnoticed prototypical behaviors that correspond

to roles in conceptual frameworks (Majchrzak et al. 2013, Yates et al. 2010). Our findings thus help to bridge these two literatures. In particular, our results indicate that there are two distinct forms of shapers. The first, *Content Shapers*, concentrates on activities associated with the reorganization of text (and to a lesser extent: adding wiki markup); while the second, *Layout Shapers*, is almost entirely about adding wiki markup. Majchrzak et al. (2013) argue that “recognizing and clarifying the role of shaping allows us to theorize new ways in which knowledge resources affect knowledge reuse,” and our empirical findings contribute towards this goal.

We also make a methodological contribution by introducing techniques that have not been previously used in the study of online production communities: studies in the area rarely examine clustering reproducibility (Lange et al. 2004) and assume, rather than validate, that clustering results represent natural groupings in the data. In addition, our method for profiling participants' activities employed machine learning that has some unique advantages over the previously-used rule-based approach (Liu and Ram 2011).

B. The Dualistic Nature of Self-Organizing Knowledge Co-Production

The main contribution of this study is in discovering the “turbulent stability” of emergent roles and unraveling the dualistic nature of self-organized knowledge co-production. This duality between individual-level turbulence and organization-level stability in emergent role behaviors contributes to our understanding of new forms of organizing for knowledge production. Scholars have called for research on the emergent nature of work in post-industrial production that takes place outside traditional boundaries and requires “assembling knowledge in ways that we have not done before while facing new tasks environments” (Okhuysen and Bechky 2009, p. 496). Schreyögg and Sydow (2010, p. 15) for instance, call for

developing theoretical frameworks that “overcome the one-sided ideals of organizational fluidity and full flexibility on the one hand and the advantages of bureaucratic replication on the other hand”. Instead, they suggest conceiving contemporary organizations in terms of dual, dialectic, or paradoxical processes.

Organizational theorists have highlighted the need to develop refined theoretical understandings of the processes underlying, enabling, and sustaining emergent work (Orlikowski 2000, Hernes 2014, Farjoun 2010). To date, the literature on online production communities has paid particular attention to the role of community-based governance mechanisms, such as norms and policies (Forte et al. 2009, Kittur et al. 2007), coordination processes (Shah 2006, Demil and Lecocq 2006), and quality control procedures (Stvilia et al. 2008). What seems to be missing from this scholarly discourse is attention to the emergent nature of co-production work itself (Faraj et al. 2011, Kane et al. 2014, Majchrzak et al. 2013, Choi et al. 2010, Panciera et al. 2009). We note that while some scholars have previously questioned the ability of the “crowd” to self-select and replenish its role players (Goldman 2009), this study illustrates that massive amounts of distinct contributors *can* enact emergent roles, shed them, and sometimes transition to new roles, and yet, at the organizational level, these roles remain stable and well-defined over time.

Our findings regarding the persistence of emergent roles demonstrate that stability in knowledge co-production work can emerge independent of formal organizational structures. In particular, we find evidence for the rise of stability at the system level (i.e. emergent role behaviors) in the face of high levels of role mobility at the individual level (i.e. transitions in and out of co-production, as well as between emergent roles). Given the high level of fluidity and mobility in participants’ role enactment during knowledge co-production (Faraj et al. 2011, Kane et al. 2014), one might expect that these high levels of change in individual’s role-taking-and-shedding will translate into unstable role behaviors. This is especially true when the organization goes through fundamental changes in structure and governance procedures (as was the case for Wikipedia) (Halfaker et al. 2012). Hence, our results about the stability in emergent roles behavior are quite surprising, and are of particular importance.

Our findings regarding the turbulent stability of emergent roles also contribute to the discussion in the literature on new forms of organizing. In particular, our results inform the debate between the bureaucratic and “open” perspectives, suggesting a more nuanced synthesis of the two approaches. To date, the literature on new forms of organizations moves between two extremes: on one hand, scholars claim that such co-production communities are built on unprecedented freedom and very little structure, and thus represent a new form of organization; on the other hand, scholars focus on how formal structures rise in such settings, transforming the organization into familiar bureaucracies (Shaw and Hill 2014, Puranam et al. 2014). We note that even in periods when the organization has become more formal and institutionalized (i.e. Wikipedia in the period between 2007-2012), it kept “production” work free from workflow constraints associated with traditional knowledge

production. Strikingly, the way in which work organically self-organizes into emergent role behaviors has remained consistent over the two distinct periods in Wikipedia’s evolution. Thus, our study stresses the value of a more hybrid and nuanced approach to the understanding of the dynamic processes underlying online production communities (Schreyögg and Sydow 2010).

C. Artifact-Centric Production

We argue that the enabling means for the turbulent stability in emergent roles is artifact-centric co-production. Our finding that emergent role behaviors remain stable over two distinct periods in Wikipedia’s life, despite fundamental organizational changes in governance structure and norms, rules out the explanation of traditional role theories that the mechanisms stabilizing role behaviors are persistent norms, policies, or social interactions. The artifact-centric perspective highlights the role of the co-produced artifact as a central mechanism facilitating tacit coordination in peer-production (Howison and Crowston 2014) and supporting “mutual adjustment” (Mintzberg 1992). We maintain that the artifact and its affordances served as a stabilizing factor amid the turbulence of individuals’ role mobility.

Orlikowski and colleagues have called for the development of a unified conceptualization that encompasses the social and the material under the label of sociomateriality (Orlikowski and Scott 2008). In this study, we provide empirical evidence in support of this perspective. More specifically, we found support for the conceptualization introduced by Faraj et al. (2011), describing emergent role-taking as a response to rising tensions around the co-production of an artifact. We use evidence from contributors’ comments indicating that, when enacting a particular emergent role, they respond to the current state and needs of the artifact. For instance, a contributor enacting the emergent role of *Layout Shaper* in response to tension fluctuations commented: “Put entries in chronological order, added list formatting, removed flash movie”, bringing empirical evidence to the ideas postulated by Faraj et al. (2011, p. 1231): “When convergence is so incomplete and temporary that ideas become disorganized, a participant may create an organizer role for herself by organizing ideas that others have posted”. Moreover, our study explored whether a product’s maturity level influences participants’ role-taking behaviors (Kane et al. 2014). Faraj and Azad (2012) emphasize the temporal dimension of the artifact-centric perspective, viewing the artifact not as a static object to which the ‘social’ responds but rather as an evolving relational construct. Our findings imply that when enacting a particular emergent role, participants consider the maturity level of the product. Our study thus provides corroboration for the artifact-centric perspective, and demonstrates the importance of the artifact in directing emergent work.

Our findings also have implications for role theories. Traditional theories of organizational roles assume a “stage” (Goffman 1963) where roles are performed, viewed, monitored, and negotiated with others. Co-production communities offer a novel interplay between “front stage” and “back stage”: at the

front stage is the visible artifact-centric co-production work, whereas at the back stage are workers, often anonymous, and their emergent roles (Zammuto et al. 2007). Such settings create a unique dynamic of choices, where role-taking decisions are less susceptible to social comparison or herding considerations (Faraj et al. 2011).

We argue that the affordances provided by the socio-technical system underlying the online community is a central factor facilitating artifact-centric co-production. Recent years have seen a growth in the investigation of these affordances (Zammuto et al. 2007, Majchrzak and Markus 2012), specifically of social media and wikis (Treem and Leonardi 2012, Majchrzak et al. 2013). These studies have tried to identify the specific features that enable (Ziaie 2015) or forestall (Majchrzak 2009, Yeo and Arazy 2012) online knowledge collaboration. We suggest that the key enabling affordances for turbulent stability of emergent roles are: work process visibility (Treem and Leonardi 2012, Wagner 2004); unconstrained workflow (Wagner 2004); choice of anonymity (Gabrielle 2012); task decomposability (Lakhani et al. 2013, Baldwin and von Hippel 2011); and open boundaries (Zammuto et al. 2007). In order to validate the boundary conditions of this study's findings, we call for future research into other co-production communities (e.g. the GitHub platform supporting open source software development).

Finally, we propose that in addition to affordances, the clear definition of the nature of the end product is key to enabling artifact-centric production (Zhu et al. 2012). The "five pillars" of Wikipedia make the nature of the co-produced artifact clear: it is an encyclopedic entry that should provide a brief overview of a topic, state facts rather than opinions, provide support for these statements, and include only copyright-free material. In fact, scholars have attributed Wikipedia's success to this clarity in artifact requirements (Hill 2013). This high level of clarity serves as a "boundary infrastructure" (Bowker and Star 1999) that stabilizes and facilitates tacit coordination. While Wikipedia hosts millions of co-produced artifacts, the understanding of artifact requirements is shared between all participants, over a decade of co-production work. We note that by-and-large, prior studies on "boundary infrastructure" and "boundary objects" were able to demonstrate their effect in enabling communication and coordination only within a restricted setting, during a relatively short period, and with a small number of participants (Carlile 2002, Kellogg et al. 2006). Our study enriches this literature by demonstrating the effects of "boundary infrastructure" in a much broader and dynamic setting.

D. Practical and Managerial Implications

Recent years have witnessed an increase in the types of knowledge-based products co-created in self-organizing online communities, such as open source software, community-based maps, product designs, and the development of scientific knowledge through the aggregation of citizen's contributions. Our findings have important practical implications for designers and administrators of these online communities. Many co-production communities struggle with the decisions of how

much structure to impose on the co-production process and whether or not to formalize roles. Our findings demonstrate that it is possible to achieve stability in the overall organization of knowledge work while avoiding explicit roles prescriptions in the production space and allowing community members freedom in self-selecting their level and form of participation. It is essential for platform designers to build into the IT platform affordances that enable this turbulent stability. For instance, guiding participants to tasks without imposing restrictions on co-production workflows requires making the evolving requirements of the artifact more visible.

More specific practical implications of our study relate to the "statistical machinery" developed for identifying participants' emergent roles. In particular, our methods can be employed to develop tools that track contributors' activities, identify for them tasks of interest (Zhang et al. 2014), and offer them "career guidance". That is, rather than simply encouraging participants to become more involved – which is implied by extant frameworks such as 'Legitimate Peripheral Participation' (Lave and Wenger 1991) or 'Reader-to-Leader' (Preece and Shneiderman 2009) – we propose that participants be offered specific, personalized non-binding guidance regarding the nature of tasks most relevant for them.

Beyond online communities, key principles from the community-based peer-production model have recently begun "spilling over" into traditional organizations (Lifshitz-Assaf 2016). Many companies use wikis as a knowledge management tool (Majchrzak et al. 2013, Arazy and Nov 2010), particularly for developing organizational encyclopedias and knowledge sharing tools (Arazy and Gellatly 2013, 2016), adopting, in part the organic processes that typify wiki-based collaboration. Similarly, some technology companies participate in open source software development.

However, only few have gone through the transformation needed in order to truly adopt the principles of peer production for their internal research and development processes (Lifshitz-Assaf 2016, Levina et al. 2014) or for their organizational design (Puranam and Håkansson 2015, Van De Kamp 2014).

V. CONCLUSION

Following a call to investigate the temporal dimension in complex and evolving socio-technical systems (Orlikowski and Yates 2002), our study explored the temporal dynamics of emergent roles. We investigated a large number of co-production projects within one setting (i.e. Wikipedia). The advantage of this research method is that it allowed for substantial variation and at the same time controlled for exogenous factors that might have hindered a cross-organizational study. However, we do acknowledge potential concerns regarding generalizability. Our findings suggest several theoretical contributions to our understanding of: (I) the nature of emergent roles; (II) the dualistic nature of self-organizing knowledge co-production; and (III) artifact-centric co-production. Notwithstanding our study's contribution, it represents only a preliminary investigation of emergent roles in online co-production communities, and thus leaves much room for future

research. We investigated a large number of co-production projects within one setting (i.e. Wikipedia). The advantage of this research method is that it allowed for substantial variation and at the same time controlled for exogenous factors that might have hindered a cross-organizational study. However, we do acknowledge the potential concerns regarding generalizability to other settings, and stress the need to extend the investigation of emergent roles to other online communities. In addition, while our study does show that individuals' transition between roles, future research is warranted in order to reveal the intricacies of contributors temporal dynamics between emergent roles and across knowledge-based products. From a methodological perspective, we propose that additional approaches may complement the statistical methods applied in this study to offer a more comprehensive description of emergent work dynamics. For example, scholars have suggested that sequencing techniques from bioinformatics and related fields could be applied to the investigation of the temporal dynamics of socio-technical systems (Gaskin et al. 2014) and in particular to the study of Wikipedia (Keegan et al. 2016). We, thus call for future research that would use sequencing methods to reveal emergent patterns of work (i.e. the 'DNA' of co-production).

Whereas our study has focused on work patterns that organically emerge, we do acknowledge that formal organizational structures may also affect the way in which self-organized work emerges. For example, the relationship between a participant's formal role in the community (namely, his position on the periphery-core continuum, power, status) and the emergent role she enacts, is a potential additional important factor. Whereas some empirical studies have demonstrated that a participant's position in the organization is not directly linked to the set of behaviors he enacts (Yates et al. 2010), other scholars have argued that for a strong connection between formal and emergent roles (Levina and Arriaga 2014). We, thus, call for future research that would explore the extent to which the temporal dynamics of formal and emergent roles are interconnected.

We also recognize that other potential explanations are plausible and alternative theoretical perspectives could be employed to shed light on the nature of emergent co-production work. Namely, this study's findings regarding the spontaneous emergence of system-level order (or stability) form complex dynamic interactions between self-organizing agents imply that Wikipedia could be viewed as a complex adaptive system. Complexity theory has been applied to explain a host of organizational phenomena (Chiles et al. 2004, McKelvey 2008), and in particular IT-mediated social participation (Nan and Lu 2014). We believe that novel insights could be gained by viewing knowledge co-production through the lens of complexity theory, and we call for future research that would employ this conceptualization for understanding of emergent work within peer-production.

In conclusion, we believe that online co-production communities represent a fascinating research area. Understanding how emergent roles are enacted in response to the needs of the artifact, as well as the mechanisms enabling such artifact-centric co-production, are key to theorizing online

knowledge collaboration. Practitioners interested in leveraging the potential of emergent work within corporate walls are encouraged to take note of our findings regarding the mechanisms for balancing openness and control.

ACKNOWLEDGMENT

This work was partially supported by SSHRC Insight Grant 435-2013-0624 and by NSF Award ACI-1322218. We thank Adam Balila for work on data processing and Carlos Fiorentino for his contribution to the graphical design of figures. We also wish to thank Prasanna Tambe and the participants of the 8th International Process Symposium for insightful comments on earlier drafts of this manuscript. The second author acknowledges support from the German Federal Ministry of Education and Research (BMBF) under the promotional reference 01UG1416B (CEDIFOR). The fifth author acknowledges support from the Volkswagen Foundation as part of the Lichtenberg-Professorship Program under grant №1/82806. Some of the experiments covered in this work have been conducted as part of the PhD thesis of the second author.

REFERENCES

- Aaltonen, A., J. Kallinikos. 2013. Coordination and Learning in Wikipedia: Revisiting the dynamics of exploitation and exploration. *Managing 'Human Resources' by Exploiting and Exploring People's Potentials* 161–192.
- Antin, J., C. Cheshire, O. Nov. 2012. Technology-Mediated Contributions: Editing Behaviors Among New Wikipedians. *Proceedings of CSCW*. 373–382.
- Arazy, O., E. Stroulia, S. Ruecker, C. Arias, C. Fiorentino, V. Ganey, T. Yau. 2010. Recognizing contributions in wikis: Authorship categories, algorithms, and visualizations. *Journal of the American Society for Information Science and Technology* 61(6) 1166–1179.
- Arazy, O., I. Gellatly. 2013. Corporate Wikis: The Effects of Owners' Motivation and Behavior on Group Members' Engagement. *Journal of Management Information Systems* 29(3) 87–116.
- Arazy, O., I. Gellatly. 2016. Motivation to share knowledge using wiki technology and the moderating effect of role perceptions. *Journal of the Association for Information Science and Technology* forthcoming.
- Arazy, O., O. Nov. 2010. Determinants of Wikipedia Quality: the Roles of Global and Local Contribution Inequality. *Proceedings of CSCW*. 233–236.
- Arazy, O., O. Nov, F. Ortega. 2014. The [Wikipedia] World is Not Flat: On The Organizational Structure of Online Production Communities. *Proceedings of the 22nd European Conference on Information Systems*.
- Arazy, O., O. Nov, R. Patterson, L. Yeo. 2011. Information Quality in Wikipedia: The Effects of Group Composition and Task Conflict. *Journal of Management Information Systems* 27(4) 71–98.
- Arazy, O., F. Ortega, O. Nov, L. Yeo, A. Balila. 2015. Functional Roles and Career Paths in Wikipedia. *Proceedings of CSCW*.
- Arthur, D., S. Vassilvitskii. 2007. K-means++: The Advantages of Careful Seeding. *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms*. 1027–1035.
- Baldwin, C., E. von Hippel. 2011. Modeling a Paradigm Shift: From Producer Innovation to User and Open Collaborative Innovation. *Organization Science* 22(6) 1399 – 1417.

- Bayá, A. E., P. M. Granitto. 2013. How Many Clusters: A Validation Index for Arbitrary-Shaped Clusters. *IEEE/ACM Trans. on Computational Biology and Bioinformatics* **10**(2) 401–414.
- Bechky, B. A. 2006. Gaffers, Gofers, and Grips: Role-Based Coordination in Temporary Organizations. *Organization Science* **17**(1) 3–21.
- Benkler, Y. 2006. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. Yale University Press.
- Bolici, F., J. Howison, K. Crowston. 2016. Stigmergic coordination in FLOSS development teams: Integrating explicit and implicit mechanisms. *Cognitive Systems Research* **38** 14–22.
- Bowker, G. C., S. L. Star. 1999. *Sorting things out: Classification and its consequences*. MIT Press.
- Burke, M., R. Kraut. 2008. Mopping Up: Modeling Wikipedia Promotion Decisions. *Proceedings of CSCW*. 27–36.
- Butler, B., E. Joyce, J. Pike. 2008. Don't look now, but we've created a bureaucracy: the nature and roles of policies and rules in Wikipedia. *Proceedings of the CHI conference on human factors in computing systems*. 1101–1110.
- Carlile, P. R. 2002. A pragmatic view of knowledge and boundaries: Boundary objects in new product development. *Organization Science* **13**(4) 442–455.
- Cetina, K. K. 1999. *Epistemic Cultures: How the Sciences Make Knowledge*. Harvard University Press.
- Chandler, A. D. 1962. *Strategy and structure*. MIT Press, Cambridge.
- Chiles, T. H., A. D. Meyer, T. J. Hench. 2004. Organizational Emergence: The Origin and Transformation of Branson, Missouri's Musical Theaters. *Organization Science* **15**(5) 499–519.
- Choi, B., K. Alexander, R. E. Kraut, J. M. Levine. 2010. Socialization Tactics in Wikipedia and Their Effects. *Proceedings of CSCW*. 107–116.
- Cohen, L. E. 2013. Assembling Jobs: A Model of How Tasks Are Bundled Into and Across Jobs. *Organization Science* **24**(2) 432–454.
- Daxenberger, J., I. Gurevych. 2013. Automatically Classifying Edit Categories in Wikipedia Revisions. *Proceedings of EMNLP*. 578–589.
- Demil, B., X. Lecocq. 2006. Neither Market nor Hierarchy nor Network: The Emergence of Bazaar Governance. *Organization Studies* **27**(10) 1447–1466.
- Edwards, A. L. 1979. *Multiple regression and the analysis of variance and covariance*. WH Freeman.
- Faraj, S., S. Jarvenpaa, A. Majchrzak. 2011. Knowledge collaboration in online communities. *Organization Science* **22**(5) 1224 – 1239.
- Faraj, S., B. Azad. 2012. The materiality of technology: An affordance perspective. P. M. Leonardi, B. A. Nardi, J. Kallinikos, eds., *Materiality and organizing: Social interaction in a technological world*. Oxford University Press, 237–258.
- Farjoun, M. 2010. Beyond Dualism: Stability and Change As a Duality. *Academy of Management Review* **35**(2) 202–225.
- Fisher, D., M. Smith, H. T. Welsler. 2006. You are who you talk to: Detecting roles in usenet newsgroups. *Proceedings of the Annual Hawaii International Conference on System Sciences*. 59.2.
- Forte, A., V. Larco, A. Bruckman. 2009. Decentralization in Wikipedia Governance. *Journal of Management Information Systems* **26**(1) 49–72.
- Gabrielle, C. 2012. Our Weirdness Is Free, The logic of Anonymous - online army, agent of chaos, and seeker of justice. URL https://www.canopycanopycanopy.com/contents/our_weirdness_is_free.
- Galison, P. 1997. *Image and logic: A material culture of microphysics*. University of Chicago Press.
- Gaskin, J., N. Berente, K. Lyytinen, Y. Yoo. 2014. Toward Generalizable Sociomaterial Inquiry: A Computational Approach for Zooming In and Out of Sociomaterial Routines. *MIS Quarterly* **38**(3) 849–871.
- Gleave, E., H. T. Welsler, T. M. Lento, M. Smith. 2009. A conceptual and operational definition of 'social role' in online community. *Processings of the Hawaii International Conference on System Sciences*. 1–11.
- Goffman, E. 1963. *Stigma: Notes on the management of spoiled identity*. Prentice Hall.
- Goffman, E. 1961. *Encounters: Two Studies in the Sociology of Interaction*. Bobbs-Merrill.
- Goldman, E. 2009. Wikipedia's Labor Squeeze and its Consequences. *Journal of Telecommunications and High Technology Law* **8**.
- Halfaker, A., R. S. Geiger, J. T. Morgan, J. Riedl. 2012. The Rise and Decline of an Open Collaboration System: How Wikipedia's Reaction to Popularity Is Causing Its Decline. *American Behavioral Scientist* **57**(5) 664–688.
- Hallerstede, S. 2013. *Managing the Lifecycle of Open Innovation Platforms*. Springer Science & Business Media.
- He, J., A.-H. Tan, C.-L. Tan, S.-Y. Sung. 2004. On Quantitative Evaluation of Clustering Systems. *Clustering and Information Retrieval*. Springer US, 105–133.
- Hernes, T. 2014. *A Process Theory of Organization*. Oxford University Press.
- Hill, B. M. 2013. Essays on volunteer mobilization in peer production. Ph.D. thesis, Massachusetts Institute of Technology.
- Howison, J., K. Crowston. 2014. Collaboration through superposition: How the IT artifact as an object of collaboration affords technical interdependence without organizational interdependence. *MIS Quarterly* **38**(1) 29–50.
- Jain, A. K., M. N. Murty, P. J. Flynn. 1999. Data Clustering: A Review. *ACM Computing Surveys* **31**(3) 264–323.
- Kane, G., J. Johnson, A. Majchrzak. 2014. Emergent Life Cycle: The Tension Between Knowledge Change and Knowledge Retention in Open Online Coproduction Communities. *Management Science* **60**(12) 3026 – 3048.
- Katz, D., R. Kahn. 1978. *The Social Psychology of Organizations*. Wiley.
- Keegan, B., S. Lev, O. Arazy. 2016. Sequence Analysis Methods for Analyzing Organizational Routines in Online Knowledge Collaborations. *Proceedings of CSCW*. San Francisco, CA, USA, to appear.
- Kellogg, K. C., W. J. Orlikowski, J. Yates. 2006. Life in the trading zone: Structuring coordination across boundaries in postbureaucratic organizations. *Organization Science* **17**(1) 22–24.
- Kittur, A., E. Chi, B. Suh. 2009. What's in Wikipedia? Mapping Topics and Conflict Using Socially Annotated Category Structure. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 1509–1512.
- Kittur, A., R. E. Kraut. 2010. Beyond Wikipedia: Coordination and Conflict in Online Production Groups. *Proceedings of CSCW*. 215–224.
- Kittur, A., B. Suh, B. A. Pendleton, H. Chi. 2007. He Says, She Says: Conflict and Coordination in Wikipedia. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 453–462.
- Kriplean, T., I. Beschastnikh, D. W. McDonald. 2008. Articulations of Wikiwork: Uncovering Valued Work in Wikipedia Through Barnstars. *Proceedings of CSCW*. 47–56.
- Lakhani, K. R., H. Lifshitz-Assaf, M. Tushman. 2013. Open Innovation and Organizational Boundaries: Task Decomposition,

- Knowledge Distribution and the Locus of Innovation. *Handbook of Economic Organization: Integrating Economic and Organization Theory*. Edward Elgar Publishing, 355–382.
- Lakhani, K. R., J. A. Panetta. 2007. The Principles of Distributed Innovation. *Innovations* 2(3) 97–112.
- Lange, T., V. Roth, M. L. Braun, J. M. Buhmann. 2004. Stability-Based Validation of Clustering Solutions. *Neural computation* 16(6) 1299–1323.
- Langley, A., C. Smallman, H. Tsoukas, A. H. V. de Ven. 2013. Process Studies of Change in Organization and Management: Unveiling Temporality, Activity, and Flow. *Academy of Management Review* 38(1) 1–13.
- Lave, J., E. Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press.
- Leonardi, P. M., S. R. Barley. 2008. Materiality and change: Challenges to building better theory about technology and organizing. *Information and Organization* 18(3) 159–176.
- Levina, N., A. Fayard, E. Gkeredakis. 2014. Organizational Impacts of Crowdsourcing: What Happens with “Not Invented Here” Ideas? *Proceedings of the Collective Intelligence Conference*.
- Levina, N., M. Arriaga. 2014. Distinction and status production on user-generated content platforms: Using Bourdieu’s theory of cultural production to understand social dynamics in online fields. *Information Systems Research* 25(3) 468–488.
- Lifshitz-Assaf, H. 2016. Dismantling Knowledge Boundaries at NASA: From Problem Solvers to Solution Seekers. Available at SSRN: <http://ssrn.com/abstract=2431717>.
- Liu, J., S. Ram. 2011. Who Does What: Collaboration Patterns in the Wikipedia and Their Impact on Article Quality. *ACM Trans. Management Inf. Syst.* 2(2) 11.
- Majchrzak, A. 2009. Comment: Where is the theory in wikis? *MIS Quarterly* 33(1) 18–20.
- Majchrzak, A., M. L. Markus. 2012. Technology affordances and constraints in management information systems (MIS). E. Kessler, ed., *Encyclopedia of Management Theory*. Sage Publications.
- Majchrzak, A., C. Wagner, D. Yates. 2013. The impact of shaping on knowledge reuse for organizational improvement with Wikis. *MIS Quarterly* 37(2) 455–470.
- McKelvey, B. 2008. Emergent strategy via complexity leadership: Using complexity science and adaptive tension to build distributed intelligence. *Complexity Leadership Part I: Conceptual Foundations*. Information Age Publishing, 225–268.
- Mintzberg, H. 1992. *Structure in fives: Designing effective organizations*. Prentice Hall.
- Nan, N., Y. Lu. 2014. Harnessing the Power of Self-Organization in an Online Community During Organizational Crisis. *MIS Quarterly* 38(4) 1135–1157.
- Okhuysen, G. A., B. A. Bechky. 2009. Coordination in Organizations: An Integrative Perspective. *Academy of Management Annals* 3(1) 463–502.
- O’Mahony, S., K. Lakhani. 2011. Organizations in the Shadow of Communities. *Research in the Sociology of Organizations*. Emerald Group Publishing, 3–36.
- Oreg, S., O. Nov. 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.
- Orlikowski, W. J. 2000. Using Technology and Constituting Structures: A Practice Lens for Studying Technology in Organizations. *Organization Science* 11(4) 404 – 428.
- Orlikowski, W. J., S. V. Scott. 2008. 10 Sociomateriality: Challenging the Separation of Technology, Work and Organization. *The Academy of Management Annals* 2(1) 433–474.
- Orlikowski, W. J., J. Yates. 2002. It’s About Time: Temporal Structuring in Organizations. *Organization Science* 13(6) 684 – 700.
- Ortega, F., J. M. Gonzalez-Barahona, G. Robles. 2008. On the Inequality of Contributions to Wikipedia. *Proceedings of the Annual Hawaii International Conference on System Sciences*.
- Panciera, K., A. Halfaker, L. Terveen. 2009. Wikipedians are born, not made: A study of power editors on Wikipedia. *Proceedings of the ACM 2009 International Conference on Supporting Group Work*. 51–60.
- Platt, J. C. 1998. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines. Tech. rep., Microsoft Research.
- Preece, J., B. Shneiderman. 2009. The Reader-to-Leader Framework: Motivating Technology-Mediated Social Participation. *AIS Trans. on Human-Computer Interaction* 1(1) 13–32.
- Puranam, P., O. Alexy, M. Reitzig. 2014. What’s “New” About New Forms of Organizing? *Academy of Management Review* 39(2) 162–180.
- Puranam, P., D. D. Håkansson. 2015. Valve’s Way. *Journal of Organization Design* 4(2) 2–4.
- Quinlan, J. R. 1993. *C4.5: Programs for Machine Learning*. Morgan Kaufmann.
- Ransbotham, S., G. Kane. 2011. Membership Turnover and Collaboration Success in Online Communities: Explaining Rises and Falls from Grace in Wikipedia. *MIS Quarterly* 35(3) 613–627.
- Schreyögg, G., J. Sydow. 2010. CROSSROADS-Organizing for Fluidity? Dilemmas of New Organizational Forms. *Organization Science* 21(6) 1251–1262.
- Shah, S. K. 2006. Motivation, Governance, and the Viability of Hybrid Forms in Open Source Software Development. *Management Science* 52(7) 1000–1014.
- Shaw, A., B. M. Hill. 2014. Laboratories of Oligarchy? How the Iron Law Extends to Peer Production. *Journal of Communication* 64(2) 215–238.
- Stvilia, B., M. B. Twidale, L. C. Smith, L. Gasser. 2008. Information Quality Work Organization in Wikipedia. *Journal of the American Society for Information Science and Technology* 59(6) 983–1001.
- Treem, J. W., P. M. Leonardi. 2012. Social Media Use in Organizations: Exploring the Affordances of Visibility, Persistence, Editability, and Association. *Communication Yearbook* 36 143–189.
- Tsoukas, H., R. Chia. 2002. On Organizational Becoming: Rethinking Organizational Change. *Organization Science* 13(5) 567–582.
- Tuertscher, P., R. Garud, A. Kumaraswamy. 2014. Justification and interlaced knowledge at ATLAS, CERN. *Organization Science* 25(6) 1579–1608.
- Turner, J. H. 1986. *The Structure of Sociological Theory*. University of Chicago Press.
- Turner, R. H. 1978. The Role and the Person. *American Journal of Sociology* 84(1) 1–23.
- Van De Kamp, P. 2014. Holacracy - A Radical Approach to Organizational Design. *Elements of the Software Development Process - Influences on Project Success and Failure*. University of Amsterdam, 13–26.
- Von Krogh, G., E. Von Hippel. 2006. The Promise of Research on Open Source Software. *Management Science* 52(7) 975–983.
- Wagner, C. 2004. Wiki: A Technology for Conversational Knowledge Management and Group Collaboration. *Communications of the Association for Information Systems* 13(1) 265–289.
- Welser, H., D. Cosley, G. Kossinets. 2011. Finding Social Roles in Wikipedia. *Proceedings of the 2011 iConference*. 122–129.

- Williams, L. J., H. Abdi. 2010. Fisher's least significant difference (LSD) test. N. J. Salkind, ed., *Encyclopedia of Research Design*. Sage Publications.
- Yates, D., C. Wagner, A. Majchrzak. 2010. Factors Affecting Shapers of Organizational Wikis. *Journal of the Association for Information Science and Technology* **61**(3) 543–554.
- Yeo, M. L., O. Arazy. 2012. What Makes Corporate Wikis Work? Wiki Affordances and Their Suitability for Corporate Knowledge Work. K. Peffers, M. Rothenberger, B. Kuechler, eds., *Design Science Research in Information Systems. Advances in Theory and Practice*. Springer Berlin Heidelberg, Berlin, Heidelberg, 174–190.
- Zammuto, R. F., T. L. Griffith, A. Majchrzak, D. J. Dougherty, S. Faraj. 2007. Information Technology and the Changing Fabric of Organization. *Organization Science* **18**(5) 749–762.
- Zhang, H., S. Zhang, Z. Wu, L. Huang, Y. Ma. 2014. Predicting Wikipedia Editor's Editing Interest Based on Factor Graph Model. *IEEE International Congress on Big Data*. 382–389.
- Zhu, H., R. Kraut, A. Kittur. 2012. Organizing Without Formal Organization: Group Identification, Goal Setting and Social Modeling in Directing Online Production. *Proceedings of CSCW*. 935–944.
- Ziaie, P. 2015. A Model for Context in the Design of Open Production Communities. *ACM Computing Surveys* **47**(2) 29.