Identifying Shared Leadership in Wikipedia

Haiyi Zhu, Robert Kraut, Yi-Chia Wang, Aniket Kittur
Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213
{haiyiz, robert.kraut, yichiaw, nkittur}@cs.cmu.edu

ABSTRACT
In this paper, we introduce a method to measure shared leadership in Wikipedia as a step in developing a new model of online leadership. We show that editors with varying degrees of engagement and from peripheral as well as central roles all act like leaders, but that core and peripheral editors show different profiles of leadership behavior. Specifically, we developed machine learning models to automatically identify four types of leadership behaviors from 4 million messages sent between Wikipedia editors. We found strong evidence of shared leadership in Wikipedia, with editors in peripheral roles producing a large proportion of leadership behaviors.

Author Keywords
Shared leadership, Wikipedia, Applied machine learning

ACM Classification Keywords
H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces.

General Terms
Measurement.

INTRODUCTION

Leadership is a central concept in organizational theory and remains a key issue in organizational life. Good leaders are important for successful groups in conventional organizations [2]. Who are the leaders in online organizations? Many online production communities, such as Apache, Linux, and Wikipedia, have a small core of participants who do the lion's share of the production work [4][6]. Are they also the leaders? It is plausible that in online environments that value broad participation, leadership may involve many more people, rather than being concentrated in a small set of formal leaders. For example, Forte and Bruckman's research reveals a trend of increasing decentralization in Wikipedia governance [4]

In this paper, we introduce a model of "shared leadership" [8] to further understand leadership in online production communities. We conjecture that leadership in online production communities is a collective activity; leadership behaviors are performed by members at all levels, from peripheral as well as central roles in online production communities.

Leadership in conventional organizations is often defined behaviorally, in terms of the functions that formal leaders serve in helping a group accomplish its tasks (task-focused) and facilitate team interaction or development (person-focus) [2]. These leadership behaviors include supervising, rewarding, warning and socializing workers. To understand how leadership is accomplished in online groups, one needs a way to measure leadership behaviors.

This paper describes the development and validation of tools for the automated measurement of these four types of shared leadership behavior in Wikipedia, one of the most popular online production communities. With these tools we quantify leadership in Wikipedia, showing evidence that it is highly distributed among members.

Specifically, we examined leadership behavior by coding messages Wikipedia editors left on each other's personal profile pages. We used machine learning to classify messages into four categories of leadership behaviors: providing positive feedback, providing negative feedback, directing someone to work on a particular task, and exchanging social information. These four categories can be detected with high accuracy (89% on average), by using standard text-classification algorithms operating on 21 features designed from domain knowledge about how Wikipedia operates. The agreement between the machine classifications and human judgments (Kappa = 0.7) is quite close to inter-human agreement (Kappa = 0.8).

Then we applied the classification model to approximately 4 million messages between Wikipedia editors. The results suggest that leadership in Wikipedia is distributed among editors with different levels of engagement: much leadership behavior is done by regular editors who are not in more formal leadership roles (e.g. administrators) in Wikipedia. Furthermore, we identify some differences in leadership styles for editors with differing involvement.

The paper lays a foundation for future research to quantitatively study shared leadership (such as the impact of leadership behaviors on the success of articles and

projects) in Wikipedia and other online production communities. Our research also demonstrates how appropriately incorporating domain knowledge into machine learning algorithms can enhance researchers' ability to process large-scale data sets.

SHARED LEADERSHIP

The concept of shared leadership, which separates leadership behavior from explicit leadership roles, was developed in the mid-1990s in response to the increasing use of self-managed teams in conventional organizations [9]. Unlike vertical leadership in a hierarchical managerial system, shared leadership is defined as "a dynamic, interactive influence process among individuals in groups for which the objective is to lead one another to the achievement of group or organizational goals" [8]. Pearce et al. [8] summarize three main characteristics of shared leadership as: 1) distributed and interdependent among people at all levels; 2) a social process embedded in the social context in which it occurs; and 3) focusing on the particular social interactions which lead to mutual learning, greater shared understanding, and eventually positive actions.

Four Types of Leadership Behaviors

In this paper, we measure leadership by examining in the messages sent between editors the four specific behaviors derived from Pearce et al.'s theoretical model of shared leadership [9]: providing positive feedback, providing negative feedback, directing, and social exchange. Definitions and example messages are shown in Table 1.

AUTOMATED MEASUREMENT

Without automated coding of behavior, research on leadership in online communities is restricted to small samples. For example, Choi et al [3] hand coded communication among approximately 300 individuals in 12 Wikipedia projects, even though there are approximately 40,000 members of the 2392 Wikiprojects. In this paper, we demonstrate the possibilities of going beyond these small samples by using automated coding of leadership behaviors.

Leadership Behaviors	Leadership Type and Message Example
Providing Positive Feedback (Positive)	Transactional Leadership: Energize people through acknowledging work and provides rewards
	"I'm so impressed. This is a very fine article!"
Providing Negative Feedback (Negative)	Aversive Leadership: Regulate people through reprimands.
	"If you continue in this manner you will be blocked from editing without further warning. Please stop, and consider improving rather than damaging the work of others."
Directing (Directive)	Directive Leadership: Direct people through issuing instructions, commands, assigning tasks, setting goals.
	"Here is a new article on a former airport I thought you might want to check out."
Social exchange (Social)	Transformational Leadership: Promote emotional engagement through for example talking nice, starting off-topic conversation, etc.
	"Drop me a line on my talk page sometime, we'll get a coffee over at Hot Rize or the new King Kocoa…"

Table 1. Four leadership behaviors and their corresponding leadership types and example messages

Classification Approach

We used a machine learning approach to classify messages into types of leadership behaviors. Specifically, we trained statistical models on a small set of human-coded data and evaluated it using a separate set of human-coded data. Then we applied the model to a larger data set that had not been human coded. A machine learning approach has three main components - training sets, representation of messages for machine learners (feature sets), and training algorithms, which we explain in more detail in the following sections.

Training Sets

We hand coded 500 messages into each of the four leadership behaviors defined in Table 1 to provide training data for the model. Messages could be assigned to multiple categories if they exhibited more than one leadership behavior. To assess the reliability of the coding, two human judges annotated 100 messages. The Cohen's Kappa measure of inter-judge agreement averaged across the four categories was 0.82 (positive 0.81, negative 0.80, directive 0.79, social 0.88), which is very high [11].

Representation of Messages (Feature Sets)

We first considered surface text features, such as unigrams (individual words), bigrams, and part-of-speech unigrams (unigrams tagged with part of speech information) extracted using TagHelper [10]. With Support Vector Machines (SVM) as our baseline learning method, we tested different kinds of surface text features. The results were acceptable, with prediction accuracies around 80%.

We then used features based on domain knowledge, realizing that message senders tend to frequently use certain words and phrase patterns to express different intents. We identified 21 domain knowledge features in five categories[1]:

- *Strong/weak, positive/negative polarity words.* Four features based on the combination of strength and polarity derived from the subjectivity lexicon of OpinionFinder [11].
- *Strong positive adjectives.* Seventeen strong positive adjectives used in praise, such as "excellent", "great", and "impressive".
- *Negation.* Seventeen negation words and phrases (e.g., "not", "shouldn't", "doesn't").
- *Negative jargon.* Nineteen Wikipedia-specific negative words such "vandalism" and "blocked". *Causative/subjunctive verbs.* Twenty-seven causative or

[1] For exact vocabulary used for all features, please refer to http://www.cs.cmu.edu/~haiyiz/FeaturesVocabulary.htm

subjunctive verbs including "make", "suggest", *"recommend", "wish" and* "need".

- *<You+modal>*. Sentences starting with a pronoun "you" immediately followed by a modal word (e.g., "should", "might", "must") or vice versa.
- *Acknowledgements.* Phrase patterns of "thank you/thanks for".
- *Smiley.* Textual expressions such as :), ;).
- *Greetings.* Greeting words/phrases, such as "hello", "congratulations", and "happy birthday".
- *He/she.* Number of "he, him, his, she, her".
- *Length.* Number of word tokens in a message.
- *Variants of the following words/phrases was included as a separate feature:* "if you", "newsletter", "Wikiproject", "congrats", "welcome", and "please"+ verb.

Learning Algorithms

Since the four categories are not exclusive, we formulated the classification task as four binary decision problems. We conducted experiments on the training set with various learning algorithms implemented in Weka [5], including decision trees, Adaboost on decision trees, Naïve Bayes, and linear SVM. We chose to use linear SVM because it worked consistently well on all four categories[2].

Classification Results

We report results of ten-fold cross-validation of the trained model in Table 2[3].

	Positive	Negative	Directive	Social
Acc.	0.91	0.87	0.86	0.92
Kappa	0.75	0.48	0.71	0.80

Table 2. Training results of ten-fold cross-validation using the SVM algorithm and 21 domain knowledge features

Kappa, which represents agreement between machine and human judges, is moderate for negative feedback, but is very substantial or excellent for the other three categories [11]. The SVM classifiers calculate the weighted sum of feature counts plus intercept for each message, then determines whether the message belongs to the category depending on whether the sum is larger than 0 or not. Table 3 reports features with large weights (>0.1 or <-0.1).

	Positive	Negative	Directive	Social
Feature (weight)	*Barnstar* (+2.00) *Acknowledge* (+0.67)	*Negative Jargon* (+0.29) *<Please+ verb>* (+0.11) *Barnstar* (-0.12)	*<You+ modal>* (+2.00) *<Please+ verb>* (+1.80) *<If+you>* (+0.10)	*Greeting* (+2.00) *Smiley* (+1.00) *Congrats* (+0.33) *Barnstar* (-0.25)
Intercept	-1.00	-1.00	-1.00	-1.00

Table 3. Features with large weight in the SVM Models

[2] The Adaboost model worked slightly better on negative feedback category, but was outperformed by SVM on the other categories.

[3] We also applied our models on another 100 annotated messages to assess whether the models overfit the training set. The result on this extra test set was quite good, with an average kappa of 0.78.

APPLICATION OF THE MEASUREMENT

We then applied the SVM model to Wikipedia data to examine who produces leadership behaviors. We used a complete download provided by the Wikimedia Foundation from Wikipedia's inception to January 2008 (approximately 182 million revisions). To handle this data volume, we used the Yahoo! M45 computing cluster running Hadoop. We identified 4,014,925 messages posted by users to other users' talk pages (we excluded replies and messages sent by bots). The approximately 4 million messages were sent by 130,000 distinct users (who had edited Wikipedia for an average of 13.6 months) and were received by 1.1 million distinct users (who averaged 10.8 months of editing). We compared leadership behaviors enacted by editors in different roles from peripheral participants to central ones.

Roles in Wikipedia

Bryant et al.'s research [1] and our own observations suggest several observable behaviors that allow one to distinguish peripheral participants in Wikipedia from more central ones. We illustrate two possible dimensions in Figure 1.

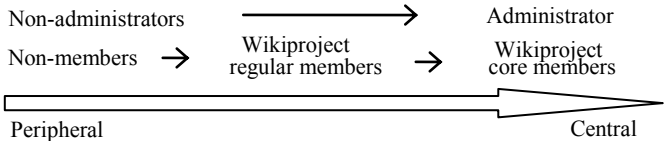

Figure 1. Peripheral and central roles in Wikipedia

Administrators versus non-administrators. Editors' privileges are one way to distinguish peripheral from central roles. Administrators have successfully passed a rigorous peer review process and are given special privileges, such as blocking malicious editors.

Wikiproject core members, Wikiproject regular members, and non-members. Participation in community subgroups is another measure of centrality. Wikiprojects are subgroups of editors organized around a topic, such as Wikiproject Medicine. We define Wikiproject core members as those who either founded a project or are one of its top three contributors, as measured by edits to the project pages. Regular members are those who joined at least one project but are not core members. Non-members are editors who contribute to Wikipedia but never joined a Wikiproject.

Who produces leadership behaviors

Table 4 compares the degree to which editors in different roles produce the four types of leadership behaviors. Interestingly, although people in central positions such as administrators and Wikiproject members send many more leadership-related messages per person than others, non-central editors still contribute a large proportion of the aggregated total leadership behavior (e.g., 1.5 million directive messages sent by non-admins vs. 0.8 million sent by admins). This means that if researchers only examine directive leadership behaviors from central figures such as administrators, almost two-thirds of directive behaviors will

	Admin Vs Non-admin		Wikiproject member Vs Non-member	
Num of People	1723	131,848	25019	108552
Positive				
Per person	154.7	4.3	28.9	1.0
Aggregated	267K	569K	723K	113K
Negative				
Per person	155.9	3.9	27.8	0.8
Aggregated	269K	509K	696K	81K
Directive				
Per person	483.8	11.4	78.0	3.1
Aggregated	834K	1503K	2001K	336K
Social				
Per person	244.0	4.6	32.7	1.6
Aggregated	386K	602K	819K	170K
Overall Wikipedia activities per person	16977.7	573.7	3387.7	185.6

Table 4. Comparing number of people, aggregated/per person leadership messages, overall Wikipedia activities for editors in different roles

be ignored (64.3%). Furthermore, when we investigate leadership behaviors as the proportion of overall activity (calculated as the total number of edits in any Wikipedia page), we find that the more central editors perform more leadership behaviors per person because they are generally more active. However, these differences are not huge. For example, 2.8% of administrators' work consists of sending directive messages compared to 2.0% for non-administrators. This slight difference calls into question traditional models of leadership, where leaders devote a much larger percentage of their time to directing and supervising others than non-leaders [7], and suggests a shared leadership model in which leadership activity is distributed among contributors in different roles. In other words, people in central roles are not exclusively dedicated to "leading" nor are people in non-central roles merely acting as "workers".

Although editors in different roles all perform leadership behaviors, there are some interesting differences in leadership styles for different editor types. When we focus on Wikiproject members, we see that non-core members (regular members) are more likely to perform task-oriented leadership behaviors (providing positive and negative feedback and directing), while founders and other core members are more likely to engage in socially-oriented leadership (Table 5). This suggests that the role of core members in Wikiprojects may be less task-focused and more person-focused, with social or motivational messages to keep members active. Further work is needed to understand differences between leadership styles and editor types, as well as the effects they have on those receiving messages; our model may serve as a tool for enabling such future research at a large scale.

	Positive	Negative	Directive	Social
Wikiproject core members	18.0%	18.2%	57.3%	28.2%
Wikiproject regular members	23.9%	22.5%	62.3%	23.2%

Table 5. Percentage of each type of leadership messages among all the messages sent by Wikiproject core members and regular members

CONCLUSION & DISCUSSION

In this paper we developed a machine learning model to automatically identify four types of leadership behaviors from messages sent between Wikipedia editors. We found strong evidence of shared leadership in Wikipedia, showing that a large proportion of leadership behaviors are performed by editors in peripheral as well as central roles. These results, however, represent only one way of operationalizing leadership behaviors, and we hope to spur an evidence-based dialog about what behaviors and roles constitute leadership in online communities. Our results have practical applications for researchers interested in automatically identifying leadership behaviors and quantitatively studying their effects in social media.

ACKNOWLEDGMENTS

This work was supported by NSF grants IIS-0968484 and OCI-0943148.

REFERENCES

1. Bryant, S. L., Forte, A., and Bruckman, A. Becoming Wikipedian: transformation of participation in a collaborative online encyclopedia. In *GROUP '2005,* ACM Press (2005).
2. Burke, C., Stagl, K., Klein, C., Goodwin, G., Salas, E., and Halpin, S. What type of leadership behaviors are functional in teams? A meta-analysis. *The Leadership Quarterly*, 17(3). (2006).
3. Choi, B.R., Alexander, K., Kraut, R.E. and Levine, J.M. (2010). Socialization Tactics in Wikipedia and Their Effects. In *CSCW'2010*, ACM Press (2010).
4. Forte, A., and Bruckman, A. Scaling Consensus: Increasing Decentralization in Wikipedia Governance. *Proceedings of HICCS*. (2008).
5. Hall, M., Frank E., Holmes G., Pfahringer B., Reutemann P., and Witten I.H. The WEKA Data Mining Software: An Update; *SIGKDD Explorations*, 11(1) (2009).
6. Kittur, A., Chi, E., Pendleton, B.A., Suh, B. and Mytkowicz, T. Power of the Few vs. Wisdom of the Crowd: Wikipedia and the Rise of the Bourgeoisie. In *CHI' 2008.* ACM Press (2008)
7. Mintzberg, H. The Manager's Job: Folklore and Fact. *Harvard Business Review*, 53 (4). 49-61 (1990).
8. Pearce, C.L., and Conger, J.A. Shared Leadership: Reframing the Hows and Whys of Leadership. SAGE Publications, Inc, 2002.
9. Pearce. C.L., and Sims H.P. Vertical Versus Shared Leadership as Predictors of the Effectiveness of Change Management Teams: An Examination of Aversive, Directive, Transactional, Transformational, and Empowering Leader Behaviors, *Group Dynamics*, 6(2) (2002).
10. Rose, C. P., Wang, Y.C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., and Fischer, F. *International Journal of Computer Supported Collaborative Learning*. 3(3) (2008).
11. Stemler. S. An overview of content analysis. *Practical Assessment, Research & Evaluation*, 7(17) (2001).
12. Wilson T., Wiebe J., and Hoffmann P. Recognizing Contextual Polarity: An Exploration of Features for Phrase-Level Sentiment Analysis, *Computational Linguistics* 35(3) (2009).