THE [WIKIPEDIA] WORLD IS NOT FLAT: ON THE ORGANIZATIONAL STRUCTURE OF ONLINE PRODUCTION COMMUNITIES

Complete Research

Arazy, Ofer, University of Alberta, Edmonton, Alberta, Canada, ofer.arazy@ualberta.ca

Nov, Oded, NYU Polytechnic, Brooklyn, NY, USA, onov@nyu.edu

Ortega, Felipe, University Rey Juan Carlos, Madrid, Spain, felipe.ortega@urjc.es

Abstract

The creation and maintenance of online production communities depend on the complex ecology created by the interaction of social roles, and these roles are essential for the governance of the community. This study investigates the organizational structure of one of the most notable peer-production projects: Wikipedia. While online communities have often been depicted as 'flat' and egalitarian, recent studies of Wikipedia suggest that it has developed a cumbersome beaurocratic structure that includes a hierarchy of organizational role. The objective of this study is, thus, to empirically study the organization of roles in Wikipedia and the hierarchy formed through their power relationships. Our research method employs Wikipedia's formal set of access privileges as indicators of roles, and analyses all 4,902,643 Wikipedia members (of which 10,496 hold special access privileges). Applying statistical techniques traditionally employed to validate the psychometric properties of scales, we find that Wikipedia has an intricate ecology of roles. Our analysis of power relationships within these twelve roles reveals Wikipedia's organizational hierarchy. Implications for theory and practice are discussed.

Keywords: *Peer Production, Wikipedia, Access Privileges, Social Roles, Organizational Structure, Hierarchy.*

1 Introduction

Recent years have seen the emergence of a new community-based model for the production of knowledge-based goods, for example open-source software such as the Linux operating system or question-and-answer communities such as StackExchange (http://stackexchange.com/). Wikipedia, the peer-produced online encyclopedia built on wiki technology, is often cited as the most notable example of this community-based model (Benkler 2006). Wikipedia was able to recruit thousands of volunteers who contribute their knowledge, develop extensive policies and mechanisms for governing the collaborative authoring process, and produce millions of encyclopedic entries across a large number of languages. Wikipedia has become a prodigious one-stop-shop for knowledge: its articles

are frequently at the top of search results and it is one of the world's most visited websites1. The success of Wikipedia has attracted the attention of organizational researchers and information systems scholars, who have been investigating Wikipedia's organizational model (Forte et al. 2009; Arazy et al. 2011; Ransbotham and Kane 2011; Oreg and Nov 2008).

Organizational structure – namely, the organization of roles and responsibilities – is a central issue in the governance of knowledge production on Wikipedia. Peer-production projects need to attract and retain volunteer participants with various interests and skill-sets, so that these contributors will share the responsibility for carrying out project tasks (including both production and administrative duties). As participants become more involved in the project and gain the community's trust, they gradually move from the periphery to the community core, gaining access to more sensitive and influential decisions (Amrit and van Hillegersberg 2010). As Wikipedia grew, it developed a set of roles, defined through their access privileges, rights and responsibilities. These roles were not prescribed or directed by Wikipedia's founders, rather they grew organically from the community.

The classification of participants into roles and the power relationships between roles are essential elements of Wikipedia's governance model. These roles play an important part in the central processes that underlie Wikipedia's operation. First, roles help organize the production process, such that more trusted users can perform more sensitive tasks. For example, unregistered participants can contribute content but cannot create new articles, and tasks such as splitting an article into two separate wiki pages can only be performed by participants with special roles. Second, roles are essential for the management of work activities, as certain roles can grant others with rights, giving participants the ability to perform additional tasks. Third, roles are important in establishing of Wikipedia's policies and norms, as certain decisions related to the enactment of procedures are only open to privileged roles. Finally, roles are essential in policing and enforcing the norms (Butler et al. 2008), such that some roles are particularly designed to: manage quality assurance tasks (Stvilia et al. 2008); fight vandals (Gleave et al. 2009); protect against biases and external influences (e.g. when firms are trying to influence the information concerning their products); and manage conflicts (Arazy et al. 2013; Arazy et al. 2011).

Despite the importance of roles in Wikipedia's governance model, to date there has been little research of the various roles in Wikipedia and their power relationships. Wikipedia's own description of the user access levels2 is quite confusing, including a large number of privileges (or "flags"), some of which relate to tools that are no longer in use (e.g. 'afttest', 'afttest-hide'), some privileges that are associated with external actors rather than active participants in Wikipedia's collaborative-authoring process (e.g. 'researcher', 'educator'), and some that are very similar in terms of their role in the community (e.g., 'checkuser' and 'oversight'). Furthermore, it is difficult to discern the ladder of power between roles from Wikipedia's own description of access levels. Prior research in the area does provide some description of Wikipedia roles, but these earlier studies have often been more interested in particular Wikipedia procedures, analyzing only those roles that are relevant for that procedure (Stvilia et al. 2008; Forte et al. 2009). This past research, therefore, provides an incomplete picture of participants' access levels and does not advance our understanding of how different privileges are organized to form Wikipedia's 'organizational chart'.

The objective of this study is, thus, to enhance our understanding of the organizational structure of mature online production communities, through the analysis of Wikipedia, one of the most prominent examples of peer production. The setting for our empirical study is the English version of Wikipedia, and we employ a dual method approach for determining roles and their organization by: (I) collecting

[1] Wikipedia is consistently ranks in the top ten sites according to Alexa (http://www.alexa.com/topsites), and considerably higher than any other reference site. The number of unique visitors in October 2013 was 530,470.
[2] http://en.wikipedia.org/wiki/Wikipedia:User_access_levels

and analyzing large-scale empirical data of all contributors to Wikipedia and (II) examining in details the various Wikipedia pages describing access privileges, their rights and responsibilities, and the procedure for granting these privileges. Our results delineate Wikipedia's organizational structure, categorizing the many access privileges into a smaller set of higher-order roles and describing the hierarchical relationships between these roles. Our findings inform theoretical frameworks of the organization of online production communities, and yield import insights for practitioners involved in managing online communities.

2 Related Work

2.1 Social Roles and Organizational Structure

The concept of social role has been the subject of extensive analysis in sociology for over a century (c.f. (Blau et al. 1995). The meaning of 'social role' is broad and varies between areas. Social roles encapsulate the social context, history of actions, structures of interaction, and the attributes people bring to the interaction by providing a meaning system, which both constrains and enables action (Merton 1968). The importance of this concept lies in its utility: the classification of types of social relations and behaviors into a smaller set of roles reduces the analytic complexity of social systems and facilitates the comparative study of populations across time and setting (Lerner 2005). Roles – and more generally, social life – could be understood through two primary dimensions: structure and culture. The structural definition of roles pertains to commonalities in behavior patterns; while the cultural dimension refers to roles that are recognized in social (or organizational) settings and differ in terms of their accessibility (the extent to which it is easy for one to accept the role) and situational contingency (the contextual factors that affect action). Roles are resources that help people accomplish their goals and are tools used in the establishment of social structure (Baker and Faulkner 1991), thus social (or organizational) structure could be viewed as an ecology of roles.

Formal organizations are generally understood to be systems of coordinated and controlled activities that arise when work is embedded in complex networks of technical relations and boundary-spanning exchanges (Meyer and Rowan 1977). Organizational structure affects organizational action in two primary ways. First, it provides the foundation on which standard operating procedures and routines rest. Second, it determines which individuals play which roles and who gets to participate in which decision-making processes. Thus the structure determines the extent to which individuals at different organizational positions are able to influence the organization's actions (Miles et al. 1978). The organizational literature describes various organizational structures, such as functional, divisional, and matrix structures. Recent technological advancements have given rise to new organizational forms such as network and virtual organizations (Amaral and Uzzi 2007). Although extant models of organizational roles and structures may provide a starting point for understanding organizational structure in online production communities, these models are not necessarily applicable to the peer-production context. In fact, growing evidence shows that the structure and processes of online production communities deviate from the predictions of traditional theories of group work (Arazy et al. 2011; Von Krogh and Von Hippel 2006), highlighting the need to develop refined theories of organizational roles and structure for this particular setting.

2.2 Roles and Organizational Structure in Online Production Communities

Online communities, and in particular production communities, have been investigated extensively in recent years. Relevant to our inquiry of roles and organizational structure within peer-production are prior works that have studied community members' roles. Previous investigations of online communities have discovered that users often follow very distinctive patterns of activity, playing roles in their online community (Welser et al. 2007). Production communities are often described in terms of a core-periphery structure, which entails a dense, cohesive core, and a sparse, unconnected periphery (Borgatti and Everett 2000). Contributors play different roles, with the majority of contributors, who are not very active, at the community's periphery, and a small minority who take on

additional responsibilities and privileges constituting the core (Long and Siau 2007). Most prior studies in the area focused on the quantity and types of activities performed by participants. For example, Kriplean et al. (2008) divided Wikipedia production activities into broad categories, including: editing work, social and community support actions, border patrol, administration, collaborative actions, etc; Arazy et al. (2010) have classified wiki editing tasks into several categories, and employed this classification to assess a user's overall contribution; Liu and Ram (2011) identified a similar set of actions in Wikipedia and defined roles based on participants' activity patterns (e.g. all-round editors, watchdogs, starters, content justifiers, copy editors, and cleaners); Arazy et al. (2011) categorized contributors into two primary classes: administrative (or community) vs. content-oriented; and Welser et al. (2011) applied a combination of interpretive and network analysis methods to identify four key roles: substantive experts, technical editors, vandal fighters, and social networkers. Much less work has been devoted to studying contributors' formal duties (Butler et al. 2007). A few prior works provide a description of Wikipedia's formal roles (Stvilia et al. 2008; Forte et al. 2009; Niederer and Van Dijck 2010), but these prior studies have often been primarily interested in a particular Wikipedia mechanism, analyzing only those roles that are relevant for that procedure. For example, Stvilia et al. (2008) provided a detailed review of quality assurance (QA) processes within Wikipedia and described three of the roles that are relevant for QA work: *registered users*, *anonymous users*, and *administrators*; Forte et al. (2009), in their analysis of Wikipedia governance mechanisms, described 6 access levels: *Administrator*, *Bureaucrat*, *Steward*, *Oversight*, *Checkuser*, and *Developer*; and Niederer and Van Dijck (2010) who investigated the role of software robots provide a partial description of nine access rights, which they organize in a hierarchy of privileges. While these prior studies are relevant for our investigation, their analysis does not span the entire spectrum of Wikipedia roles.

In recent years there have been several attempts to provide a conceptual framework of organizational roles within online communities arriving from both industry and academia. In industry, one noticeable framework is Forrester Research's 'Social Technographics Ladder' (Li et al. 2007) that proposes seven ordinal stages of participation: *inactives*, *spectators*, *joiners*, *collectors*, *critics*, *conversationalists*, and *creators*. In academia, Preece and Shneiderman (2009) provide an overall framework of participation in online communities by synthesizing prior works in the area. They organize activities in successive levels of involvement, termed the "Reader to Leader" (R2L) framework. In this funnel-like process new community participants join and begin their career as readers or "lurkers". Some of these readers then become more active and make minor contributions, and some of these active participants go further and take on leadership and facilitative roles. Interestingly, despite its influence on research of online communities, the R2L framework has not been examined empirically, and the extent to which organizational structure in online communities corresponds to this framework is not clear.

In sum, while organizational structure is a pivotal topic in the study of firms, to date there has been relatively little work on the structure of online production communities. Existing empirical studies of Wikipedia tend to focus on participants' transient activities as enacted roles at the moment, rather than on their formal role (Kane et al. in press). The few studies that do address issues related to formal roles and responsibilities, do not provide a comprehensive account of Wikipedia's organizational structure. Existing conceptualizations of roles in online communities are for the most part linear and focus on the axis between the periphery and the community's core, overlooking distinctions that may exist between different roles that occupy a similar position on the core-periphery axis. Furthermore, existing conceptualizations tend to view online production communities as egalitarian, ignoring the power relations that may exist between roles and the organizational hierarchy that emerges from these power relations. For example, studies of open source software describe the progression between roles as a lateral movement (Dahlander and O'Mahony 2011; O'Mahony and Ferraro 2007).

3 Research Method and Results

In order to address the gaps in the literature and advance our understanding of the organizational structure of online communities, we performed an empirical study of Wikipedia, investigating how different privileges are organized to form Wikipedia's 'organizational chart'. The focus of this empirical investigation is the English Wikipedia and the community of editors contributing to it. We base our data gathering and analysis on (I) Wikipedia's system logs, as harvesting these logs can reveal important insights about members' ongoing behavior in its natural setting, and (II) a detailed analysis of Wikipedia's internal procedures, and in particular the wiki pages devoted to the various roles and access privileges[3]. Our method includes two stages: first, we analyze the Wikipedia user access levels and group them into higher-level roles; second, we organize these roles according to their power relations to form Wikipedia's 'organizational chart'.

3.1 Delineating Wikipedia Roles

Our analysis of roles is based on Wikipedia's definitions of access privileges. Over time, the Wikipedia community has developed a comprehensive set of procedures for governing the collaborative editing process, including a well-defined scheme of access privileges (Butler et al. 2008). Each contributor may hold multiple access privileges. Wikipedia's formalized access privileges enable us to examine both the cultural and structural dimensions in the definition of social roles. First, social roles are cultural objects that are recognized and accepted by the community (Callero 1994); Wikipedia's access privileges were developed by the community through deliberation, are presented on the community's website, and are formal aspect of the Wikipedia institution. Second, people who consistently adopt particular roles develop distinctive modes of participating in social settings, which result in patterns of behavior and relations, thus social roles constrain and enable behavior. Wikipedia's access privileges – by their definition – constrain and enable action; thus, users that hold similar access privileges are bound to exhibit some commonalities in behavior patterns (especially those users that hold the special access permissions).

The first step of our study required that we review the various access privileges and screen for only those privileges that are relevant for our analysis, i.e. human actors that are active contributors. As part of this screening we excluded: software bots (Niederer and Van Dijck 2010; Geiger 2013) access levels that are no longer used (e.g. related to the now inactive User Feedback feature: 'afttest' and 'afttest-hide'); and special access privileges given to external parties (e.g. researchers or educators). Table 1 below lists the set of nineteen access privileges included in our analysis. Note that while some access privileges are local to the English Wikipedia, some are defined at a global level.

While access privileges portray a detailed picture of the permissions to perform tasks, our interest is in higher-level classification of organizational roles, where several privileges may be associated with a single role. The next step was therefore to (a) group access permissions into a smaller set of classes, each class corresponding to an organizational role; and (b) identify the power relationships between these various roles, thus delineating Wikipedia's hierarchical organizational structure.

[3] http://en.wikipedia.org/wiki/Wikipedia:User_access_levels and the links to then pages describing in detail each privilege.

Access level (aka)	**L / G**	**Description**
* (Unregistered User)	L	Contributors who have not created an account are identified by their IP address. Unregistered users may read all standard Wikipedia pages and edit pages that are not protected (with few restrictions) and they may create most types of talk pages.
user (New User)	L	Users that can immediately create pages in almost any namespace (with fewer restrictions on editing activities). They are also granted very limited clerical access privileges (e.g. marking others' edits as 'small') and may interact with other users.
autoconfirmed	L	Users that are given access to actions otherwise restricted (move pages, edit semi-protected pages, upload files, or upload a new version of an existing file).
confirmed	L	Accounts that require some intervention in order to be approved and have the same rights as the 'autoconfirmed' pseudo-group.
IPpblock-exempt	L	Users that were inappropriately affected by a block intended to prevent vandalism or disruption and were manually exempt from the block.
filemover	L	Users that are experienced in working with files and are allowed to rename them, with some restrictions (cannot move a file if the same file exists on Commons).
accountcreator	L	Users who are not affected by account creation limits (e.g. 6 daily account creation per IP address), and can create accounts for others.
rollback	L	Users who may revert revisions using the rollback feature.
reviewer	L	Users who are able to review other users' edits to articles placed under 'pending changes protection'. The reviewing process is intended to ensure edits don't contain vandalism, policy violations or other obviously inappropriate content.
autoreviewer (AutoPatrolled)	L	User that are highly active and who could be trusted not to submit inappropriate material, such that it is more efficient to mark their edits as approved preemptively. They can also have their pages automatically patrolled on the New Pages List.
abusefilter (EditFilterManager)	L	Users who can create, modify, and delete edit filters (i.e. set specific controls on user activity and create automated reactions for certain behaviors).
sysop (Administrator)	L	Users have access to a number of tools to allow them to carry out additional functions (e.g. page deletion, page protection, blocking and unblocking, modify protected pages).
bureaucrat	L	Users that are allowed to perform certain actions on other users' accounts (e.g. rename users). They can add other users to the 'sysop' and 'bureaucrat' groups (but cannot remove them) and can add/remove from the 'bot' group.
oversight	L	Users who can hide revisions of pages from all users, and can see hidden revisions. This expunges information from any form of usual access (even by administrators) and is used within strict limits to remove defamatory material, protect privacy, and remove serious copyright violations.
checkuser	L	Users that can view the IP addresses used by other users. This is used to establish whether two or more accounts are being operated by one individual or group of people, and then to protect Wikipedia against disruptive or abusive behavior.
steward	G	Stewards are appointed globally across all public Wikimedia wikis. They may grant and revoke any permission to or from any user or group, and generally act only when there is no user on a particular wiki that can make the necessary change.
importer & transwiki	G	Highly-trusted users that are permitted to move pages between Wikimedia wikis (for instance, to copy an article written in English on another-language wiki, while preserving full edit history).
ombudsman	G	Ombudsmen investigate complaints about violations of privacy policy on any Wikimedia project. They have additional (global) rights, such as: search deleted pages, check user's IP addresses (checkuser), and view deleted entries.
founder	G	Class reserved for Jimmy Wales. Founder is responsible for defining high-level policies and norms and for overall direction of the community

Table 1. Wikipedia's access privileges (L=English Wikipedia; G=the entire Wikipedia).

In line with the definition of roles as representing commonalities in action, our method for grouping access privileges was based on Wikipedians' actual practices, assuming that if Wikipedians tend to hold a set of privileges together, then these privileges are likely to represent a single higher-order role. We first extracted from Wikipedia's API the list of all users and the access privileges associated with the English Wikipedia, as well as meta-wiki level privileges (e.g. ombudsman, import). Our analysis was based on data extracted on April, 2012. It included 4,902,643 Wikipedia members, 10,496 of which had special privileges. We performed exploratory factor analysis (EFA) of the special privileges[4], making no prior assumptions on the relationships between privileges. Anonymous and new users (i.e. 'user' and 'autoconfirmed') were excluded from this analysis: for those starting as anonymous users, it is not possible to track additional privileges that were granted later, since there is no way to associate the IP address used originally by the anonymous users with the user name they later registered. New users were excluded from the factor analysis due to the following reasons: (a) to make the analysis more tractable (scaling down from close to five million to just over ten thousand users) and (b) new users represent a role class that is distinct from the classes associated with higher privileges. Table 2 presents results of the factor analysis.

Access Privileges	**# cases**	**Factors**						
		A	**B**	**C**	**D**	**E**	**F**	**G**
confirmed	260	**.73**						
ipblockexempt	269	**.51**	.34					
filemover	258		**.68**					
accountcreator	96		**.70**					
rollbacker	4,511	.42	.27	-.55		-.54		
reviewer	5,514	-.28		**.64**		-.36		
autoreviewer	2,682			**.75**				
checkuser	42				**.86**			
oversight	37				**.87**			
abusefilter	142					**.58**		
sysop	1,479					**.82**		
bureaucrat	34					.31		
import_transwiki	3						**.83**	
steward	37						**.63**	.37
ombudsman	6							**.90**
founder	1				.46			

Table 2. Results of exploratory Factor analysis after Varimax rotation; for clarity, values below 0.25 suppressed; values above 0.5 in bold; circled represent discrepancies.

[4] For the factor analysis, we used SPSS version 21 (www.ibm.com/software/analytics/spss/)

The factor analysis grouped the special privileges onto seven high-level factors, together explaining 58% of the variance in the data. While the EFA provided a useful starting point for grouping access privileges into role categories and in most cases were privileges that represent a closely-related role were grouped together (e.g. 'checkuser' and 'oversight'), there were some borderline cases where privileges did not map clearly onto a factor. Namely, three privileges did not load on any factor above the 0.5 threshold: 'rollbacker' (max loading 0.27), 'bureaucrat' (0.31) and 'founder' (0.46). An additional access privilege – 'abusefilter' – loaded with 0.58 on Factor E together with 'sysop', but the two access privileges do not seem to represent the same role (correlation between the two privileges was 0.26). In order to resolve these discrepancies and determine the groupings of privileges onto higher-order roles, we analyzed in detail Wikipedia's descriptions of the various privileges, considering the rights and responsibilities associated with each user privilege. In doing so, we combine the structural approach that was based on quantitative analysis of participants' use of access privileges with an interpretative approach that captures the behavioral notion of a role. This hybrid approach to role definition makes it possible to identify meaningful roles in online communities (Gleave et al. 2009). Through the interpretive process, we resolved some of the discrepancies in the EFA. Namely: 'bureaucrat' was associated with Factor E (together with 'sysop'), and new role categories were created for the 'rollabacker', 'abusefilter' and 'founder' privileges. Table 3 presents the results of this analysis, grouping access privileges into role classes and assigning a name to each role class.

Role	**Description**	**Access Privileges**
Unregistered Users	Non community members	*
New Registered Users	Newly registered users	user
		autoonfirmed
Manually Registered Users	New users who had to be manually registered to bypassed some restrictions	confirmed
		IPpblock-exempt
Technical Administration	Privileged users responsible for the administration of the technical aspects (e.g. user accounts, files)	filemover
		accountcreator
Border Patrol	Users responsible for fighting vandalism by reverting malicious edits	rollback
Quality Assurance	Privileged users responsible patrolling Wikipedia and for ensuring content quality	reviewer
		autoreviewer
QA Technicians	Users who develop automated tools (i.e. edit filters) to assist quality assurance work	abusefilter
Administrators	Highly involved users that are responsible for the social administration of the English Wikipedia community	sysop
		bureaucrat
Security Force	Highly trusted users who are working to keep malicious users out and combat intentional manipulations of content	oversight
		checkuser
Directors	Key users responsible for oversight of the Wikimedia organization	steward
		importer & transwiki
Privacy Commissioner	High-ranking users who investigate complaints about violations of privacy policy	ombudsman
Benevolent Dictator	Jimmy Wales; responsible for defining high-level policies and norms and for overall direction of the community	founder

Table 3. The organization of access privileges into higher-order role classes.

In order to ensure that that our revised grouping of role classes is a natural one, we wanted to verify that (I) each class is characterized by high internal consistency (i.e. access privileges associated with the same role are correlated) and (II) the different groups represent distinct classes (i.e. different role classes are not correlated with one another). We applied several techniques which are commonly used in behavioral studies, where each access privilege is viewed as a measurement item of a scale and each role class is viewed as a construct (see Table 4 below for details). In these analyses, we employed Partial Least Squares (PLS) (Abdi 2003) with the software tool SmartPLS (). To assess internal consistency, we analyzed the composite reliability of the multi-item role categories (Directors, Executives, Administrators, QA, Tech. Administration and Manual Registered Users), and found the values to be above the 0.5 threshold (0.65, 0.78, 0.56, 0.74, 0.71, and 0.66, respectively), demonstrating reasonable internal consistency. Convergent validity for all role 'constructs' was also good. The average variance extracted (AVE) for the multi-item role classes were above the commonly used 0.5 threshold (0.53, 0.59, 0.51, 0.59, 0.56, and 0.50, respectively). Discriminant validity was examined by comparing the square root of the AVE (RAVE) of a particular role class 'construct' (presented in Table 4 on the diagonal) and the correlation between that 'construct' and other role classes 'constructs' (Fornell and Larcker 1987) (presented in the off-diagonal positions of Table 4). The constructs' RAVEs are 0.70 or higher (0.73, 0.77, 0.71, 0.77, 0.75 and 0.70, respectively), and the RAVE for every role class is higher than the correlation between that 'construct' and all other role classes. In addition, the correlations between latent constructs did not exceed the recommended threshold of 0.5. Thus, our results demonstrate good discriminant validity. It should be noted that we applied here the strict threshold used to evaluate the psychometric properties of scales used in behavioral studies (Fornell and Larcker 1981; Fornell and Larcker 1987), although in our case where we are only seeking to evaluate the naturalness of role classes groupings there should not be as rigorous restrictions. Still, our results pass the threshold used in behavioral studies to assess constructs' convergent and discriminant validity.

Role Class	Composite Reliability	AVE	*Benevolent Dictator*	*Directors*	*Privacy Commission*	*Executives*	*Admins*	*QA Technicians*	*QA*	*Border Patrol*	*Tech. Admin*	*Man. Reg. Users*
Benev. Dictator	1.00	1.00	**1.00**									
Directors	0.65	0.53	0.00	**0.73**								
Privacy Comm.	1.00	1.00	0.00	0.12	**1.00**							
Executives	0.93	0.87	0.17	0.06	0.04	**0.93**						
Admins	0.51	0.56	0.02	0.03	0.00	0.16	**0.71**					
QA Technician	1.00	1.00	0.00	0.06	0.00	0.21	0.26	**1.00**				
QA	0.74	0.59	-0.01	-0.05	-0.03	-0.07	-0.45	-0.12	**0.77**			
Border Patrol	1.00	1.00	-0.01	-0.02	-0.02	-0.06	-0.35	-0.09	-0.13	**1.00**		
Tech. Admin	0.71	0.56	0.00	-0.01	0.00	-0.01	-0.07	0.01	0.10	0.13	**0.75**	
Man. Reg. User	0.66	0.50	0.00	-0.01	-0.01	-0.02	-0.09	-0.02	-0.18	-0.13	0.02	**0.70**

Table 4. Composite Reliability, AVE, RAVE (on the Diagonal) and inter-construct correlations.

3.2 Determining Wikipedia's Organizational Structure

After confirming that our grouping of access privileges into role classes exhibits convergent and discriminant validity, we analyzed the superior-subordinate relationships between role classes to determine Wikipedia's organizational chart. Roles are tools employed to establish social structure and defined power relationships (Baker and Faulkner 1991). For example, in Wikipedia a participant may rely on the role of an administrator in order to achieve social order and to sanction 'newbies' for failure to follow local norms. Examining social role ecologies, defined by the balance and relationships between roles, allows us to further our overall understanding of the organizational structure on online communities (Gleave et al. 2009). Our analysis is based primarily on the relationships between access privileges, as they are described on Wikipedia's pages. These descriptions of users' privileges suggest there are various types of superior-subordinate relationships.

Superior	Subordinate	Relation	Details
Benevolent Dictator	*Directors, Privacy Com.*	Access Scope	Benevolent Dictator can choose to receive subordinate privileges (as well as all user access privileges). Benevolent Dictator has the exclusive right to make some higher-level decisions on community norms and policies
Directors	*Security Force*	Granting Access	Stewards are responsible for granting and revoking oversight and checkuser access levels (no other group can make these changes). Stewards can also act as checkusers or oversighters, on wikis which do not have active local members of those groups
Privacy Commissioners	*Security Force*	Scope Granting	The ombudsman acts as a mediator on complaints pertaining to checkusers and others. The ombudsman is also in charge of educating checkusers or others about the Foundation's privacy policy When the privacy policy has been breached, the ombudsman may recommend to the Board the removal of checkusers access privileges
Directors	*Administrators*	Granting	Steward can grant (and remove) sysop or bureaucrat access levels on wikis which do not have any local administrators. Stewards can also act as administrators on wikis which do not have active local administrators
Security Force	*Administrators*	Promotion Scope	Checkusers and oversight require more stringent application and approval process (including the requirement to prove identity). While the application process does not formally require being an administrator, in practice nearly all checkuser and oversight members were previously administrators Oversight members may select whether the RevisionDelete extension will be used as a suppression action that prevents administrator access, or as an administrator action that any administrator can see and modify (administrators only have access to the latter)
Administrators	*QA Tech., QA, Border Patrol, Tech. Admin.*	Access Granting Scope	Administrators are given access to these subordinate privileges. Superiors can grant and remove these subordinate privileges. Administrators have rights not available to subordinate roles (e.g. administrators can move a file if the same file name exists on Commons, but filemovers can't)
QA Tech., QA, Border Patrol, Tech. Admin.	*Man. Reg. Users, New Reg. Users*	Scope Promotion	Members of the superior classes have all of the privileges of the subordinate classes and more. To be part of the superior classes, one needs to be a registered user, and apply for additional privileges
Man. & New Reg. Users	*Unregistered Users*	Scope	Superiors have all of the privileges of the subordinate privileges and more

Table 5. Superior-subordinate relationships between role classes.

First, a particular privilege can contain all of the rights of another access privilege and more, indicating a higher rank of the former privilege (e.g. 'filemovers' have the rights of 'users' and in addition can rename files); we refer to this relation type as "*Scope*". Second, a user who receives certain access privilege can automatically (or semi-automatically) be granted an additional privilege, also indicating higher position of the former privilege in the organizational ladder (e.g. 'sysop' is automatically granted other privileges such as 'rollback'); we term this relation "*Access*". Third, certain privileges have the rights of adding/removing other privileges, suggesting that the former privileges hold more organizational power (e.g. 'stewards' can add or delete access privileges such as 'oversight' or 'checkuser') ; we refer to this relationship type "*Granting*". Fourth, promotion to a higher access privilege may require as a prerequisite a lower access privilege (e.g. application to the 'checkuser' position is in practice open only to administrators, and also requires that the applicant provides a proof of identity); we term this relationship "*Promotion*". Table 5 describes the superior-subordinate relations between role classes, and Figure 1 depicts the hierarchy of user groups resulting from these power relationships, as well as the number of participants performing the various roles.

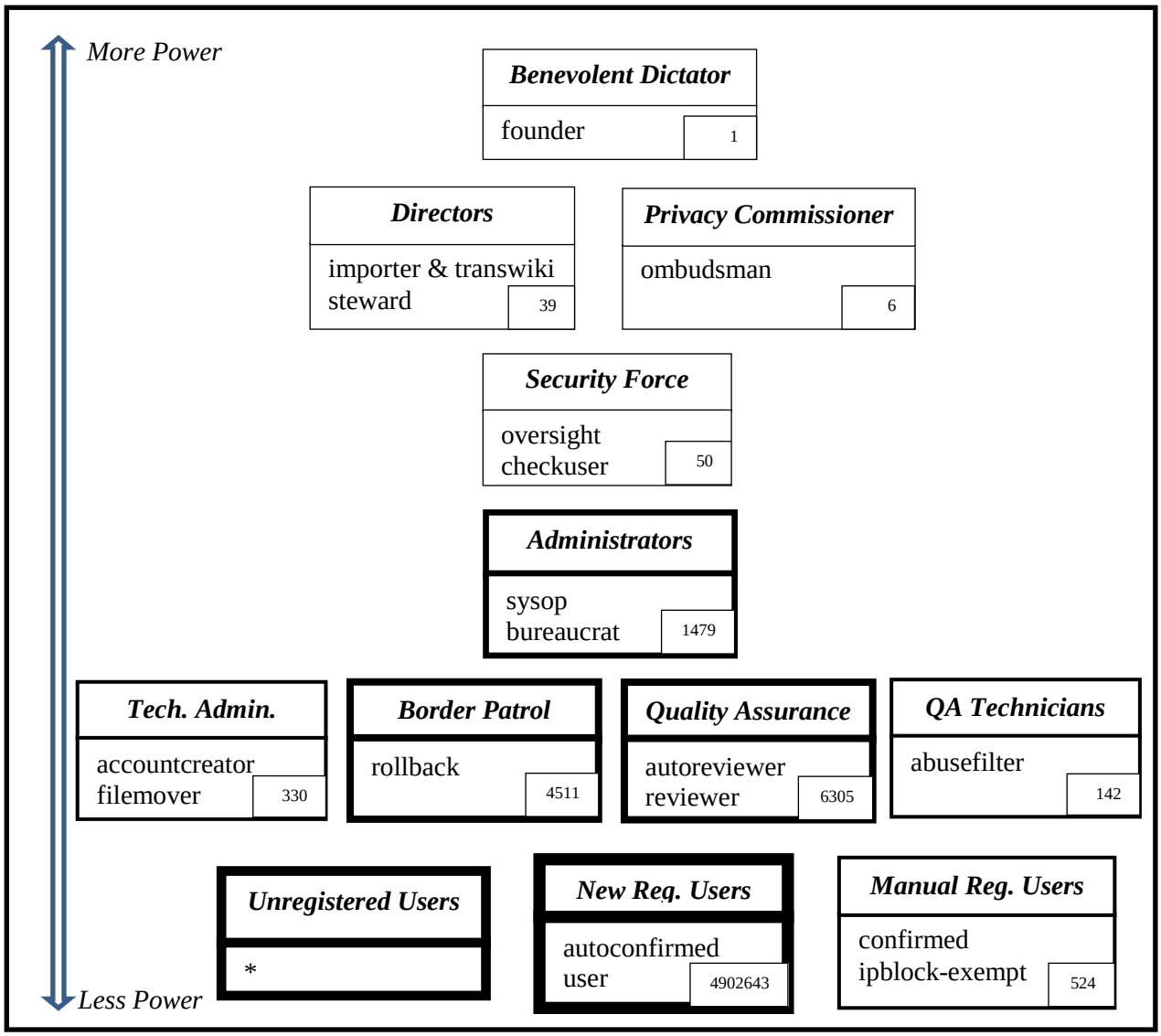

Figure 1. Wikipedia's organizational chart. Thickness of borders corresponds to the number of participants performing the role. Note that it is not possible to determine the number of unregistered participants, since an IP address is cannot be linked to a single user.

4 Discussion and Conclusion

Peer production initiatives represent a novel organizational model that is playing an increasing role in today's knowledge-based economy. Recent studies have demonstrated that work processes within online production communities deviate from the predictions of traditional organizational theory, highlighting the need to advance our understanding of the community-based model. The creation and maintenance of online production communities depend on the complex ecology created by the interaction of social roles. Roles within online communities are essential for quality assurance, coordination, and conflict resolution processes. An understanding of these social roles and clearly defining them allows us to simplify complex systems, compare between different social settings, and thus is a critical step in the analysis of work processes within online communities (Gleave et al. 2009).

Online production communities, and Wikipedia in particular, are based on egalitarian principles. Their organizational structure has been likened to the layers of an onion, the outside representing occasional contributors at the periphery and the internal layers as the active community core. Online communities often reject traditional notions of power and expertise and describe transitions from the periphery to the core in terms of increasing responsibilities, rather than power. That it, insiders are viewed as facilitators of the community, rather than managers.

However, such online communities – despite the public reputation for openness and democratic practices - do not avoid the formation of hierarchy. (Dean 2001) critiqued the popular notion that ''the infotainment society'' realizes democratic ideas by erasing political hierarchy, and argued that online communities are dominated by tech-savvy individuals forming a ''technocracy''. Recent accounts of Wikipedia describe the beaurocratic structure that has arisen in recent years (Butler et al. 2008). For example, Gleave et al. (2009) refers to the community's core as "positions of power and authority" and the emerging organization as a "hierarchy of administrators and arbiters"; Hartelius (2010) describes Wikipedia as a ''technocratic hierarchy" and the peer-production process as "contentious struggle over authority" (p. 507); and Niederer and Van Dijck (2010) write that "Wikipedia has a refined hierarchical structure in which contributing administrators, registered users, anonymous users and 'bots' all have a distinct rank in an orderly system (p. 1369) and refer to Wikipedia as "disciplinary system of power distribution" (p. 1373); they call the core contributors "elite" and describe Wikipedia roles in terms of their power and authority. The argument is that such a hierarchical structure is essential for coordinating the massive task of Wikipedia work (Burke and Kraut 2008) and that "without the implementation of this strict hierarchical content management system, Wikipedia would most likely have become a chaotic experiment" (Niederer and Van Dijck 2010) (p. 1372).

The primary contribution of our study is in delineating Wikipedia's organizational structure. We performed a large-scale analysis of all of Wikipedia contributors to find that Wikipedian's could be classified into twelve role classes. The grouping of access privileges into these higher-order role classes allows us to tie the results of our study to more general constructs that transcend the particular settings of our study. Furthermore, our analysis demonstrated how these role classes are organized in power relationships, forming Wikipedia's hierarchy.

Most of the prior accounts of roles in online communities have used interpretive methods for role discovery. Although these studies have identified several important roles in online groups, there are additional roles that cannot be identified with a strictly interpretive approach. Alternative approaches for role identification that have been employed to study Wikipedia include the analysis of: social network (Gleave et al. 2009; Welser et al. 2011); editing tasks (Liu and Ram 2011); and personal awards (Kriplean et al. 2008). Our approach has been based on the analysis of Wikipedia access privileges and has the advantages of: (I) working with roles that have a clear meaning and are recognizable in the community (as opposed to the network- and task-based approaches that produce patterns that are difficult to interpret); (II) lending itself to the analysis of power relationships (difficult

with the other approaches); and (III) providing a comprehensive picture that includes all community members (as opposed to approaches that are based on awards that are granted to only few members).

Our findings have important implications for both theory and practice. In terms of theory, our findings describe the organizational structure that characterizes a mature and large community such as Wikipedia, which is substantially more elaborate than extant one-dimensional models, e.g. core-periphery model (Amrit and van Hillegersberg 2010) or the 'reader-to-leader' framework (Preece and Shneiderman 2009). In addition, to the best of our knowledge our work provides the first empirical account for the hierarchical structure of Wikipedia, contributing to organizational theory on community governance. Findings from our study can also inform practitioners, as the mapping of participants to one or more role categories can help designers and managers of computer-mediated social spaces to focus on meta-level management tasks, such as monitoring the relative proportions of roles (Gleave et al. 2009). Particularly relevant for the Wikipedia community, our findings can help contributors who often get lost in the beaurocratic maze to understand the complex social structure of their community. Furthermore, our results may force Wikipedians to come to terms with the discrepancy between their ideal community structure and the hierarchy that emerged in practice. An understanding of Wikipedia's structure is of particular importance, given the recent suggestions that Wikipedia is in decline as a result of its cumbersome role ecology (Halfaker et al. 2013).

In conclusion, our study delineates organizational roles within Wikipedia and provides an account of then power relationships between these roles. As a preliminary investigation of Wikipedia's organizational structure, our work has several limitations. First, while access privileges provide a useful identifier for roles, other types of data (e.g. details on users' personal pages, personal awards, activity patterns) could be employed to provide a fuller representation of roles within Wikipedia. Second, our study was performed within the context of one community and further research is warranted in order to generalize our findings. We plan to address these limitations in our future research and intend to develop a refined framework of power relationships in online communities.

Acknowledgements

The authors thank Adam Balila for his work on data extraction and analysis. This research was supported in part by the Canadian Social Science and Humanities Research Council (SSHRC; Ref. IG 435-2013-0624), as well as by the Corporate & Community project funded by the Spanish Ministry of Economy and Competitiveness (MINNECO INNPACTO; Ref. IPT-2011-0871-430000).

5 References

Abdi, H. (2003). Partial Least Squares Regression. In Encyclopedia for Research Methods for the Social Sciences, Lewis-Beck, Bryman, Futing (eds), Thousand Oaks, Sage CA:7109-7795

Amaral, L.A.N., Uzzi, B. (2007). Complex systems - a new paradigm for the integrative study of management, physical, and technological systems. Management Science, 53 (7):1033-1035

Amrit, C., van Hillegersberg, J. (2010). Exploring the impact of socio-technical core-periphery structures in open source software development. Journal of Information Technology, 25 (2):216-229

Arazy, O., Nov, O., Patterson, R., Yeo, L. (2011). Information quality in Wikipedia: The effects of group composition and task conflict. Journal of Management Information Systems (JMIS), 27 (4):71 - 98

Arazy, O., Stroulia, E., Ruecker, S., Arias, C., Fiorentino, C., Ganev, V., Yau, T. (2010). Recognizing contributions in wikis: Authorship categories, algorithms, and visualizations. Journal of the American Society for Information Science and Technology (JASIST), 61 (6):1166-1179

Arazy, O., Yeo, L., Nov, O. (2013). Stay on the Wikipedia Task: when task-related disagreements slip into personal and procedural conflicts. Journal of the American Society for Information Science and Technology (JASIST), 64 (8):1634–1648

Baker, W.E., Faulkner, R.R. (1991). Role as resource in the Hollywood film industry. American Journal of Sociology 97 (2):279-309

Benkler, Y. (2006). The wealth of networks: How social production transforms markets and freedom. Yale Univ Press,

Blau, J.R., Coser, R.L., Goodman, N. (1995). Social Roles & Social Institutions. Transaction Publishers, New Jersey, USA

Borgatti, S.P., Everett, M.G. (2000). Models of core/periphery structures. Social Networks, 21 (4):375-395

Burke, M., Kraut, R. Mopping up: modeling wikipedia promotion decisions. In: Proceedings of the 2008 ACM conference on Computer supported cooperative work (CSCW), San Diego, California, USA, November 8–12, 2008 2008. ACM, pp 27-36

Butler, B., Joyce, E., Pike, J. Don't look now, but we've created a bureaucracy: the nature and roles of policies and rules in wikipedia. In: Proceedings of the CHI conference on human factors in computing systems, Florence, Italy, April 5-10, 2008 2008. ACM, pp 1101-1110

Butler, B., Sproull, S., Kiesler, S., Kraut, R.E. (2007). Community effort in online communities: who does the work and why. In: Weisband S (ed) Leadership at a Distance Research in TechnologicallySupported Work. Lawrence Erlbaum Associates, London UK, pp 171-194

Callero, P.L. (1994). From role-playing to role-using: Understanding role as resource. Social Psychology Quarterly, 57 (3):228-243

Dahlander, L., O'Mahony, S. (2011). Progressing to the center: Coordinating project work. Organization Science, 22 (4):961-979

Dean, J. (2001). From Technocracy to technoculture. Theory & Event 5 (1)

Fornell, C., Larcker, D. (1981) Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research, 18 (1):39-50

Fornell, C., Larcker, D. (1987). A second generation of multivariate analysis: Classification of methods and implications for marketing research. Review of marketing 51:407-450

Forte, A., Larco, V., Bruckman, A. (2009). Decentralization in Wikipedia governance. Journal of Management Information Systems (JMIS), 26 (1):49-72

Geiger, R.S. Are computers merely supporting cooperative work: towards an ethnography of bot development. In: Proceedings of the 16th conference on computer supported cooperative work (CSCW'2013), San Antonio, Texas, USA, February 23-27, 2013 2013. ACM, pp 51-56

Gleave, E., Welser, H.T., Lento, T.M., Smith, M.A. A conceptual and operational definition of'social role'in online community. In: 42nd Hawaii International Conference on System Sciences (HICSS'09), 2009. IEEE, pp 1-11

Halfaker, A., Keyes, O., Taraborelli, D. Making peripheral participation legitimate: reader engagement experiments in wikipedia. In: Proceedings of the 16th conference on computer supported cooperative work (CSCW'2013), San Antonio, Texas, USA, February 23-27, 2013 2013. ACM, pp 849-860

Hartelius, E.J. (2010) Wikipedia and the emergence of dialogic expertise. Southern Communication Journal, 75 (5):505-526

Kane, G.C., Johnson, J., Majchrzak, A. (in press). Emergent Lifecycle: The Tension Between Knowledge Change And Knowledge Retention In Open Online Co-production Communities. Management Science.

Kriplean, T., Beschastnikh, I., McDonald, D.W. Articulations of wikiwork: uncovering valued work in wikipedia through barnstars. In: Proceedings of the 2008 ACM conference on Computer supported cooperative work, San Diego, CA, USA, 2008. ACM, 1460573, pp 47-56. doi:http://doi.acm.org/10.1145/1460563.1460573

Lerner, J. (2005). Role assignments. In: Brandes U, Erlebach T (eds) Network analysis. Springer, Berlin, Germany, pp 216-252

Li, C., Bernoff, J., Fiorentino, R., Glass, S. (2007). Social Technographics.

Liu, J., Ram, S. (2011). Who does what: Collaboration patterns in the Wikipedia and their impact on article quality. ACM Transactions on Management Information Systems, 2 (2):1-23. doi:10.1145/1985347.1985352

Long, Y., Siau, K. (2007). Social network structures in open source software development teams. Journal of Database Management, 18 (2):25-40

Merton, R.K. (1968). Social theory and social structure. Simon and Schuster, New York, NY, USA

Meyer, J.W., Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. American Journal of Sociology, 83 (2):340-363

Miles, R.E., Snow, C.C., Meyer, A.D., Coleman, H.J. (1978). Organizational strategy, structure, and process. Academy of Management Review, 3 (3):546-562

Niederer, S., Van Dijck, J. (2010). Wisdom of the crowd or technicity of content? Wikipedia as a sociotechnical system. New Media & Society, 12 (8):1368-1387

O'Mahony, S., Ferraro, F. (2007). The emergence of governance in an open source community. Academy of Management Journal, 50 (5):1079-1106

Oreg, S., Nov, O. (2008). Exploring motivations for contributing to open source initiatives: The roles of contribution context and personal values. Computers in Human Behavior, 24 (5):2055-2073

Preece, J., Shneiderman, B. (2009). The Reader-to-Leader Framework: Motivating technology-mediated social participation. AIS Transactions on Human-Computer Interaction, 1 (1):13-32

Ransbotham, S., Kane, G.C. (2011). Membership Turnover and Collaboration Success in Online Communities: Explaining Rises and Falls from Grace in Wikipedia. MIS Quarterly, 35 (3):613-627

Stvilia, B., Twidale, M., Smith, L., Gasser, L. (2008). Information quality work organization in Wikipedia. Journal of the American Society for Information Science and Technology (JASIST), 59 (6):983-1001

Von Krogh, G., Von Hippel, E. (2006). The Promise of Research on Open Source Software. Management Science, 52 (7):975-983

Welser, H.T., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gay, G., Smith, M. Finding social roles in Wikipedia. In: Proceedings of the 2011 iConference, Seattle, WA, USA, February 8-11, 2011 2011. ACM, pp 122-129

Welser, H.T., Gleave, E., Fisher, D., Smith, M. (2007). Visualizing the signatures of social roles in online discussion groups. Journal of Social Structure, 8 (2):1-32