The Behavior and Network Position of Peer Production Founders

Jeremy Foote(✉) and Noshir Contractor

Northwestern University, Evanston, IL, USA
jdfoote@u.northwestern.edu

Abstract. Online peer production projects, such as Wikipedia and open-source software, have become important producers of cultural and technological goods. While much research has been done on the way that large existing projects work, little is known about how projects get started or who starts them. Nor is it clear how much influence founders have on the future trajectory of a community. We measure the behavior and social networks of 60,959 users on Wikia.com over a two month period. We compare the activity, local network positions, and global network positions of future founders and non-founders. We then explore the relationship between these measures and the relative growth of a founder's wikis. We suggest hypotheses for future research based on this exploratory analysis.

1 Introduction

The surprising success of online peer production (OPP) projects like Wikipedia and open source software has shown that groups of motivated volunteers can successfully create high-quality goods without formal hierarchy. Scholars across a number of disciplines have studied why these projects work, sparking new research on the role of firms, intellectual property rights, and individual motivations in producing shared goods [1–4]. In this project, we focus on one aspect of OPP projects that has escaped much scholarly attention: its founding.

Founders have been shown to be influential in the similar context of entrepreneurship. Researchers have found both that people differ in their propensity to become an entrepreneur [5–9] and that the attributes and experiences of founders relate to a firm's success [10–14]. There are a number of reasons to think that OPP founders differ from entrepreneurs, such as the lower costs, risks, and benefits of founding. Learning about OPP founders can help us to understand why projects grow (or don't) in this increasingly important context.

We use measures of editing behavior and measures of social capital and social integration from over 60,000 contributors to Wikia.com to explore how founders differ from non-founders as well as how founders of high-growth projects differ from founders of low-growth projects. Our exploratory results suggest that compared to non-founders, founders are typically novelty-seeking and have low social capital. However, founders who are successful at creating larger communities have a different set of attributes: they have less diverse experience but are

G. Chowdhury et al. (Eds.): iConference 2018, LNCS 10766, pp. 99–106, 2018.
https://doi.org/10.1007/978-3-319-78105-1_12

actually more integrated in social networks, suggesting that they may use their social capital to recruit others.

2 Related Work

2.1 Entrepreneurship

Researchers have studied both who decides to become an entrepreneur and what attributes and experiences of entrepreneurs correlate with firm success. They have found that people with diverse experience and skills are more likely to become entrepreneurs [5,7,15], as are those who have worked with other entrepreneurs [8]. Given that someone has chosen to start a new firm, founders with more experience [10–12] have more successful firms, as do those with larger and more diverse social networks [14].

2.2 Peer Production

At first blush, online peer production projects seem very dissimilar to firms. They are composed of volunteers, without formal hierarchies and without paychecks. Indeed, much research on OPP focuses on how these organizations can work using structures and incentives so different from firms ([16] provides a survey). Surprisingly, much of this research has found that these supposedly decentralized, leaderless organizations are actually quite structured. For example, researchers have found that people select into different social and behavioral "roles" (like jobs) that are persistent over time [17], including leadership roles [18]. In other words, OPP looks more like firms than we would expect.

We argue that at its core the decision to start a new OPP project is very similar to the decision to start a new business. There are different costs to creating each type of organization, different tools for recruiting and encouraging contributors, and different benefits that accrue to founders. However, both founders and entrepreneurs organize the efforts of a group of people to create shared outputs. Indeed, the outputs of OPP projects and firms often directly compete.

Despite the similarity in context, very little research on peer production founders exists. Survey-based research suggests that OPP founders have diverse expectations and modest goals [19]. In the related context of online communities, researchers found that active and well-connected founders started longer-lived groups [20]. This paper extends these earlier works by comparing founders and non-founders and by exploring the relationship between founder attributes and the growth of a community rather than its survival.

Based on the entrepreneurship research, we focus on two questions:

RQ1: How do founders differ from non-founders?
RQ2: How do founders of large and small communities differ?

3 Methodological Approach

Our data comes from a dump of all edits to all wikis on Wikia.com as of April 2010. While wikis are used for different purposes, many of the wikis on Wikia are used for knowledge aggregation. The most popular wikis aggregate information about popular media, such as Disney movies or the Harry Potter books. These wiki projects represent an important strand of OPP but differ markedly from other OPP projects like open source software development.

We collected data at three different points in time (Fig. 1). We gathered behavior and network measures from the 60,959 users who made at least one edit from March to April 2009. We then identified the founders of the 16,904 wikis created in May and June 2009, and measured the number of unique contributors to each of these wikis as of April 2010. We used these data snapshots to compare the activity and network positions of those who became founders in May or June and those who did not, and to predict the relative growth of a founder's wikis between founding and April 2010.

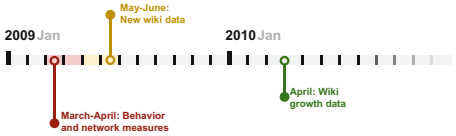

Fig. 1. Timeline of data collection

Specifically, we measured behavior and attributes which have been found to relate to founding propensity or organization success in entrepreneurship. We created a number of *experience* and *activity* measures, such as tenure on Wikia, lifetime edits, recent edits, days since editing, and number of days with at least one edit. For *diversity of experience*, we measured the number of wikis that a user had contributed to and the Gini coefficient of edits per wiki. Finally, we measured the user's *founding expertise* and experience by measuring how many wikis users had started in the past, how often they started new pages, how often they participated in administering wikis, and the earliest point at which they contributed to a wiki (e.g., a value of 3 means they were the third editor on a wiki).

In order to measure a user's *social capital*, we created two different types of social networks based on two kinds of activities that occur on Wikia. We used article pages to create undirected collaboration networks, where edges are formed between an editor and the previous five editors of that page[1]. Communication networks are similar but are directed: users are assumed to be talking *to* the previous five editors on a talk page or *to* the owner of a user talk page. We created unweighted networks for each wiki and created each unweighted global network by combining the wiki networks. For each of the two network types, we measured the degree centrality, betweenness centrality, and PageRank of each user. There is, of course, no direct way to measure social capital, but these are somewhat overlapping measures of how prestigious and involved a user is in the network [22]. We also calculated "coreness", which is a measure of how integrated

[1] Similar analyses, (e.g., [21]) connect all editors, but a cutoff more closely approximates the social interactions we are interested in.

a user is [23]. We measured these values for each user in the global networks as well as in the wiki-level network for the wiki they were most active in.

We defined founders as the first three editors to a given project, since many founders start new wikis as groups [19]. We measured growth as the number of unique contributors to the wiki from the time of its founding until April 2010. Because a given user might start multiple wikis during our data collection, we used the median community size for all foundings that a user was a part of. We also included a control for when the median wiki was started.

Because there is so little research specifically on peer production founders we took an exploratory and hypothesis-generating approach. Our predictors had a high degree of multicollinearity, so we used elastic net regularized regression to identify candidate predictors [24]. We used the smaller set of predictors produced by this procedure in regression models for each of our research questions.

4 Results

There is one interesting result which doesn't require any modeling. We learned that founding is a rare activity for established users. Of the 60,959 users who were active in March and April 2009, only 822 founded a new wiki in May or June. That is not to say that founding is rare overall. There were 8,487 founders in May and June, meaning that nearly 90% of wikis were founded by new users.

For the rest of the analyses, we remove users with 'bot' in their username (N = 104) and low-activity users who don't appear in both the communication and collaboration networks (N = 45,775). There were also Wikia employees who we could not identify automatically, so we remove all users with edits to more than 100 different wikis (N = 24). This leaves us with a sample of 15,184 users and 470 founders.

For our first research question, we try to predict which users would become founders. Figure 2 shows the results of a logistic regression with the predictors identified by the elastic net procedure. Given our earlier finding that founders are likely to be new users, it is not surprising to see that tenure has a negative relationship with founding. However, both activity (edits) and diversity of activity (wikis edited, Gini of edits per wiki) predict becoming a founder, as was found to be the case in the entrepreneurship literature. This suggests that founders are *either* new users *or* active users. Founding experience is mixed, with creating pages and founding other wikis as positive predictors but being an early participant as a negative predictor. Social capital is also mixed, with communication network indegree as a positive predictor while PageRank and integration (coreness) are negative predictors. Overall, the difference between founders and non-founders is quite pronounced along many dimensions. On the right side of Fig. 2 we show a few density plots comparing the number of edits and the number of wikis edited, respectively. In both cases, there are clear differences between founders and non-founders.

In Fig. 3 we show the results of a negative binomial regression predicting the median number of contributors that a founder's wikis received as of April

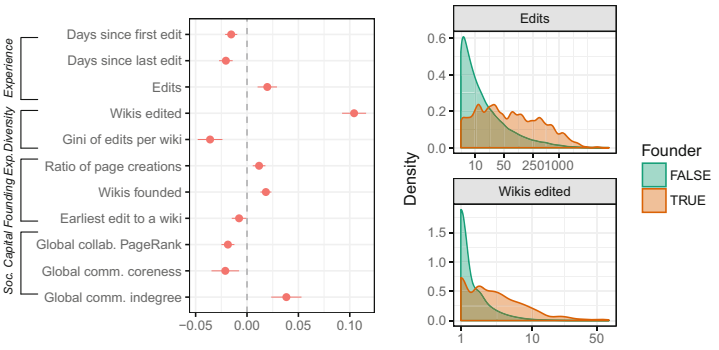

Fig. 2. Left: Scaled coefficient estimates with 95% confidence intervals for predicting of whether a user becomes a founder ($N = 15{,}184$). Right: Density plots of the number of edits and the number of wikis edited in March and April 2009 by those who become founders in May and June 2009 (blue) and those who do not (red). (Color figure online)

2010. This relationship is much noisier and much more difficult to model. The scatterplots on the right show two of the strongest predictors of growth and even these are incredibly noisy. This suggests that founder behavior and attributes do not have a strong effect on community growth. That being said, the analysis does offer a few insights. First, there is a benefit to experience, both tenure and editing activity, while experience as an admin has a negative relationship to community growth. Social capital is mixed, although most predictors show a positive relationship with growth. The median edit size of a founder's edits positively predicts growth. Finally, it is worth noting that the regularization step eliminated diversity measures from this model, suggesting that there is no significant relationship between diversity of experience and community growth in this dataset.

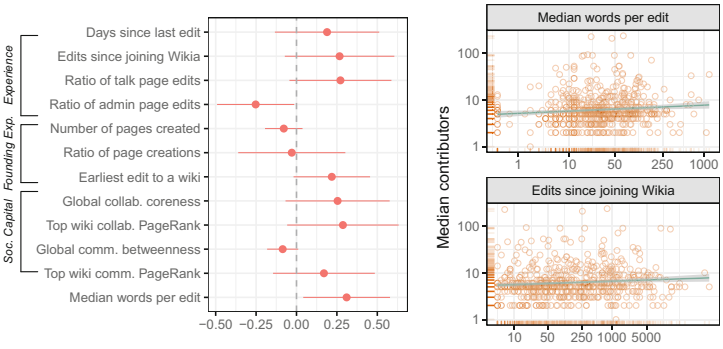

Fig. 3. Left: Scaled coefficient estimates with 95% confidence intervals, predicting the number of contributors to a community ($N = 470$). Right: Scatterplots of the median words per edit and total edits on the x axes and median community size on the y axes.

5 Discussion

We believe that the results of this exploratory study provide a foundation for future work. For example, founders in our study edited many different wikis and were more likely to start new pages. However, neither of these behaviors were good predictors of community growth. This suggests that many founders seek novelty and new experiences but then abandon their communities quickly to move on to the next new thing. Future research should look more explicitly at novelty-seeking behavior.

Our findings regarding social capital suggest additional hypotheses to be tested. We saw that integration in the global network was negatively related to becoming a founder but positively related to community growth. One explanation is that those who are on the periphery are discontent and thus more likely to start new communities but are ironically less able to grow their communities since growth requires spending social capital which they do not have.

Finally, we saw that a founder's typical edit size was the strongest predictor of growth. The reason for this relationship is not obvious, but one explanation is that people who make large edits typically add lots of information to a page and that wikis with more information are able to attract new contributors. Future work could test this hypothesis through experiments or causal inference.

6 Conclusion

There are important limitations to this work. Most obviously, we only look at founders who have previous activity on the site. Nearly 90% of the founders in our dataset were new to the site. This is consistent with research which shows that new communities can be founded as a way of learning about a site [19], but this also means that our approach cannot tell us anything about the majority of founders. Our results regarding the propensity to found and community growth only apply to existing members and the attributes and impact of new member founders are likely to differ. In addition, while we study a very large population of projects, they are all wikis on the same platform and all of the data comes only from activity on that platform. Interviews or surveys could help us to gain a richer understanding of how founding decisions are made and whether our measures accurately represent the constructs as proposed. Finally, similar analyses in contexts such as GitHub are needed to test generalizability.

Despite these limitations, we believe that this dataset and analysis provide unique insight into an important phenomenon and establish a set of intuitions for future work on peer production and public goods production to build upon.

Acknowledgments. This work was supported by the Army Research Office (W911NF-14-10686) and the NSF (IIS-1617468, IIS-1617129). The authors also wish to thank the iConference reviewers for their very helpful comments.

References

1. Benkler, Y.: The Wealth of Networks: How Social Production Transforms Markets and Freedom. Yale University Press, New Haven (2006)
2. Levine, S.S., Prietula, M.J.: Open collaboration for innovation: principles and performance. Organ. Sci. **25**(5), 1414–1433 (2013)
3. Raymond, E.S.: The Cathedral & the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary. O'Reilly Media Inc., Sebastopol (2001)
4. von Krogh, G., von Hippel, E.: The promise of research on open source software. Manag. Sci. **52**(7), 975–983 (2006)
5. Backes-Gellner, U., Moog, P.: The disposition to become an entrepreneur and the jacks-of-all-trades in social and human capital. J. Socio-Econ. **47**, 55–72 (2013)
6. Dobrev, S.D., Barnett, W.P.: Organizational roles and transition to entrepreneurship. Acad. Manag. J. **48**(3), 433–449 (2005)
7. Lazear, E.P.: Balanced skills and entrepreneurship. Am. Econ. Rev. **94**(2), 208–211 (2004)
8. Nanda, R., Sørensen, J.B.: Workplace peers and entrepreneurship. Manag. Sci. **56**(7), 1116–1126 (2010)
9. Zhao, H., Seibert, S.E., Lumpkin, G.: The relationship of personality to entrepreneurial intentions and performance: a meta-analytic review. J. Manag. **36**(2), 381–404 (2010)
10. Cassar, G.: Industry and startup experience on entrepreneur forecast performance in new firms. J. Bus. Ventur. **29**(1), 137–151 (2014)
11. Eisenhardt, K.M., Schoonhoven, C.B.: Organizational growth: linking founding team, strategy, environment, and growth among U.S. semiconductor ventures, 1978–1988. Adm. Sci. Q. **35**(3), 504–529 (1990)
12. Jo, H., Lee, J.: The relationship between an entrepreneur's background and performance in a new venture. Technovation **16**(4), 161–211 (1996)
13. Ostgaard, T.A., Birley, S.: New venture growth and personal networks. J. Bus. Res. **36**(1), 37–50 (1996)
14. Stam, W., Arzlanian, S., Elfring, T.: Social capital of entrepreneurs and small firm performance: a meta-analysis of contextual and methodological moderators. J. Bus. Ventur. **29**(1), 152–173 (2014)
15. Wagner, J.: Testing Lazear's jack-of-all-trades view of entrepreneurship with German micro data. Appl. Econ. Lett. **10**(11), 687–689 (2003)
16. Benkler, Y.: Peer production and cooperation. In: Bauer, J.M., Latzer, M. (eds.) Handbook on the Economics of the Internet. Edward Elgar, Cheltenham (2016)
17. Welser, H.T., Cosley, D., Kossinets, G., Lin, A., Dokshin, F., Gay, G., Smith, M.: Finding social roles in Wikipedia. In: Proceedings of the 2011 iConference, iConference 2011, pp. 122–129. ACM, New York (2011)
18. Zhu, H., Kraut, R., Kittur, A.: Effectiveness of shared leadership in online communities. In: Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work, CSCW 2012, pp. 407–416. ACM, New York (2012)
19. Foote, J., Gergle, D., Shaw, A.: Starting online communities: motivations and goals of wiki founders. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI 2017), pp. 6376–6380. ACM, New York (2017)
20. Kraut, R.E., Fiore, A.T.: The role of founders in building online groups. In: Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, CSCW 2014, Baltimore, Maryland, USA, pp. 722–732. ACM (2014)

21. Zhang, X., Wang, C.: Network positions and contributions to online public goods: the case of Chinese wikipedia. J. Manag. Inf. Syst. **29**(2), 11–40 (2012)
22. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge (1994)
23. Barberá, P., Wang, N., Bonneau, R., Jost, J.T., Nagler, J., Tucker, J., González-Bailón, S.: The critical periphery in the growth of social protests. Plos One **10**(11), e0143611 (2015)
24. Zou, H., Hastie, T.: Regularization and variable selection via the elastic net. J. Roy. Stat. Soc.: Ser. B (Stat. Methodol.) **67**(2), 301–320 (2005)