RESEARCH STATEMENT NATHAN TEGROTENHUIS

Introduction: Peer production is the distributed process of online collaboration behind Free/Open Source Software and Wikipedia and has been described as among the most important organizational innovations made possible by the Internet [1]. Although peer production organizations are frequently compared to traditional markets and firms, most research on peer production has looked to internal factors like users, interface design, and social networks to explain project success. Strikingly, we know that the survival of traditional organizations is significantly influenced by *environmental pressures* within populations of organizations such as legitimization and competition [2]. My goal is to study how environmental pressures explain the success and failure of peer production organizations. My research will help peer production projects, map differences between peer production and traditional organizations, and contribute to a broader understanding of collective action, organizational communication, and communal public goods.

I am particularly interested in how peer production contributors work simultaneously in multiple organizations. Although this might intensify competition between organizations for volunteers' time, it can also lead to complementarities as contributors' experience with similar projects may increase their skill or motivation. Using social scientific methods and models from organizational ecology, I will build upon a small number of previous studies in computer science to understand the distinct manifestations of competition and complementarities in peer production [3].

My research will bring methodological innovation and "big data" techniques to organization research and will use collections of digital traces from millions of individuals working across many thousands of organizations over time. Unlike previous studies of competition and complementarity, I will use these datasets to look both within and across organizations across time.

Hypotheses: I already have access to data from three of the largest and most impactful peer production community platforms: over 80,000 wikis from the wiki provider Wikia, 716,000 communities hosted within the website Reddit, and 40,000 distinct software projects hosted within the Debian project. I will test three hypotheses that describe major untested theories in organizational communication research and the study of online communities.

1. As predicted by organizational ecology and resource mobilization theory, competition over niches will deny most new projects volunteer labor and cause them to fail.
2. Complementarities between communities will lead to rapid community growth in successful organizations.
3. Unlike traditional organizations, peer production projects will change to complement and cooperate with each other, rather than compete.

To test the first hypothesis, I will first measure levels of community activity within individual communities over time to identify when a community is born or dies, use digital traces to identify individuals participating in different communities, and use machine learning analyses of communities to identify organization niches. Using these inputs, I will use survival analyses to model organization lifespans to see if niches and volunteer scarcity presents hazards to organizations. I will use hierarchical models to control for differences between organization populations.

Technical Skills and Mentoring: My mentor and advisor Dr. Benjamin Mako Hill is an expert in peer production communities, quantitative social science, and organizational communication. I am the newest member of the *Community Data Science Collective* – a multi-institutional research

group including faculty at Northwestern University which will provide learning opportunities, collaborators, and forums for sharing research. In previous work, the group has parsed datasets from each of the peer production communities I hope to study. In pilot research addressing the first hypothesis above, we have found some evidence of a positive relationship between the number of edits to the Wikia and Wikipedia encyclopedias over time.

Thanks to my professional background, I will be able to take advantage of the abundant computational resources at the University of Washington to use the very large datasets I've described. I will take advanced statistics and econometrics courses to enable me to build on these skills.

Intellectual Merit: My work will advance our understanding of *organizational communication* by adapting established theories from organizational ecology and resource mobilization to peer production. I hope to be the first to to measure, at a large scale, both the aggregate effects of ecological forces and the behavior of individuals subject to them. My project will advance understanding of social movements and volunteer labor because peer production is a remarkable instance of *collective action* [4] and will contribute to the literature on *communal public goods* by investigating the challenges of firms incentivizing their production to see if commercial or bureaucratic characteristics of the workplace necessarily contradict the incentives for peer production [5].

Broader Impacts: Peer produced projects such as Wikipedia and free/open source software are now vital technologies for the national interest and global economy. Wikipedia and Stack Exchange are important resources for STEM education. Apache and GNU/Linux technologies form the backbone of information infrastructure. I will engage in *STEM leadership* by reaching out to peer production communities in the hope of *informing decisions* that affect the directions of their organizations. My work will also be in pursuit of my personal goal to advance *quantitative and computational methods* in the social sciences. I will incorporate this research into teaching courses, volunteering at public data science workshops in Seattle, and through collaboration with other social science researchers. In the course of my work, I will *construct software* for the collection, organization and analysis of large datasets. I will publish both my software and my datasets so that other researchers and programmers can learn from and reuse my work.

Criteria For Success: My work will be successful in so far as it builds new salient theories, performs persuasive tests of existing influential theories, and develops new computational tools, datasets, and scientific methods.

REFERENCES

[1] Yochai Benkler, Aaron Shaw, and Benjamin Mako Hill. *Collective Intelligence*, chapter Peer Production: A Form of Collective Intelligence. MIT Press, 2015.

[2] Michael T. Hannan and John Freeman. The population ecology of organizations. *American journal of sociology*, pages 929–964, 1977.

[3] Haiyi Zhu, Jilin Chen, Tara Matthews, Aditya Pal, Hernan Badenes, and Robert E. Kraut. Selecting an effective niche: an ecological view of the success of online communities. In *Studying Online Communities*, pages 301–310. ACM Press, 2014.

[4] Mancur Olson. *The logic of collective action*. Harvard University Press, 1965.

[5] Janet Fulk, Andrew J. Flanagin, Michael E. Kalman, Peter R. Monge, and Timothy Ryan. Connective and communal public goods in interactive communication systems. *Communication Theory*, 6(1):60–87, 1996.