Online Communities, Teams Characteristics, and Knowledge Quality

Amira Rezgui, Nicolas Jullien

▶ **To cite this version:**

Amira Rezgui, Nicolas Jullien. Online Communities, Teams Characteristics, and Knowledge Quality. 17th Working Conference on Virtual Enterprises (PRO-VE), Oct 2016, Porto, Portugal. pp.35-47, 10.1007/978-3-319-45390-3_4 . hal-01532045

HAL Id: hal-01532045
https://hal.science/hal-01532045

Submitted on 2 Jun 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Online Communities, Teams Characteristics, and Knowledge Quality

Amira Rezgui, Nicolas Jullien[1]

[1] Telecom Bretagne-UBL, M@rsouin-LEGO, Brest, France
{Amira.Rezgui, Nicolas.Jullien}telecom-bretagne.eu

Abstract. Understanding the characteristics of a "good" team and members diversity affects the outcomes group in a question of growing importance for the organizations, for their competitive advantage relies more and more on innovation, produced by virtual cooperation on knowledge production. In this study, we propose a method to forecast the future quality of an online knowledge production community -or online epistemic community- by studying the composition of the group who initiated them (the "core members" of, in this case, an article). First, we set a team building period which is defined as the period of 120 days after article creation in order to construct this "core members" group. Second, we explore the effects on article quality of both group and member diversity. Core members' characteristics are learned from their previous behavior. The analysis is based on the French Wikipedia project. Our results show that the most important attributes of initial core member to have a high quality article are average reputation, diversity's contribution, participation, and group size. We also find no significant effects of experience diversity and reputation during the team building period.

Keywords: Wikipedia, Epistemic community, Article Quality, Charac-teristics of the team, Core member, group composition, diversity.

1 Introduction

Computer technologies have enabled new forms of open collaboration within online epistemic communities [1], such as open source software projects, or Wikipedia, the largest free and open access online encyclopedia. These communities, as an online self organizing group, perform a wide range of activities such as writing Wikipedia articles and developing softwares, but also prioritize the work of the participants (via bug lists or articles in need of improvement), or manage the global organization of the project. If this organizational model of online epistemic community, or community of creation as named by [2], is viewed as central in the generation of new, innovative knowledge by and forrms, the path to successful community building is still risky and uncertain, and as for

business building, most of the attempts fail, no matter how hundreds of thousands of dollars were spent.

The question of how to integrate newcomers, but also what a good (efficient and effective) team means are intensively discussed in the recent literature. Many crucial factors, at various stages of the group composition, are stressed: motivation of participants [3], governance structure [8], culture and ideology [4], social structure and network ties [6] [7] and social identity [8] [9]. In general, these studies highlight the reasons individuals participate in online self-organizing groups and the manner individual efforts are organized to accomplish the goals of the group. In particular they shed light on the processes contributing to Wikipedia's success and the role of coordination along with contribution inequality. They are less prolix on the characteristics of the members who comprise the groups, and how their diversity affects the quality of the knowledge produced by their virtual work.

To understand the production of article as common knowledge, we first base our analysis on the Institutional Analysis and Development framework described in Fig.1. This framework distinguishes the characteristics of the community, or the "inputs" ("biophysical characteristics", "attribute of the community", "rules-in-use") from "the action arena" which constrain the way people interact, leading to "outcomes" [10].

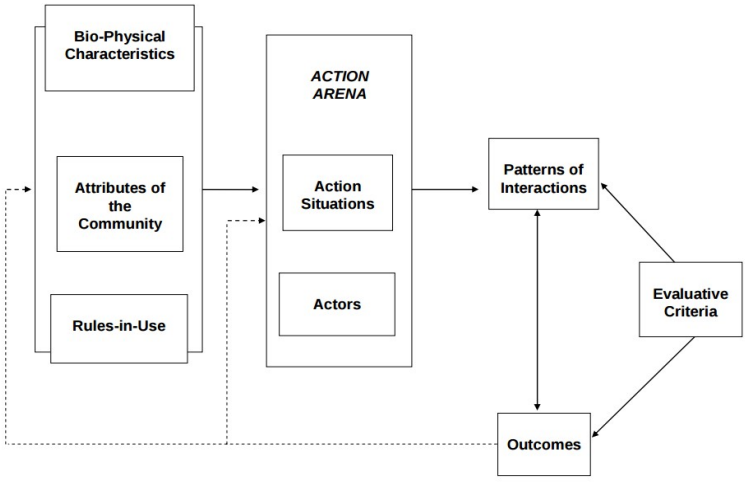

Fig. 1. Institutional Analysis and Development Framework [10]

Based on this framework, we cannot overlook neither the effect of the process nor the effect of the community attributes on the quality of the articles. Before talking about the process and the action arena, we need to identify the appropriate group based on member attributes, because they affect greatly how members work with one another and how effective their collaboration efforts will be. Research in traditional organizations has linked group characteristics and member diversity to performance [11]. However,the

negative effects of these factors may be curtailed in teams mediated by collaborative technologies because of reductive capabilities such as visual anonymity and equality of participation. Similarly, their positive effects may be enhanced because of additive capabilities such as coordination support and electronic trial. So, in one side, two key factors are to be considered: group characteristics in general and member diversity. More specifically, the focus should be placed on characteristics reflecting the extent to which group members share similar or different attributes.

On the other hand, another distinctive feature of online groups is their self-organizing nature. Unlike work groups or virtual teams in organizations whose memberships are determined by organizational design and managerial oversight, group composition in online self-organization is driven by members' voluntary participation. So it is important to look at membership constitution based on the career of the contributors before the creation of the article. This factor plays an equally important role in the success of online collaboration e orts. However, this ideas is not developed in the case of Wikipedia as an online epistemic community. The majority of the existing studies on the subject focus on what happen after article creation. They link knowledge quality to group characteristics within the same project but neglect what may have happened before the studied project.

Insights from our study will advance knowledge of conditions under which this crowding group thrives in online collaboration to produce useful knowledge. In this article, we inspect two research questions:

1. What are the characteristics of a "good" team to produce new knowledge?
2. How does member diversity or similarity a ect the outcomes that a group produced in an online epistemic community?

As a primary contribution of our work we developed a set of theoretical propositions about initial structure in terms of group composition and member diversity in virtual teams based on what happened before article creation. We further integrated various social theories to develop these propositions by considering how structure could be instantiated in shared mental models and the specific behaviors that contribute to building such models.

The paper continues as follows: the next section reviews related work on group composition and core members. The following section develops hypotheses regarding the relationships between these constructs and the quality of knowledge produced. Then, we describe our research methods following with our results. Finally, we discuss our findings and highlight implications for both theory and practice, before our concluding remarks with possible future research directions.

2 Literature Review

According to Tan et al. [12], group composition is an important determinant of project success in community-based projects, and there is an extensive body of literature that describes the core-periphery structure in online communities (speci cally, peer production communities) [13]. In developing our hypotheses, we build on two sets of literature. The first is the group composition literature and the second is the extensive core members and their relationships with knowledge quality. We explain what effects we expect to persist in Wikipedia and extend these frameworks where necessary.

2.1 Group composition

Study research in team composition focuses on the attributes of team members and the impact of attributes combination on processes and outcomes [14]. There are two important dimensions of group composition that have been investigated in literature: the mean level of group's characteristic, and the diversity within the group [13].

In one side, the mean level is considered as the proportion of group members who possess a simple attributes combination, which is averaging lower level units to represent a higher level construct [15][16]. Emergence form underling the average of member attributes is designed as a summary index and relies on some measures of central tendency of members' characteristic [14].

On the other hand, diversity in group work is defined as heterogeneity or dissimilarity of individual attributes that arises from the di erence of any attribute from others [9]. These attributes can be social demographic characteristics (age, gender, race, and nationality), informational ones (tenure, experience, education and functional areas) or deeper individual differences (personality, values and beliefs). The relationship between team diversity and outcomes has been extensively [9].

This line of research has focused on the dimensions of the group's activity, based on what is done during article creation. It does not explain how member's previous activity, in itself, would determine the quality of a group's outputs. To forecast knowledge quality, we investigate these two dimensions of group composition, focusing on the core members on the the basis of their previous behavior.

2.2 Core member constitution

Some studies on epistemic communities, especially in the case of open source software, stressed the importance of the core members [17] [13] [18]. In this context, they define the core-periphery structure in terms of members' activity in the project, where a small group of highly active core members are responsible for most of the contribution to the project and a large and loosely coupled group of periphery members support the others.

The core members are generally more active than other members, and their activity is often spread across many tasks. The presence of such a core set has become the signature of successful online communities.

These studies conduct to making the hypothesis that in Wikipedia projects too, there is a core-periphery structure, similar to the one found in open software production, and that the initiators of the projects are its 'natural' core members, and are key to its success. In this study, we examine the question of how to identify core members of an article and if its characteristics impact the quality of the article produced.

3 Research Hypotheses

While most studies look at the career of contributions after the first contribution, we focus on what happened before. We define the first contribution as the first time an editor makes a revision and proposes a new article as a new and original knowledge.

Based on these definitions we define article creation life cycle as described in Fig.2. The first contribution is a clear separator, which makes it possible to define three main periods. The learning period is the period before the first contribution where we learn about the members attributes based on their history. The teaming period is the period of four months after the first contribution where we constitute the core members. The active period is the period just after the teaming period when teams start activities [1]. These definitions help to refine our main hypothesis:

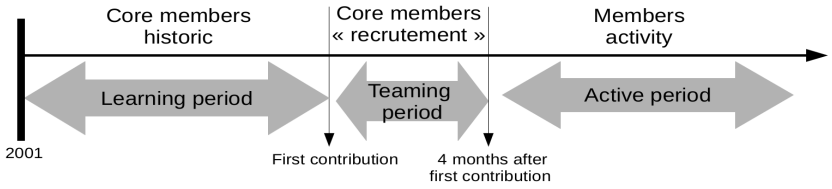

Fig. 2. Article life cycle among Wikipedia

H:"The teaming period carries important information about the initial group composition using what we learned about the members in the first period which impacts in turn the quality of knowledge produced by the group."

Our main hypothesis argued that group composition is an important determinant of project success in Wikipedia as a community-based platform. The manner in which self-formed groups attract and recruit participants can have a profound effect on group composition, which in turn in uences knowledge quality. There are several dimensions of group composition that have been investigated. In our study a key factor to consider is member characteristics, within the virtual work environment, at both individual and group levels. In addition, we provide valuable insights into their diversity which reflects the extent to which group members share similar or different attributes.

3.1 Effects of initial group size on the quality of knowledge produced by a virtual group

Many studies have been trying to determine whether small or large groups are more likely to cooperate on a project and produce knowledge. Group size impact quality in two ways. In one hand, large groups ensure a large set of opinions and knowledge and a faster time to corrected errors and discovered incomplete information [19]. On the other hand, small groups often lack the resources that large groups can extent. These limited resources, make difficult to give additional resources to producing article within Wikipedia as a collective action [20]. Ostrom suggested further research on collective action to focus on the hypothesized curvilinear effects of group size [10]. Based on Ostrom proposition we hypothesize that:

H1: "Group size in the teaming period is an important determinant of project success in online epistemic community which has a curvilinear effects on article quality."

3.2 Effects of diversity of member characteristics within the virtual group on the quality of knowledge produced

Some studies on team dynamic, show that heterogeneous teams are more productive than heterogeneous isolated workers in the case of low-skilled worker. Our study of diversity in Wikipedia expands the diversity literature to the context of virtual teams. We test and confirm theoretical propositions of the effects of diversity to know if diversity should be encouraged or discouraged.

3.2.1. Effects of experience disparity on the quality of knowledge produced by a virtual group.

The members comprising a team may be classified according to their experience. Some agents are newcomers having a little experience and skills. The other are old-timers or incumbents, persons with identi able talents named.

Evidence from virtual organizations has shown that, although old-timers are more experienced and skilled than newcomers, their e ort is generally lower [21]. Having a blend of incumbents and newcomers ensures a sufficient group experience to establish and maintain task structure, and in the same time acquires new ideas and information to complete the task [9]. In the meantime, when experience disparity increases, old-timers and new-comers may have di erent collaboration work views for example on article scope and interpretations of Wikipedia policies. So experience diversity may reduce communication and social integration and thus it has been linked to increased conflicts. Therefore, our next hypothesis:

H2: "There is a non linear relationship between experience diversity and article quality. Article quality increases as experience diversity increases. However, beyond certain levels of experience diversity, article quality will decrease."

3.2.2 Effects of reputation diversity on the quality of knowledge produced by a virtual group

User reputations are computed according to the number of their past contributions, the quality of produced articles and the quantity of succeeding edits (see [22]). Reputation systems are considered one of the primary factors for success of online communities [18]. By exploring German Wikipedia, [22] showed that high quality articles are not necessarily written by a huge number of people, but the most important is to be written by contributors with reputation for high quality contributions. On the other hand, [23] found that the highest quality contributions come from the vast numbers of anonymous who contribute infrequently. So we must find an adjustment between user with high level reputation and anonymous users.

H3: "High-quality content in Wikipedia comes from means level of reputation distributed among members during the teaming period. In addition there is a non linear relationship between reputation diversity and article quality. Article quality increases as reputation diversity increases. However, beyond certain levels of reputation diversity, article quality will decrease."

4 Research Method

4.1 Wikipedia case study

To answer our research questions and to verify the validity of our hypotheses, we chose Wikipedia as data source. Wikipedia, is one of the most heralded success stories of peer collaboration and has become a notable example of the online epistemic community.
The first important characteristic of Wikipedia data availability. Many useful data records publicly available online, includes useful information about Wikipedia [24]. Furthermore, the case study will be Wikipedia, because of its connection with firms' knowledge management and production challenges. Another important aspect of Wikipedia is that MediaWiki site is maintained for every different language in Wikipedia.

4.2 Data collection

The best way to retrieve large portions or the whole set of activity data from any Wikipedia language is using the database dump files. These dump files contain precise information about all actions performed in any Wikipedia language. Dump files can be retrieved from the Wikimedia Downloads center.

For our purposes, we are interested in French wikipedia. We first retrieve French Wikipedia XML database dump le "pages-meta-history.xml.7z" from the set of available dump files. We use data extracted on December 12, 2015. This data contains the complete meta data of every version of all French articles from the beginning of the online encyclopedia (January 2001) to December 2015.

Once dump file is loaded, we use WikiDAT and Media wiki API for data extraction. WikiDAT is a tool for Wikipedia data analytics, based on Python and R and using MySQL database. It is aim to create an extensible toolkit for Wikipedia using Python and R to automate the extraction of Wikipedia data into 5 different tables of MySQL database (page, people, revision, revision hash, logging).

The MediaWiki action API is a web service that allows the collection of data and meta-data from the latest Wikipedia's dump and it's available for several languages, in particular, French, the one we're concerned with. It is a project maintained by the Mediawiki and contains a well structured documentation to be able to query data and can be returned as JSON. In our project, we used it to retrieve the articles that are part in the categories "Featured Articles" and "Good Articles".

4.3 Data preparation

4.3.1 Data selection

We preprocessed the XML records in the raw data using Wiki-DAT into a tabular data set representing 7833289 articles and 114907858 edits. We used MediaWiki API to retrieve article quality; a qualified article needs to belongs to one of two classes: "Featured Articles" and "Good Articles". There are 2920 qualified article (1% of total number) and we randomly sampled 2500 non qualified articles. We finally analyzed these articles based on their revision and the historic of their members' revision on other articles. As a first contribution we start by analyzing 100 articles from 114907878 editors' revision made.

4.3.2 Variable measure

In this part, Table 1 describes the different variables we consider for modeling our hypothesis.

4.4 Data analysis

In this paper, we examined WikiProjects from 2001 to 2015 in order to understand how group composition and diversity characteristics affect the quality of created knowledge. For data analysis, we used Random Forest in order to select the most relevant group attributes leading to successful articles. We created a predictive model of random forest algorithm.

Table 1. *Variables measure*

Independent Variables		
Variables		**Metrics**
Individual members' characteristics	# contribution	The number of edits made by each editors beyond current article.
	length contribution	The length of edits made by each editors beyond current article.
	# participation	The number of other article edited by each editors.
	# Efficient Contribution	The number of previous edits made by each editors in qualified articles(FA or GA) and beyond current article.
	Length Efficient Contribution	The length of previous edits made by each editors in qualified articles (FA or GA) beyond current article.
	# Efficient Participation	The number of qualified articles (FA or GA) edited by each editors.
	# Experience (how long the editor had been a member of	The number of days elapsed from a member's first edit in Wikipedia to his last revision [25]
	Reputation	we compute the reputation of authors based on their contribution to excellent pages. We then compute the rating of a page based on the reputation of the contributing authors [22]: $Reputation_{eb} = \frac{efficient_{contribution}}{Total_{contribution}}$ $Reputation_{pb} = \frac{efficient_{participation}}{Total_{participation}}$
Group members' characteristics	Group size	The number of unique editors who have contributed to the article during the teaming period (4 months after article inception) [26] [20] .
	Maximum individual charactersitics	Maximum of members individual characteristics in each Wikipedia project.
	Minimum individual charactersitics	Minimum of members individual characteristics in each Wikipedia project.
	Average individual charactersitics	Average of members individual characteristics in each Wikipedia project
	Diversity	Coefficient of variation of all members individual attribute in each Wikipedia project [27]: $diversity = \frac{\sqrt{attribute_i - attribute_{mean}}}{attribute_{mean}}$
Dependent Variables		
Article Quality		Wikipedia's internal quality categorization schema which assigns articles to a set of 7 distinct categories (Stub-> FA). We parse variables on article quality using the MediaWiki API. We used it to retrieve which articles are part of the categories "Featured Articles" and "Good Articles" [28]. For the quality variable, we use Wikipedia's own quality grading as our metric and we classify article on «Qualify Articles»(FA and GA) and «Non Qualify Article» (the rest).

This algorithm is a method for classification that operates by constructing a multitude of decision trees during training time, and outputting ordered attributes by importance [29].

For our predictive model, we separated the dataset in train and test sets. The train set consisted in a random 70% of all the articles and the test set contained the remaining 30%. Then, the predictive model was trained on the train set and applied on the test set to predict article quality. We compared those predictions to the real value of article quality. To deduce the accuracy value of each predictive model we compute the ratio of the number of good predictions over the number of predictions. This process was done a 5000 times to smooth over the extreme cases.

5 Results

Descriptive statistics are presented in Fig.3 for French Wikipedia. The random forest method gives values for quantifying the importance of an attribute for the quality of the prediction. The variable importance plot is a critical output of the random forest algorithm. For each variable in your matrix it tells how important that variable is in classifying the data. The plot shows each variable on the y-axis, and their importance on the x-axis. They are ordered top-to-bottom as most important to least important. Therefore, the most important variables are at the top and an estimate of their importance is given by the position of the dot on the x-axis. It computes the average decrease of accuracy of each tree in the forest when a given attribute is not used. Higher this value is, more important is this attribute for the prediction. To decide how many important variables to choose, we should look for a large break between variables. According to this metric Fig.3 shows that the large break is between max_Nb_participation and diversity_participation_FA.

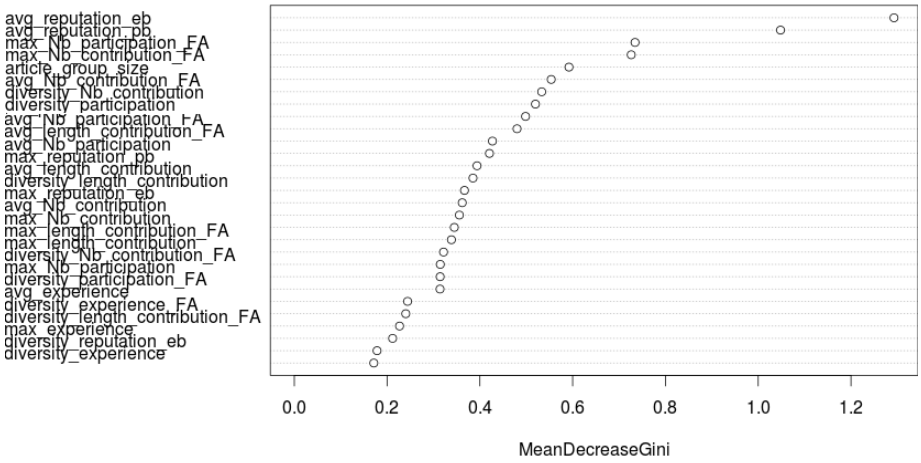

Fig. 3. *Attributes' importance of French Wikipedia*

In our model there are many attributes and some of them may be useless, being very correlated to others. Hence, we calculated the Pearson correlation on all the attribute's pairs as mentioned in Fig. 4 and we removed one element of every pair for which the absolute value of the correlation is over 0.8. So the most important attributes in general and that standing out to predict a successful French Wikipedia project are average reputation, group size, diversity contribution and participation, average experience.

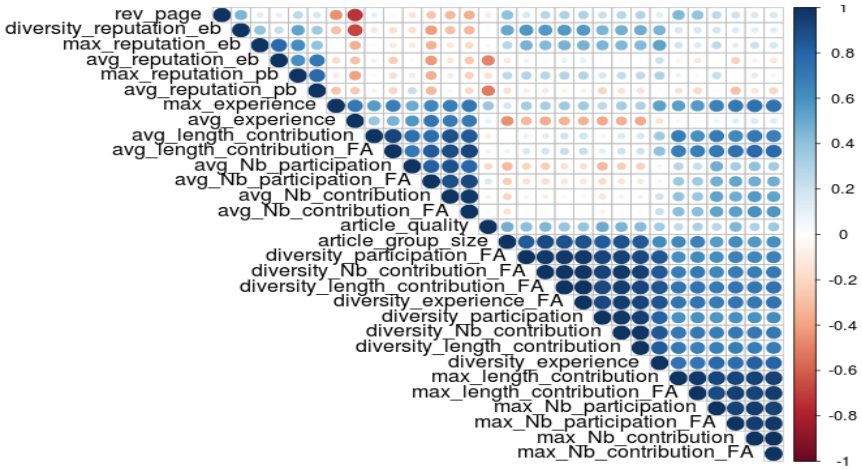

Fig. 4. *Pearson correlation on all the attribute's pairs of French Wikipedia*

6 Discussion

The results shown in section 5 indicate that the most important variable is average reputation. This means that the average reputation of the authors who edit qualified articles is higher than the reputation of authors contributing to other articles. Similarly, the editors who wrote the qualified articles participate during teaming periods of qualified ones. In the same time, average reputation is more important than maximum reputation. More specifically, average reputation is more important than reputation diversity which doesn't exert any significant importance to predict qualified articles. At the beginning of the core member recruitment process, recruiting editors with high or heterogeneous reputation is not necessary. But the most important is to recruit well-known editors based either on their edits or on the pages they edited.

The results regarding the effects of experience stand in contrast to the findings of [9] which posit that high experience or tenure disparity leads to high productivity. Experience diversity does not have any significant impact article quality during teaming period. This means that core members do not need to have a diverse experience to produce a qualified article. In contrast, average experience distributed among core members have a slight importance article quality.

Experience in Wikipedia is sometimes viewed as conferring social status. Old editors and newcomers may refer disparagingly to each other, may cause conflict in a WikiProject that reduces performance. In particular conflicts at the beginning of article creation, and during the teaming period, can be particularly damaging to online volunteer groups. As a result, when members get frustrated, they are more likely to leave or stop contributing to group effort. High tenure diversity of core member increasing conflict is consistent with prior research on offine groups. But, status inequalities during the teaming period can be less salient in online groups. This may be particularly true in online volunteer groups like Wikipedia where most editors participate at the beginning on equal experience, without much difference in privilege or rank.

Likewise, our findings suggest that people in online volunteer groups still categorize their peers based on experience and treat them differently. But in a first step of teaming period, preferably having equal experience distributed among members lead to a high quality article.

Interestingly, we found that while diversity in experience and reputation does not has a significant impact on article quality, the number of contributions, and their length, matter, in a single sense: the more diverse, the longer, the better. Our results show that group size is also an important attribute to qualify pro-duced article. This can be understood as the fact that the team which rapidly attracts contributors (within this 4-months period) have a better chance of success, something to be related to the fact that a well-know indicator of article quality is the number of contributors who have participated in their redaction. This is coherent with the finding related to the control variable. We found that article age exerted a positive impact on all constructs by generating the increase of teaming period on all article. The construction of an article of quality is, mainly a question of stock accumulation (here edits), and the longer the period is, the better the chance that new edits have been made.

7 Conclusion and Future Works

The primary contribution of our work is to develop a set of theoretical propositions about structure involvement in term of group composition in a virtual team. We further integrate various social theories to develop these propositions by considering how structure can be instantiated in shared mental models and the specific behaviors that contribute to building such models.

First, we show the importance of studying group composition in online open collaboration. Although most existing research on online collaboration has focused on motivation, governance, and social structure, our results suggest that the attributes of group members are another important factor that infuences the success of these groups. Our findings, on one hand, confirm the importance of diversity in online collaboration and, on the other hand, suggest that the impact of diversity depends upon member attributes and the degree to which an attribute is accessible and salient online.

As a second step in our model, we will work on the manner members organize their activities. We will study in particular the different kinds of leadership and their effect on

article quality. Because there are interesting suggestions that can be gleaned from the nascent literature on leadership in virtual teams, we will present three order leadership that seems likely to be more e ective.

Reference

1. Patrick Cohendet, Frederic Creplet, and Olivier Dupou•et. Communities of practice and epistemic communities: a renewed approach of organisational learning within the rm. In Workshop on Economics and Heterogeneous Interacting Agents, 2001.
2. Francesco Rullani and Stefan Hae iger. The periphery on stage: The intra-organizational dynamics in online communities of creation. Research Policy, 42(4):941{953, 2013.
3. Je rey A Roberts, Il-Horn Hann, and Sandra A Slaughter. Understanding the motivations, participation, and performance of open source software de-velopers: A longitudinal study of the apache projects. Management science, 52(7):984{999, 2006.
4. Brian Butler, Elisabeth Joyce, and Jacqueline Pike. Don't look now, but we've created a bureaucracy: the nature and roles of policies and rules in wikipedia. In Proceedings of the SIGCHI conference on human factors in computing systems, pages 1101 ACM, 2008.
5. Katherine J Stewart and Sanjay Gosain. The impact of ideology on e ec-tiveness in open source software development teams. Mis Quarterly, pages 291{314, 2006.
6. Sherae L Daniel and E Ilana Diamant. Network e ects in oss development: The impact of users and developers on project performance. ICIS 2008 Proceedings, page 122, 2008.
7. Rajdeep Grewal, Gary L Lilien, and Girish Mallapragada. Location, loca-tion, location: How network embeddedness a ects project success in open source systems. Management Science, 52(7):1043{1056, 2006.
8. Richard P Bagozzi and Utpal M Dholakia. Open source software user com-munities: A study of participation in linux user groups. Management sci-ence, 52(7):1099{1115, 2006.
9. Yuqing Ren, Jilin Chen, and John Riedl. The impact and evolution of group diversity in online open collaboration. Management Science, 2015.
10. Charlotte Hess and Elinor Ostrom. Understanding Knowledge as a Com-mons. From Theory to Practice. MIT Press, 2006.
11. Sujin K Horwitz and Irwin B Horwitz. The e ects of team diversity on team outcomes: A meta-analytic review of team demography. Journal of management, 33(6):987{1015, 2007.
12. Yong Tan, Vijay Mookerjee, and Param Singh. Social capital, structural holes and team composition: Collaborative networks of the open source software community. ICIS 2007 Proceedings, page 155, 2007.
13. Ofer Arazy, Oded Nov, Raymond Patterson, and Lisa Yeo. Information quality in wikipedia: The e ects of group composition and task con ict. Journal of Management Information Systems, 27(4):71{98, 2011.
14. John Mathieu, M Travis Maynard, Tammy Rapp, and Lucy Gilson. Team e ectiveness 1997-2007: A review of recent advancements and a glimpse into the future. Journal of management, 34(3):410{476, 2008.
15. RL Moreland and JM Levine. Group composition: Explaining similarities and di erences among group members. The Sage handbook of social psy-chology, pages 367{380, 2003.

16. Richard L Moreland, JM Levine, and ML Wingert. Creating the ideal group: Composition e ects at work. Understanding group behavior, 2:11{35, 2013.
17. Kevin Crowston, Kangning Wei, Qing Li, and James Howison. Core and periphery in free/libre and open source software team communications. In System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on, volume 6, pages 118a{118a. IEEE, 2006.
18. Amiangshu Bosu and Je rey C Carver. Impact of developer reputation on code review outcomes in oss projects: An empirical investigation. In Proceedings of the 8th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, page 33. ACM, 2014.
19. James Surowiecki. The wisdom of crowds. Anchor, 2005.
20. Wu Yang, Wei Liu, Andres Vi~na, Mao-Ning Tuanmu, Guangming He, Thomas Dietz, and Jianguo Liu. Nonlinear e ects of group size on col-lective action and resource outcomes. Proceedings of the National Academy of Sciences, 110(27):10916{10921, 2013.
21. Loxley Sijia Wang, Jilin Chen, Yuqing Ren, and John Riedl. Searching for the goldilocks zone: trade-o s in managing online volunteer groups. In Pro-ceedings of the ACM 2012 conference on Computer Supported Cooperative Work, pages 989{998. ACM, 2012.
22. Klaus Stein and Claudia Hess. Does it matter who contributes: a study on featured articles in the german wikipedia. In Proceedings of the eighteenth conference on Hypertext and hypermedia, pages 171{174. ACM, 2007.
23. Denise Anthony, Sean W Smith, and Timothy Williamson. Reputation and reliability in collective goods the case of the online encyclopedia wikipedia. Rationality and Society, 21(3):283{306, 2009.
24. Felipe Ortega. Wikipedia: A quantitative analysis. PhD thesis, PhD thesis. Universidad Rey Juan Carlos, Madrid,.(Cit. on p.), 2009.
25. Jilin Chen, Yuqing Ren, and John Riedl. The e ects of diversity on group productivity and member withdrawal in online volunteer groups. In Proceed-ings of the SIGCHI Conference on Human Factors in Computing Systems, pages 821{830. ACM, 2010.
26. Xiaoquan Michael Zhang and Feng Zhu. Group size and incentives to con-tribute: A natural experiment at chinese wikipedia. American Economic Review, Forthcoming, pages 07{22, 2010.
27. Arthur G Bedeian and Kevin W Mossholder. On the use of the coe cient of variation as a measure of diversity. Organizational Research Methods, 3(3):285{297, 2000.
28. Aleksi Aaltonen and Stephan Seiler. Cumulative growth in user-generated content production: Evidence from wikipedia. Management Science, 2015.
29. Leo Breiman. Random forests. Machine learning, 45(1):5{32, 2001.