Summer 10-6-2011

GOVERNING SOCIAL PRODUCTION IN THE INTERNET: THE CASE OF WIKIPEDIA

Aleksi Aaltonen

Giovan Francesco Lanzara

Follow this and additional works at: http://aisel.aisnet.org/ecis2011

GOVERNING SOCIAL PRODUCTION IN THE INTERNET: THE CASE OF WIKIPEDIA

Aaltonen, Aleksi, The London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK, v.a.aaltonen@lse.ac.uk

Lanzara, Giovan Francesco, University of Bologna, Via Zamboni, 33 - 40126 Bologna - Partita IVA, Italy, giovan.lanzara@unibo.it

Abstract

In this paper we propose a theoretical framework to understand the governance of internet-mediated social production. Focusing on one of the most popular websites and reference tools, Wikipedia, we undertake an exploratory theoretical analysis to clarify the structure and mechanisms driving the endogenous change of a large-scale social production system. We argue that the popular transactions costs approach underpinning many of the analyses is an insufficient framework for unpacking the evolutionary character of governance. The evolution of Wikipedia and its shifting modes of governance can be better framed as a process of building a collective capability, namely the capability of editing and managing a new kind of encyclopedia. We understand Wikipedia evolution as a learning phenomenon that gives over time rise to governance mechanisms and structures as endogenous responses to the problems and conditions that the ongoing development of Wikipedia itself has produced over the years. Finally, we put forward five empirical hypotheses to test the theoretical framework.

Keywords: social production, governance, Wikipedia, Capability Theory

1   Introduction

This paper aims to understand the evolving governance of complex social production over the internet. Our analysis focuses on Wikipedia, the pet example of popular oeuvre mapping the shifting institutional landscape in the beginning of the 21st century. The free online encyclopedia that anyone can edit, as the website describes itself, offers an intriguing phenomenon as well as an exceptional dataset to study how production takes place seemingly outside the market or the firm-based coordination of transactions. The question we ask is not in itself new, and indeed there are a number of empirical studies about Wikipedia on technology-related fields such as computer science, information systems, artificial intelligence, and human-computer interaction. Organization studies and institutional analysis have, on the contrary, been slow to take up the opportunity to explore the arguably new form of organizing. So far, the fields have not generated a theoretically valid account on how such vast productive experiments mediated by information and communication technologies actually come about and work. This is a significant shortcoming that impedes the investigations of social production.

Following Benkler (2006), we understand social production as an activity directed towards producing specific outputs in the absence of governance mechanisms based either on the price mechanism or a managerial structure for coordination. It has been argued that the internet and online environments are powerful enablers of social production, as they foster peer-based collaboration of volunteers (Benkler 2006). According to the standard accepted theories of economic behaviour such model of production should not be there. Yet it seems to work quite effectively, to the puzzlement of scholars who set out to search for explanations. We lack satisfactory answers on how a vast and distributed endeavour such as Wikipedia can sustain itself and achieve concerted performance when standard governance mechanisms do not, for instance, assign individual and collective tasks ex ante, enforce formal contracts of sorts, or define stable organizational membership in the classical sense. The system lacks many of the basic characteristics of modern organizations, and indeed we have only a vague idea of what kind of collective arrangement Wikipedia represents. Suggested partly by our own difficulties in handling the phenomenon empirically, we argue that the shortcomings in the empirical literature largely boil down to a lack of successful theoretical framing of the research object.

To address this problem, we adopt a perspective based on Capability Theory – a knowledge-based approach that has been developed within the theory of the firm (Dosi et al. 2000; Helfat et al. 2007; Kogut & Zander 1992; Kogut & Zander 1996; Langlois & Foss 1999). From this perspective an organization is seen as an accumulated repository of experience, skills, knowledge and practices, which make for the distinct capabilities that coalesce into an organized phenomenon. We propose to extend Capability Theory to the phenomenon of social production and argue that an emerging system of social production involves a process of building a specific collective capability to aggregate, evaluate, integrate and diffuse knowledge as a common resource. This capability is enacted, leveraged, and sustained by the specific features of the internet infrastructure that enable the swift conversion of online sociality into a productive force. We see our approach as complementary rather than in opposition to the economizing approach based on transaction cost considerations wittingly or unwittingly assumed in numerous previous studies (e.g. Benkler 2006; Shirky 2008; Tapscott & Williams 2006). In order to make the framework pertinent to empirical investigation, we suggest five hypotheses to test its propositions.

The paper is structured as follows. First, we briefly review previous literature on Wikipedia to discuss relevant findings and substantiate our argument about the lack of understanding on how governance has evolved over the years. Second, we discuss the limitations of the transaction costs argument and introduce Capability Theory as a more promising framework to account for the emergence of social production in general. Third, we use Capability Theory to build a tentative theoretical narrative of the evolution of Wikipedia governance. Fourth, we summarize the empirical hypotheses to test the validity of Capability Theory perspective.

2 Previous literature

The popularity of Wikipedia and its freely available database have inspired a growing body of literature from airport books to peer-reviewed articles and academic dissertations. In terms of questions motivating this study the relevant literature can be divided into three categories. First, there are fashionable books that use Wikipedia as a vignette for assumed changes taking place in society and organizations. Second, the epistemic quality of peer-production and the editorial processes have been scrutinized in studies assessing how well the encyclopedia stacks up against various yardsticks. Third, there are attempts to tackle the question of system coordination and governance.

2.1 New peer-based arrangements

In the popular literature Wikipedia is usually taken to represent a new kind of productive arrangement beyond markets and hierarchies (e.g. Benkler 2006; Tapscott & Williams 2006; Weinberger 2007; Zittrain 2008). These works tend to emphasise the communitarian ethos of large-scale peer interaction, under which the utilitarian view of the individual agent as a self-interested maximizer is replaced with a socialized agent who exercises his basic desire to share (Shirky 2008). The increasing weight of rule-making and governance-related activities in Wikipedia are occasionally acknowledged in these studies, but a serious analysis of the evolving mode of governance tends to escape the popular writings that rely on relatively simple arguments. The timing of publications allowing the authors to observe only the early stages of Wikipedia may have also contributed to the problem. Can informal norms and peer-to-peer interaction hold together such a massive system of collective production? Wikipedia would seem to express a number of design principles found in long-surviving, self-organized resource regimes (Ostrom 1990; Hess & Ostrom 2007). A significant and increasing proportion of activity in Wikipedia is not about sharing bits of knowledge by contributing to the encyclopedia content, but oriented towards maintaining and developing the resource regime itself (Beschastnikh et al. 2008). A careful reading of recent studies suggests that the mode of governance in Wikipedia has evolved over the years, and that governing commons-based peer production online might be a rather more complicated affair than the popular narrative conveys. Browsing through the different realms of Wikipedia reveals a myriad of activities oriented towards envisioning, shaping and maintaining the collective endeavour beyond the primary content production function.

2.2 Encyclopedic quality and editorial processes

The questions concerning the quality of a peer-produced encyclopedia have resulted in a number of studies on the epistemic characteristics, editorial processes and the coordination of work in Wikipedia. In general, the encyclopedia must balance between the requirements of extensive coverage based on open participation, mobilization and accessibility on the one hand, and reliability based on quality control, system coordination and regular maintenance on the other (Anthony et al. 2009; Halavais & Lackaff 2008; Spinellis & Louridas 2008). Kittur and Kraut (2008) demonstrate how coordination work represents a significant workload beyond immediate contributions to the encyclopedia content and the debate on individual articles. While blatant vandalism is a relatively uncomplicated matter to deal with, there are often cases when contributors put forward competing edits whose relative merits cannot be assessed without recourse to broader policies and guidelines (Viégas et al. 2004). Reaching an epistemic agreement on the article content is often based on a host of rules that need to be actively maintained and enforced. All in all, articles can go through various phases during their evolution (Gorgeon & Swanson 2009) ending up into a relatively formal Featured Article process reminiscent of traditional enterprise workflows (Viégas et al. 2007). Most importantly, the gradual maturation of the encyclopedia product has resulted in an overall shift in the meaning of quality from emphasising the coverage to emphasising the reliability of articles. It would seem reasonable to assume that such a shift may have implications on the overall task structure and the kind of contributors valued by Wikipedia.

2.3 The governance of social production

From the perspective of this paper the most relevant group of studies are the handful of empirical investigations into the various governance mechanisms in Wikipedia (Beschastnikh et al. 2008; Forte et al. 2009; Spek et al. 2006; Almeida et al. 2007; Burke & Kraut 2008; Bryant et al. 2005; Anthony et al. 2009; Leskovec et al. 2010; Kittur et al. 2007; Kittur et al. 2009). In general, these studies point to the emergence of increasingly prominent governance functions over the years; identify clear differences between novice contributors, subject experts and a group of people committed to maintaining Wikipedia itself; and attribute a minor role to anonymous users in the governance of the system. Furthermore, different types of administrative work do not merely inconspicuously exist in Wikipedia, but are readily acknowledged and appreciated by Wikipedians (McDonald et al. 2009).

Governance is required to deal with occasional external shocks and vandalism, but, more importantly, it seems to have grown endogenously as a response to the changing nature of system. Various mechanisms have become necessary to cope with the increasing complexity resulting from the growth of Wikipedia itself. For instance, Kittur et al. (2009) analyse interdependencies in editorial work and conclude that increasing the number of contributors to an article does not necessarily result in better performance. In order to be effective, peer interactions need to adhere to a division of labour and to rules that transcend the editing of individual pages. Forte et al. (2009) observe that the central policies and guidelines have been generally in place since 2005, but at the same time more refined, local scopes of policymaking known as WikiProjects have emerged around specific subject areas – resulting occasionally in conflicting jurisdictions within the system. Consistent with these findings Beschastnikh et al. (2008) notice a general shift from the development of rules to their enforcement.

The variety of research designs and disciplinary underpinnings make it difficult to integrate the findings from previous studies into a state-of-the-art account of Wikipedia governance, yet they reveal the evolving nature of the research object. It is clear, however, that simplistic views such as Shirky's (2008), that sees Wikipedia through the lens of self-assembled groupings in contradistinction to formal institutional structures, do not stand up against the empirical evidence. Shirky (2008, pp. 29–30) makes a difference between Wikipedia and an institutional structure that cannot "put all its energies into pursuing its mission; it [institution] must expend considerable effort on maintaining discipline and structure, simply to keep itself viable. Self-preservation of the institution becomes job number one, while its stated goal is relegated to number two or lower, no matter what the mission statement says." In the light of previous literature, this is exactly what seems to be happening in Wikipedia. The findings point to a considerable differentiation and, indeed, institutionalization of mechanisms governing social production.

All in all, the literature suggests that Wikipedia governance has evolved over the years and looks today quite different from the early days. The system that was initiated as a loose self-managing entity has come to rely on many mechanisms found in formal organizations. It looks today like a hybrid system that comprises both the community of committed Wikipedians whose discourse and self-understanding are still rooted in the communitarian ethos and the majority of contributors who hardly perceive their relationship with the system as communal bonding (Aaltonen & Lanzara 2010). This leads to our first empirical hypotheses (H1) that the Wikipedia contributor base is divided into distinct groupings based on the amount and types of effort individuals contribute.

3   A Capability Theory perspective

Governance in Wikipedia has evolved to a significant degree endogenously by the effort of active contributors to tackle emerging issues and to safeguard the development of the encyclopedia product. However, once the governance structures and mechanisms are in place they cannot but change the nature of the system and the circumstances for its own renewal (Butler et al. 2008; Konieczny 2009). The norms that once emerged from peer-interaction now regulate that interaction and are duly imposed

on the newcomers. It is the nature of this evolution that must be explained if we want to understand Wikipedia and social production in general.

3.1   Transaction cost explanation and its limitations

A view in good currency nowadays assumes that the Internet makes large-scale social production possible by reducing transaction costs and thus making it easier for people to engage in productive exchanges and collaborative interaction. In other words, people will be attracted to collaborate to the common endeavour if the barriers to entry into joint production and the costs associated with individual transactions are low enough. Whether this unleashes the basic human desire to share (Shirky 2008) or taps into more self-interested motivations does not have to make a difference here. It is easy to posit a number of possible reasons why an individual person would contribute to an article in Wikipedia; there is a perceivable utility in making sure that your PhD supervisor, political party, hobby or topic of professional interest is adequately represented in one of the world's most popular reference tools. In addition to the extrinsic or instrumental motivations, it has been argued that, for instance, the possibility of building positive reputation in a recognized community (Anthony et al. 2009) and participating in communicative action (Hansen et al. 2009) can motivate contributions. Assuming that the cost of making contributions is low enough there is a positive value to the transaction, which would be difficult to capture by market or organizational arrangements. The all but negligible cost of making small yet meaningful contributions is crucial, since it helps to break down the natural task structure at the level of individual characters making it possible for Wikipedia to benefit from a wide range of motivations and degrees of individual commitment (Shirky 2008; Anthony et al. 2009; Bryant et al. 2005). In this respect Wikipedia could be seen as a paradigmatic example of technological disaggregation and dynamic recombination of the traditionally en bloc organizational tasks (Castells 2000; Kallinikos 2006).

From these premises the mode of governance in Wikipedia has been usually assessed against the problem of keeping transaction costs in check. It is assumed that under specific circumstances and for specific purposes the internet makes peer-based or community-based governance relatively more efficient than the alternative modes of market or hierarchical authority. More recent studies have, however, brought to attention a gradual shift from the communitarian mode of governance typical of the early days toward increasingly hierarchical arrangements taking place as the social production systems matures (Butler et al. 2008; Kittur et al. 2007). This shift has been tentatively accounted for as a response to the growing complexity of coordination and control tasks (Kittur et al. 2007; Kittur et al. 2009). The observable expansion of more structured task-role and formal rule systems is then assumed to be driven by the necessity of curbing the increasing transaction costs running from the need to maintain adequate control and, for instance, the monitoring of misbehaviour. The underlying argument is, in other words, that when task and control complexity grow hierarchy and formal structures have a superior performance compared to community and market in reducing transaction costs.

In our view the transaction costs framework provides an insightful yet incomplete theoretical basis for explaining the emergence of Wikipedia and social production in general. The approach is a powerful tool for performing a static, comparative analysis on the alternative modes of coordination and governance that are already in place, but it seems less suitable for unpacking the evolution of Wikipedia governance. The approach generally assumes the existence of a system of exchanges, in which actors hold positions according to an already established division of labour and respond to incentives to exchange products, services or intermediary inputs (Williamson 1975). Such assumptions apply, perhaps, ex post facto, when assessing the governance of Wikipedia and social production in general, that is, once the system has reached a relative maturity, but fall short in accounting for the surprising expansion of social production. To put it short, the transaction costs framework misses the kind of innovation and generativity inherent in Wikipedia. We believe this is because the argument ignores that Wikipedia is essentially a knowledge-making phenomenon. The basic processes that have brought it to existence have to do with the increasing returns and the positive externalities associated

to a large-scale knowledge system, and with the building of a collective capability to sustain that knowledge. A contribution to Wikipedia is valuable not just because it enhances the content of encyclopedia but because of the learning and the generation of new knowledge it can trigger.

3.2   Production-oriented collective capability

In addition to facilitating exchange and reducing comparative transaction costs, the internet also has a generative capacity (Zittrain 2008) that seems to us more crucial in explaining online social production. The online environment has the capacity to enable and leverage distributed, heterogeneous and unsynchronized human activity to a range of different tasks – in the case of Wikipedia to the creation of commons-based knowledge resources (Ostrom 1990; Hess & Ostrom 2007). We argue that it is such emergent collective capability at creating value rather than transactional efficiency that, under specific conditions, makes social production more effective than the market or firm-based coordination of transactions and provides Wikipedia its competitive advantage over traditional encyclopedias such as Encyclopœdia Britannica. In contrast to traditional encyclopedia production that revolves on obtaining comprehensive and completed articles from formally contracted experts and specialists, inserting merely few words or correcting a single typographic error by an anonymous user can amount to a valuable input to Wikipedia. The latter relies on the internet-based infrastructure to directly inscribe distributed social knowledge into an encyclopedic format, creating more value for more people at lower production costs and giving more open and easier access to a common good.

The idea that organized production and task-oriented coordination are an expression of knowledge-based organizational capabilities was originally proposed within the knowledge-based view of the firm (Grant 1996; Kogut & Zander 1992; Kogut & Zander 1996; Madhok 2002). These authors have argued that the transaction cost approach is indeed necessary but not sufficient to explain the nature of the firm, inasmuch as it limits itself to a negative argument about minimizing the sum of production costs and the costs associated with economic exchanges, that is, transaction costs. It considers the firm as a nexus of contracts, but ignores that it is also, and most critically, a bundle of knowledge. Firms exist because they are better than markets at creating and using organizational and technological knowledge and at fostering innovation (Nonaka & Takeuchi 1995). It is this idea we transpose to the context of social production and make a parallel argument: social production systems embody a collective capability at creating and maintaining commons-based knowledge resources, which is difficult to develop by means of the contractual forms of firms and markets. As a community or organized system Wikipedia has developed certain unique capabilities that enable it to effectively create and distribute value both in spite and because of the missing contractual bond between the agents (Kallinikos 2010).

The idea is relatively simple. Wikipedia contributors, as they edit individual articles, develop at the same time also collective practical knowledge on how to create and maintain a peer-produced encyclopedia. Importantly, the technological infrastructure makes it possible to capture this learning by embedding the medium of governance into the very same medium on which the production takes place (Aaltonen & Lanzara 2010). In Wikipedia rule-making and governance in general takes place by writing policies about writing encyclopedia articles both residing as pages on the same MediaWiki platform. The technical skills, experiences and learning acquired in the domain of production are thus readily transferable into the domain of governance. In contrast, the subject experts contributing to Encyclopœdia Britannica are unlikely to develop any new knowledge about or even understand the intricacies involved in maintaining such a body of knowledge. Even if they would the organizational arrangement itself could hardly benefit from such learning.

4   The evolution of Wikipedia governance

In this section we apply the Capability Theory perspective to understand the evolution of Wikipedia governance. Based on the selective reading of previous studies and our own empirical work we

develop a theoretical narrative to explain the evolution of Wikipedia, from which we then derive further empirical hypotheses to test the Capability Theory perspective. Our approach assumes that the development of Wikipedia (as well as other social production endeavours) is associated with the building of a collective capability at creating value – a capability which was not there at the beginning, but was developed endogenously as more and more people contributed to the encyclopedia product, learned to work together and became socialized into the system. Furthermore, we believe such capability is fundamentally mediated by the online environment and cannot be understood independently of the underlying computational infrastructure. This amounts to saying that contemporary information and communication technologies are a vital factor for the emergence of the new collective capability.
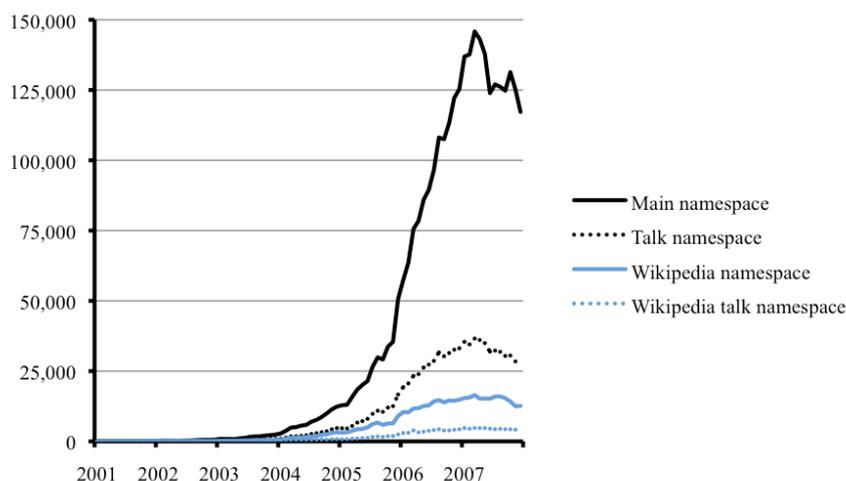

Figure 1. *The number of monthly contributors in the main, Talk, Wikipedia and Wikipedia talk namespaces 2001–2008 (adopted from Aaltonen & Lanzara 2010). Namespaces represent different kinds of artefacts (pages) in the overall morphology of Wikipedia. The encyclopedia articles reside in the main namespace while policies and guidelines are written in the Wikipedia namespace. Talk namespaces contain discussion regarding the their respective namespace.*

We divide Wikipedia evolution into three phases on the basis of trends observed in Figure 1 depicting the number of monthly contributors to the encyclopedia articles and to the domain of policy-making as well as the respective discussion forums. All in all, the growth of contributors would seem to approximate an S-shaped curve with a slow initiation, rapid growth and eventual levelling off not unlike the one found in many learning phenomena. In the following we give a brief account of each phase from the capability perspective and, more importantly, suggest dynamics that has pushed the Wikipedia from one to another.

4.1   Tapping and exploiting distributed individual capabilities (2001–2003)

The English Wikipedia was initiated in January 2001. For the first three years the system attracted only a relatively small amount of contributors. For instance, less than three thousand people made contributions in December 2003[1]. Contrary to the assumptions of the transaction cost argument, there was no clear division of labour or specialization among contributors, and an exchange system that could have been comparatively assessed did not simply exist. To wit, there was not even an established community, that was yet to take shape. Everyone was invited to do whatever he or she saw

---

[1] The figure is based on our own work with the freely available Wikipedia dataset.

fit, as nobody really knew how to successfully create a peer-produced encyclopedia online. Such knowledge or capability could not exist because no such product had been created before. Individual contributors learned by doing, exploring different approaches within an essentially weak system. By editing the article content the individual contributors started, however, to gradually accumulate knowledge and experience about producing a peer-based encyclopedia.

Transaction costs considerations were presumably negligible at the outset (and for a fairly long subsequent period) in fostering participation and driving evolution. Any coordination problem between different contributors' work was tackled directly through communication and mutual adjustment among peers (Thompson 1967). As long as the number of contributors remained sufficiently small the editors themselves easily coordinated the collective production and the editing of articles. Governance was assured by few basic principles and mostly embedded in the interaction between individuals. Indeed, it seems more likely that the early expansion was largely driven by an endogenous increasing returns mechanism that gained momentum as the system itself attained a critical mass of contributors and activity. The initiators were probably more concerned how to attract contributors and to enhance productivity rather than curbing transaction costs. The requirement of economizing on transaction costs became more relevant only later on as the generated amount of collective work and coordination requirements became more conspicuous.

In these early stages the critical capabilities required by the system were mostly the individual capabilities at writing content on the variety of topics. These capabilities were found to be available and diffused in society as a potential, unexploited cognitive surplus (Shirky 2010) that needed to be tapped, mobilized and pooled in order to assemble a rudimentary online encyclopedia. The primary invention of Wikipedia entailed thus the creation of an internet-based mechanism for harnessing such individual capabilities to build up the bulk of a commons-based knowledge resource. In this phase the variety, coverage and the rapid expansion of content were paramount, while less attention was given to coordination work.

4.2 Take off and the building of collective capability (2004–2007)

The task of collectively editing articles began to change as the number of individual contributors grew increasingly large. After some three years of slow initial growth, Wikipedia reached a critical mass of contributors that allowed it to become an increasingly powerful mechanism for mobilizing and attracting distributed resources. Figure 1 suggests that the switching point when the exponential growth took off falls somewhere in 2004. However, as a result of its increasing success and rapid growth Wikipedia had to progressively develop a set of coordinating tools to guide, evaluate and structure the skyrocketing number of contributions. Over the next three years the system learned to cope with its own complexity as individual contributors gradually externalized their knowledge into rules, guidelines and other governance instruments, which consequently took an increasingly important role in governance vis-à-vis peer-interaction. This leads to our second hypotheses (H2) that the configuration of tasks and groupings of contributors has evolved over the years.

This kind of organizing technology is different from the formal organization and constitutes an asset specific to Wikipedia. It makes the online encyclopedia not necessarily a more efficient governance mechanism, but a more effective learning system in terms of how to exploit the expanding contributor base. With respect to the firm and formal organizations in general Wikipedia embodies a swifter mechanism for accumulating and structuring knowledge, exhibiting less stickiness in converting resources to productive use, higher speed and flexibility in the editing process, better self-monitoring capabilities, and effective inscription of organizing knowledge into the evolving set of rules. Wikipedia's emerging governance structure is therefore more usefully regarded as a vehicle to manage skills and knowledge in specific contexts rather than simply serving to contain the transaction costs.

We might call this 'the collective capability development phase'. The contributors kept working on the joint production of articles, but at the same time they created and organized collective practical

knowledge on how to edit an online encyclopedia. This process resulted in the development of a different kind of collective capability, that is, the capability at editing the encyclopedia as a complex knowledge artefact that transcends its individual articles. This capability developed as an endogenous response to the growing social and task complexity of Wikipedia. New requirements emerged and became critical to the overall value of the online encyclopedia as a commons-based knowledge resource: quality, reliability, structure, usability, and consistent maintenance. After all, the value of encyclopedia resides in being a structured system of knowledge and a reliable reference tool, which goes beyond the simple accumulation of articles into a generic repository.

The cost of building the collective capability (beyond contributing to the content itself) is not zero. Not only production but also learning how to carry it out took time and effort, especially at the time when both production and editorial coordination tasks grew rapidly in complexity. From the capability theory perspective the evolution of Wikipedia towards increasingly hierarchical arrangements and a rule-based environment is understood to be driven by the expanding capability and by the necessity of capturing the benefits of collective production. Figure 1 suggests a second shift in the internal dynamics of Wikipedia early 2007.

4.3 Consolidating collective capability into role and rule structures (2007–)

Collectively produced rules (the system) became gradually more intelligent than any of its individual contributors, who increasingly needed to learn from the rules. This kind of learning was, however, less about creating new knowledge about social production and more about acquiring or internalizing the knowledge on how to contribute efficiently. Newcomers became socialized into an increasingly intelligent system with diminishing opportunities to explore and innovate with the dynamics of social production, while experienced contributors derived value from their established positions in Wikipedia. As the value of the common resources grew bigger it became then more critical to keep, protect and enhance such value over time and new capabilities had to be developed for this purpose. The focus shifted from developing effective rules to their efficient enforcement.

Throughout the years Wikipedia has undergone transformations that have presumably made it better equipped to face the challenges and problems its own success has generated. As a collective endeavour, Wikipedia has learned to enhance and protect its own collective capability in editing and managing the encyclopedia. By doing so the system has also had an unavoidable impact on the kind of human resources it elicits from its environment: not just knowledgeable individuals burning to share their knowledge on specific topics, but disciplined individuals who are willing to play a part in a complex editorial process; concerned with the management of overall quality; and ready to comply with the rules. All in all, the social production system has undergone changes as endogenous responses to the requirements of developing and enhancing an effective capability at governing the system. Instead of minimizing the transaction costs, these changes were driven by the necessity of creating a common knowledge resource and maintaining its value over time. The editorial production of a large-scale online encyclopedia came to require a support structure to could assist the contributors and stabilize the product. This had less to do with the monitoring of misbehaviour or with cutting the costs of organizing, and more with creating and managing the conditions that support the acquisition and collection of knowledge, its usability and sustainability in time.

This line of reasoning leads us to suggest that the observable tendency towards internal organization and bureaucratization can be conceptualized as an endogenous learning phenomenon: the system learns how to respond to the very same conditions that it has created by restructuring itself. More pointedly, it learns by differentiating and stabilizing skills and by inscribing collective knowledge into policies and guidelines. As Wikipedia kept expanding over the years, some contributors began to perceive the growing importance of administrative problems and started to take care of them – hence the gradual emergence of a class of editors and administrators. Some people may then tend to keep doing the things of which they have more experience, and this in turn increases and stabilizes their competence. This leads to our third and fourth hypotheses that (H3) the contributors working with

policies and guidelines represent a relatively stable minority of the overall contributor base, and (H4) the configuration of tasks and contributors has become progressively more complex and stable.

The editorial function became gradually too complex and too critical to be left to the miscellaneous content producers. It was then embedded in rules, policies and technical devices inscribing the knowledge that the system had developed about itself. These instruments both integrated the collective learning into Wikipedia and at the same time shaped any further learning by the individual contributors and the system. They contained the growing task complexity and created a moral world for the contributors thus stabilizing individual behaviour and the overall system. It is at this stage that Wikipedia, as a whole, became a knowledge commons (Hess & Ostrom 2007) and a shared value; and it is also at this stage that a full-fledged collective capability for governing the system as a whole consolidated and began to differentiate from the content-focused capability. The agglomerative function enacted by the internet infrastructure became then also a generative environment for the emergence of more fundamental collective capability of how to build and manage a peer-based online encyclopedia. Individual capabilities at producing and editing content (articles) were necessary but not sufficient to the emergence of Wikipedia's collective capability. The latter had to do with organizing the capabilities of a scattered collection of people and not those individual capabilities per se.

5 Empirical hypotheses

As we see it, the theoretical narrative presented in the previous section suggests a plausible story regarding the evolution of Wikipedia. A number of questions need, however, to be tackled empirically: Where is the collective capability is located in Wikipedia? What kind of structure does it exhibit? How has the capability changed over the years? In this section we briefly summarize the set of empirical hypotheses and metrics to answer these questions. Empirical research following these or similar hypotheses could give us hints about the plausibility of our theoretical framework. The work does not necessarily have to adopt a quantitative methodology, yet the prospect of successfully testing the hypotheses against the full historical trace of every contribution ever made to Wikipedia is intriguing. Table 1 lists four empirical hypotheses derived from the Capability Theory based explanation of Wikipedia evolution. We have also added fifth hypothesis (H5) that is intended to validate the claims about the relevant differentiation of the user base.

Hypothesis	Null hypothesis	Rationale
H1: Contributors differ both in terms how many and what kind of edits they make	Contributors differ only in their level of activity	The first task would be to assess if there is a differentiation into distinct contributor/task groupings
H2: The configuration of tasks and contributors has evolved over the existence of Wikipedia	The configuration of tasks and contributors has remained stable (despite the increasing number of contributors)	The Capability Theory argument suggests the differentiation of contributor/task groupings has evolved over time
H3: Contributors working with organizational policies and guidelines represent a stable minority of the contributor base	Contributors working with organizational code represent a random subset of the overall contributor base	The differentiation of contributor/task groupings is assumed to be accompanied by learning and, thus, increasing specialization within the overall task structure
H4: The configuration of tasks and contributors has become progressively more complex and stable (the configuration varies less from period to another towards the end of the dataset)	The configuration of tasks and contributors has become progressively more complex but does not appear to stabilize	Learning and specialization among increasingly differentiated task structure should result in stable contributor/task groupings perhaps resembling to some degree organizational roles
H5: The group of contributors who take up managerial tasks is varies less from month to month than the group of content contributors.	The turnover of contributors does not differ between those clusters who provide article content and those take up managerial tasks	Contributors whose are involved in the higher order capability of building peer-based encyclopedia are intensively committed to it

Table 1.   Hypotheses H1–H5 to test the narrative based on the Capability Theory perspective.

Wikimedia Foundation has made the databases of different Wikipedia sites freely available for research purposes. The largest of these contains all surviving edits to the English Wikipedia since the inception of system on 16 January 2001. For instance, a dump file from 30 January 2010 represents nearly 350 million edits onto 19 million pages over nine years. This raw data can be transformed into a dataset containing a comprehensive behavioural trace at the level of individual contributors and their actions, which gradually brought Wikipedia into existence. While previous studies on Wikipedia governance have, by and large, focused either on the content production, editorial coordination, or else observed the governance functions at the aggregate level, the dataset would allow at the same time more detailed and comprehensive analyses along the temporal dimension. For instance, Aaltonen and Lanzara (2010) suggest using cluster analysis to understand the differentiation of user base into groupings that revolve on particular tasks and domains within the overall system.

We suggest that the hypotheses outlined in Table 1 could be analysed quantitatively using three metrics. First, an individual Wikipedia edit can be measured according to its size and the type of page edited. The latter is relatively easy to operationalize as the namespace of the edited page. The former is somewhat more complex and computationally demanding to operationalize (e.g. Kittur et al. 2007) but can be tackled using one of the many algorithms developed for assessing the difference between two character strings. Second, the type of individual edits needs to be complemented by the information about the distribution of edits across different kinds of pages by the contributor. This is a simple aggregate metric. Using these three variables and, for instance, cluster analysis approach we believe it would be possible to shed light on the hypotheses put forward here and thus better understand Wikipedia as a knowledge-making phenomenon.

6 Concluding remarks and future work

In this paper we have proposed a theoretical framework to understand the rise of Wikipedia and its shifting mode of governance. We argue that the governance of social production needs to be understood as an evolving phenomenon, that is, by unpacking the structure or mechanisms driving its change. Drawing from our own empirical investigation and previous literature on Wikipedia, we argue that the transaction cost approach assumed in several studies is merely necessary but not sufficient theoretical scaffolding for analysing social production. Instead, we suggest a Capability Theory perspective to account for the endogenous evolutionary mechanism in Wikipedia. We see Wikipedia evolution as a learning phenomenon that over time gives rise to governance mechanisms and structures as responses to the problems and conditions that the ongoing development of Wikipedia has produced over the years. This amounts to a process of building a collective capability, that is, the capability of editing and managing an online encyclopedia. Also, our analysis points to what perhaps could be seen as Wikipedia's major challenge in the years to come, that is, the maintenance and enhancement of such collective capability. To this purpose effective governance mechanisms should be enacted that provide a dynamic balance between the requirements of system's openness and resource mobilization on the one hand and, on the other hand, the requirements of product reliability and quality control. Both bureaucratization or, alternatively, loss of structure can be equally expected as evolutionary outcomes in the future.

Given the limited space available, we have been able to discuss the suggested approach only in its main aspects without going into much detail. We have spelled out the implications of our approach in the form of a theoretical narrative on the evolution of Wikipedia and further elaborated it into empirical hypotheses that could be used to test the approach. An important aspect that we have been able to merely hint at and would certainly merit more thorough treatment is the specificity of online medium in terms of governing social production. The extreme granularity of contributions and the way the internet is able to embed the medium of governance into the medium of production are almost certainly vital to Wikipedia. We would assume, for instance, these two aspects to account to a significant degree for how the system is able to tap into a wide variety of motivations and levels of commitment as well as make it easy to transpose skills and experience acquired in the content

production to governing the system. In contrast to the quantitative, hypothesis-testing approach suggested to assess our theoretical narrative, analysing the implications of granularity and embedded governance would perhaps call for an intensive qualitative study with a sharp focus on few decisive aspects of Wikipedia.

References

Aaltonen, A. & Lanzara, G., 2010. Unpacking Wikipedia governance: The emergence of a bureaucracy of peers? Paper in 3rd Latin American and European Meeting on Organization Studies. Buenos Aires, Argentina.

Almeida, R.B., Mozafar, B. & Cho, J., 2007. On the Evolution of Wikipedia. In Proceedings of the International Conference on Weblogs and Social Media. ICWSM 2007. Boulder, Colorado, USA.

Anthony, D., Smith, S.W. & Williamson, T., 2009. Reputation and Reliability in Collective Goods. Rationality and Society, 21, pp. 283–306.

Benkler, Y., 2006. The Wealth of Networks. New Haven: Yale University Press.

Beschastnikh, I., Kriplean, T. & McDonald, D.W., 2008. Wikipedian Self-Governance in Action: Motivating the Policy Lens. In Proceedings of the 2008 AAAI International Conference on Weblogs and Social Media.

Bryant, S.L., Forte, A. & Bruckman, A., 2005. Becoming Wikipedian: Transformation of Participation in a Collaborative Online Encyclopedia. Paper presented in GROUP'05. Sanibel Island, Florida, USA.

Burke, M. & Kraut, R., 2008. Taking Up the Mop: Identifying Future Wikipedia Administrators. In CHI 2008 Proceedings. Florence, Italy.

Butler, B., Joyce, E. & Pike, J., 2008. Don't Look Now, But We've Created a Bureaucracy: The Nature and Roles of Policies and Rules in Wikipedia. Paper presented in CHI 2008. Florence, Italy: ACM.

Castells, M., 2000. The rise of the network society, 2nd ed., Oxford: Blackwell.

Dosi, G., Nelson, R. & Winter, S. eds., 2000. The Nature and Dynamics of Organizational Capabilities, Oxford University Press.

Forte, A., Larco, V. & Bruckman, A., 2009. Decentralization in Wikipedia Governance. Journal of Management Information Systems, 26(1), pp. 49–72.

Gorgeon, A. & Swanson, E.B., 2009. Organizing the Vision for Web 2.0: A Study of the Evolution of the Concept in Wikipedia. Paper presented in WikiSym'09. Orlando, Florida: ACM.

Grant, R.M., 1996. Prospering in Dynamically-competitive Environments: Organizational Capability as Knowledge Integration. Organization Science, 7(4), pp. 375-387.

Halavais, A. & Lackaff, D., 2008. An Analysis of Topical Coverage of Wikipedia. Journal of Computer-Mediated Communication, 13, pp. 429–440.

Hansen, S., Berente, N. & Lyytinen, K., 2009. Wikipedia, Critical Social Theory, and the Possibility of Rational Discourse. The Information Society, 25(1), pp. 38–59.

Helfat, C. et al., 2007. Dynamic Capabilities, Malden, MA, USA: Blackwell Publishing.

Hess, C. & Ostrom, E. eds., 2007. Understanding Knowledge as a Commons, The MIT Press.

Kallinikos, J., 2010. Bureaucracy Under Siege: On Information, Collaboration and Networks. In S. R. Clegg, M. Harris, & H. Hopfl, eds. Managing Modernity: The End of Bureaucracy? Oxford: Oxford University Press.

Kallinikos, J., 2006. The Consequences of Information, Cheltenham, UK: Edward Elgar.

Kittur, A. et al., 2007. Power of the Few vs. Wisdom of the Crowd: Wikipedia and the Rise of the Bourgeoisie.

Kittur, A. & Kraut, R.E., 2008. Harnessing the Wisdom of Crowds in Wikipedia: Quality Through Coordination. In Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work. The 2008 ACM Conference on CSCW. San Diego, California, USA.

Kittur, A., Lee, B. & Kraut, R.E., 2009. Coordination in Collective Intelligence: The Role of Team Structure and Task Interdependence. In Proceedings of the 27th International Conference on Human Factors in Computing Systems. The 27th International Conference on Human Factors in Computing Systems. Boston, MA, USA: ACM, pp. 1495–1504.

Kogut, B. & Zander, U., 1992. Knowledge of the Firm, Combinative Capabilities, and the Replication of Technology. Organization Science, 3(3), pp. 383-397.

Kogut, B. & Zander, U., 1996. What Firms Do? Organization Science, 7(5), pp. 502–518.

Konieczny, P., 2009. Governance, Organization, and Democracy on the Internet. Sociological Forum, 24(1), pp. 162–192.

Langlois, R.N. & Foss, N., 1999. Capabilities and Governance. Kyklos, 52(2), pp. 201-218.

Leskovec, J., Huttenlocher, D. & Kleinberg, J., 2010. Governance in Social Media: A Case Study of the Wikipedia Promotion Process. In Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media.

Madhok, A., 2002. Reassessing the Fundamentals and Beyond: Ronald Coase, the Transaction Cost and Resource-Based Theories of the Firm and the Institutional Structure of Production. Strategic Management Journal, 23(6), pp. 535–550.

McDonald, D.W. et al., 2009. System Design for Social Translucence in Socially Mediating Technologies. Paper presented in CHI 2009. Boston, USA.

Nonaka, I. & Takeuchi, H., 1995. The Knowledge-Creating Company, Oxford: Oxford University Press.

Ostrom, E., 1990. Governing the Commons, Cambridge, UK: Cambridge University Press.

Shirky, C., 2010. Cognitive Surplus: Creativity and Generosity in a Connected Age,

Shirky, C., 2008. Here Comes Everybody, New York: The Penguin Press.

Spek, S., Postma, E. & van den Herik, H.J., 2006. Wikipedia: organisation from a bottom-up perspective. In Proceedings of the 2006 International Symposium on Wikis. ACM.

Spinellis, D. & Louridas, P., 2008. The Collaborative Organization of Knowledge. Comm. of the ACM, 51(8), pp. 68–73.

Tapscott, D. & Williams, A.D., 2006. Wikinomics, London: Atlantic Books.

Thompson, J.D., 1967. Organizations in Action, New York: McGraw-Hill.

Viégas, F.B., Wattenberg, M. & Dave, K., 2004. Studying Cooperation and Conflict between Authors with history flow Visualizations. CHI Letters, 6(1), pp. 575–582.

Viégas, F.B., Wattenberg, M. & McKeon, M.M., 2007. The Hidden Order of Wikipedia. In Online Communities and Social Computing. Lecture Notes in Computer Science. Berlin: Springer, pp. 445–454.

Weinberger, D., 2007. Everything Is Miscellaneous, New York: Times Books.

Williamson, O.E., 1975. Markets and Hierarchies: Analysis and Antitrust Implications, New York: Free Press.

Zittrain, J., 2008. The Future of the Internet – And How to Stop It, New Haven: Yale University Press.