Over the last three decades, new communication technologies have ushered in transformative new ways of producing, organizing, and distributing digital public goods. These goods include critical Internet infrastructure (free/libre open source software or FLOSS), the world's fifth most visited website (Wikipedia), and the databases which provide smart assistants with answers to our questions (the Wikidata database). It also includes collaborative filtering systems which help millions connect with each other and identify relevant and interesting content (Reddit), as well as databases with widespread business, industrial, and humanitarian applications (OpenStreetMap). What all these information goods have in common is that they are produced collaboratively and openly by *peer production*—a term used to describe the technologically mediated mass aggregation of many small contributions from diversely motivated individuals into freely available commons.

For many, peer production's enormous success has been hard to square with its reliance on self-organization, porous organizational boundaries, and volunteer labor. My scholarship has focused on answering fundamental theoretical and empirical puzzles about the communicative and social dynamics that drive participation in the communities that produce these novel goods. Why do some attempts at peer production succeed while most fail? Why is sustainability so difficult to achieve for the rare projects that overcome these odds? My body of research has provided answers to these questions through an organization-level and lifecycle-based approach and a focus on understanding tradeoffs. In the process, my work has challenged ideas taken for granted in earlier peer production scholarship. As my research seeks to understand open and collaborative collaboration, I have attempted to infuse my own work with collaborativeness and openness. My teaching and service are both very closely aligned with my research program.

RESEARCH

My research has two major focuses. First, I conduct empirical research—most frequently using computational methods and large datasets of digital traces—in order to build a theoretical and empirical understanding of how social and communicative processes shape peer production. Although the vast majority of other peer production researchers have sought to understand the experience of individual contributors within communities, my work engages with peer production at the level of communities, projects, and populations. In the second major focus of my work, I have sought to make the type of data science studies at the heart of my research program easier to conduct—by scholars, by online community members, and by the general public.

As a computational social scientist, my work lies at the intersection of communication and computing. I have adjunct (courtesy) appointments in UW's Department of Human-Centered Engineering as well as in the Paul G. Allen School of Computer Science. There are several norms related to publishing in computing science that are worth noting. First, computing uses standards for listing authors that are common to science and engineering where the principal investigator overseeing the work is listed as the final author. Second, in computer science and related fields, most top venues for publication are conference proceedings. Published proceedings of computer science conferences are "archival" in the sense that works published in a conference, like articles in journals, are not typically published elsewhere subsequently. The most prestigious publication venues in my area of social computing are the *ACM Conference on Computer-supported Cooperative Work and Social Computing (CSCW)* and the broader *ACM Conference on Human Factors in Computing (CHI)*.

Third and finally, my work, like many other computational social scientists, is laboratory-based. Working in my lab, I have received more than 3.2 million dollars in external funding to work closely with a series of student research assistants and post-docs on research projects that are led by junior scholars but are typically designed and directed by myself and my faculty co-investigators. In a more unusual arrangement, I have taken a page from the peer production projects I study and have attempted to organize my lab so that it is virtual, collaborative, and open. What began as a joint lab with Aaron Shaw at Northwestern's Department of Communication Studies in 2014 has grown into the *Community Data Science Collective* which now also includes faculty from five universities (all our former students) and participation from dozens of other researchers.

Understanding Peer Production

Empirical studies of peer production communities form the central pillar of my research. While most other scholars studying peer production have treated individuals within a single online community as the central

unit of analysis, my work has typically looked across multiple peer production organizations [BC8]. These populations might include communities on Reddit, wikis hosted by a large hosting platform, FLOSS packages, and so on. Much of my research has sought to test established theories drawn from communication, sociology, and social computing research in these new empirical settings.

Several of my most important studies have sought to understand issues related to community governance using comprehensive contribution data drawn from the firm *Wikia* (a company recently rebranded *Fandom*). Although not as widely known as Wikipedia, Wikia was started by Wikipedia's founder and hosts hundreds of thousands of peer production knowledge bases that, together, make Wikia one of the 100 most visited websites in the world. The fact that Wikia includes many distinct communities makes it an ideal setting for the type of organization-level analyses at the heart of my work. In a paper published in the *Journal of Communication*, I tested and found disconfirming evidence for the widely cited claim that peer production is inherently resistant to the emergence of oligarchy [A10]. I did so through an empirical approach that combined "big data" techniques from computer engineering with digital trace data from more than 700 of the largest Wikia wikis. In the process, my paper provided one of the very first multiorganization quantitative tests of Robert Michel's "iron law of oligarchy" in any setting.

A follow-up publication at *CHI '18* connected these dynamics to lifecycles of community growth and decline [A24]. My analysis suggested that large peer production communities tend to grow rapidly before abruptly transitioning into periods of steady decline. Moreover, my study suggested that this decline was caused not by a stark decrease in incoming contributions but by increased newcomer rejection leading to decreased newcomer retention. In doing so, my paper established that an influential set of findings from research into English Wikipedia were common dynamics of peer production communities. It also presented correlational evidence connecting my previous findings about oligarchic leadership to decreases in community health.

A third and final study, published in *Communication Research*, provided causal evidence of the dynamics shown in the first two studies. Using digital trace data from 136 Wikia communities, the study used a quasi-experimental panel regression discontinuity design to measure the causal effect of community leaders' decision to require accounts on a set of community-level measures of activity [A34]. The study showed that although requiring accounts is effective at blocking damaging contributions from uncommitted users, it also causes a large and previously untheorized community-level decrease in contributions. This decrease in good contributions appears to be caused both by blocking would be contributors as well as by preempting a body of work from established users whose productive activity is catalyzed by newcomers' frequently low quality work.

I have conducted similar data-driven theory-testing studies in Wikia communities [A29], as well in other empirical settings that include Wikipedia [A18, A32], free and open source software [A35], online petitioning [A13], and the sharing economy [A23]. Like the third paper described above, this body of work has increasingly used field experiments and quasi-experimental designs using methods that I am among the first to introduce into communication scholarship. Because theory about peer production and digital public goods remains thin, I have also pursued a substantial body of work that has focused on building new grounded theory from ethnographic field notes, qualitative trace data, and ethnographic interviews that speak to how online communities govern themselves [A17, A27, A30] and how they handle their users' desire to remain anonymous [A29, A31]. Theories I have developed have served as frameworks for theory-testing empirical work conducted by both my lab and others.

Finally, as peer production research has developed, I have sought to synthesize an increasingly large body of scholarship on the phenomenon. To that end, I have coauthored a highly cited literature review on peer production research [BC5] as well a more recent review on Wikipedia research [BC9]. I am also the first author of a forthcoming book titled *Peer Production* that is currently under contract with MIT Press [B1]. My empirical and theoretical work on peer production has received generous support through through four grants from the National Science Foundation for a total of $2,610,510 ($1,428,284 at UW) as well as an additional $139,994 (all at UW) from the Sloan and Ford Foundations.

My empirical work on peer production has two main thrusts going forward. First, I am pursuing a series of studies that seek to understand how online communities interact with each other. Although many of my

community-level analyses have been limited to cross-community comparisons, several of my ongoing projects use methods and theories from ecology to understand how online communities interact in competitive and mutualistic ways. This work is being funded by a recent NSF award.

The second major thrust involves a focus on understanding online community governance in ways that build directly on the three studies of Wikia described above. This work will involve empirical projects that improve and test a mathematical theoretical model developed during my year as a fellow at the Center for Advanced Studies in the Behavioral Sciences at Stanford University. The model draws from the scholarship of Nobel laureate Elinor Ostrom to explain why successful knowledge commons tend to become increasingly closed to participation. As part of this work, I also hope to develop and test a set of strategies for managing the community lifecycle dynamics I have identified in empirical studies. This work is being supported by a NSF CAREER grant awarded in April 2021.

Democratizing Peer Production Research

The second focus of my research seeks to help others conduct computational and data scientific research. For example, I published an article that demonstrated and explained several computational methods for the *SAGE Handbook of Social Media* [BC7]. Like my other quantitative scholarship, the work followed standards for open and reproducible science and came with companion code, a full dataset, and detailed documentation. I have also published a series of pieces that describe new methods, once again with code and data, to help scholars avoid common pitfalls in quantitative studies of Wikipedia [A12, A14]. I have also published nine different datasets—each with detailed documentation and code—that typically reduce multi-terabyte "raw" data into clearly documented tabular formats usable by scholars without my background in computer science and engineering [D1-9]. The most successful example is a dataset of youth social media activity drawn from the Scratch online community. That work, which was described in an article published in Nature's *Scientific Data* journal [A19], was cited as the justification for the "Research Symbiont Award" which I received from the Pacific Symposium on Biocomputing in 2019. The award is given annually to "a scientist working in any field who has shared data beyond the expectations of their field."

I have also sought to democratize the ability to conduct computational work beyond academia. This has involved new scientific investigation into how people learn about data—especially in peer production communities. Much of this work has been situated in the Scratch online community where I conducted a set of studies related to remixing [A4, A6-8]. In my first year at UW, I was awarded an NSF grant for $433,262 ($124,374 at UW) that supported a set of projects focused on building new pathways to democratized data scientific learning. This work resulted in two now widely used techniques, first described in an award-winning paper published at *CSCW '16* [A15], for measuring computational learning in informal environments where quizzes and formal assessment are impossible.

This work also includes a recent publication in *New Media & Society* that seeks to unpack the dynamics of how learners in online communities communicate with each other to solve technical problems in programming [A33] as well as several other studies related to learning about data in online communities [A21, A25]. It also involved the creation of a new system that makes it easier for users of the Scratch online community to conduct their own programmatic data scientific analysis. The system, *Scratch Community Blocks*, was described in an award-winning paper published at *CHI '17* [A22]. There remains persistent inequality in participation based on gender and race in both data science and in many of the peer production communities I study. A body of my work has sought to measure, understand, and evaluate efforts to address this inequality [A5, A9, A26].

TEACHING AND MENTORING

I approach teaching and mentorship as opportunities to collaborate with students to use theoretical and methodological tools from my research to solve problems of students' choice. For example, my classroom teaching is strongly based around project-based work and typically involves a two-part structure where the first part of each course focuses on intense work to build a body of methodological or theoretical skills and the second part focuses on applying these skills to a real problem of the student's selection. For example, one of my students used theory and concepts from my computer-mediated communication class to design a strategy that was successful at attracting more than 300,000 TikTok followers.

I have taught 9 different courses a total of 19 times since arriving at UW. Students appear to enjoy my classes and have rated my courses with a median overall score of 4.7 on a 5-point scale. Students also found my courses challenging, but not overwhelming, with a median "course engagement index" of 5.2 on a 7-point scale. Figure 1 shows distributions of both measures and demonstrates that these evaluations have been consistent. My peer teaching evaluations have also been consistently strong.

I teach three types of classes. The first type maps to my first research focus and seeks to convey lessons related to online communities from communication, the social sciences, and computing. This includes an undergraduate class related to computer-mediated communication and online communities (COM 482). I have taught this course in my department's day as well as evening programs, both in person and online, and in both a 45 person version as well as a 90 person large lecture version with TA-led discussion sections. I will be teaching this material as a new class titled "Online Communities" (COM 481) beginning in the 2021–2022 academic year. I have also adapted this curriculum for the Department of Communication's Communication Leadership (Comm Lead) professional master's program where I teach a course on "Building Successful Online Communities." I have also taught a distinct class on online communities and innovation for Comm Lead.

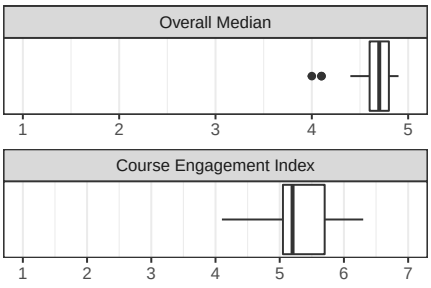

Figure 1: Boxplots of student course evaluations. Overall course evaluations are reported as estimated medians between 1 and 5. 'Course Engagement Index' captures perceived difficulty and ranges between 1 and 7. ($n = 19$)

The second category of classes corresponds to my second research focus and consists of courses related to Internet-based, statistical, and computational research methodology. This includes a class on data science designed for Comm Lead (now taught by others) as well as an MA/PhD class on Internet Research Methods (COM 528) which serves as a multimethodology methods course. I have also taught two MA/PhD classes on statistics and statistical programming. The third and final category involves courses that fill departmental needs in our PhD program. Specifically, I have co-taught the mandatory introductory communication theory course three times and the mandatory communication methods course once.

Without question, student mentorship is the most personally fulfilling part of my job. I currently advise or co-advise 6 students and have served on a total of 25 UW PhD committees (21 are active committees; 8 as a Graduate School Representative) as well as graduate committees for 8 students at other universities. I have informal mentorship roles, often focused on computational and quantitative methods, with a number of other graduate students. I meet with my advisees multiple times each week, communicate with most every day, and am closely involved in all steps of their research. In January 2018, I turned my faculty office into a meeting room and moved my own desk into my common lab space so that I could work side-by-side (literally!) with my students. Although my first advisee is still several months from graduation, he has one excellent post-doc offer. My only post-doctoral mentee is now tenure track faculty at the University of North Carolina, Chapel Hill. At least four other students I have mentored are in tenure track faculty positions at other excellent programs in communication, information science, and computer science. Many other students I have mentored are doing fantastic and impactful work in industry and in non-profits. My students' successes are the achievements of which I am most proud.

I have also supervised research work for undergraduates as part of an honors thesis, independent studies for credit, and as paid research assistants. One arrangement that I am proud of is the introduction of directed research groups (DRGs)—an idea I have borrowed from UW's Department of Human-Centered Design and Engineering—to the Department of Communication. In DRGs, I work closely with a group of undergraduate and graduate students from my lab on a collaborative research project in a class-like setting. I have run six quarters of DRGs to date. Several of the undergraduates I have mentored are now master's and PhD students at UW and elsewhere.

SERVICE AND OUTREACH

Finally, I am active in a range of service roles. At the departmental level, I chair the Statistics Concentration Committee and have served on the Diversity Committee, Colloquium Committee, Graduate Program Committee, Admissions Committee, Website Committee as well as three faculty hiring committees. At the University level, I have served on the Provost's Task Force on Data Science Education, I represent the College of Arts and Sciences on the Hyak Governing Board in charge of UW's super computer infrastructure, and I have twice co-organized the "DUB Retreat" (an annual gathering of several hundred people at UW working on human computer interaction).

I have provided service to the profession through leadership roles at the *Symposium on Open Collaboration (OpenSym)* where I acted as Program Co-Chair (equivalent to co-editing a small journal for one year) and as a member of the organization's steering committee. I have also served for several years as an Associate Chair at *CSCW* (equivalent to serving as an associate editor at a top journal) and am serving as the Panel Co-Chair for *CSCW '21*. I have co-organized the *OpenSym Doctoral Consortium*, the *Open and User Innovation Workshop* twice, one instance of the *International Workshop on Network Theory*, and have served on several conference program committees and journal editorial boards. I have also served as an external reviewer for more than three dozen journals, conferences, funders, and academic presses. I am an enthusiastic member and volunteer for the International Communication Association's Computational Methods division.

Equally important to me is the service I provide to peer production communities. This type of service is important so that my relationship to the groups I study goes beyond simply using them as sources of data and so that I can disseminate the findings from my research to those most able to use it effectively. I served on both the Board of Directors of the Free Software Foundation (the most important FLOSS NGO) and the advisory board for the Wikimedia Foundation (the NGO that runs Wikipedia) for more than a decade. I also serve as the Vice President of Cascadia Wikimedians which represents Wikipedia users in Washington, Oregon, and British Columbia. I also contribute to the peer production communities I study through technical work in the Debian and Ubuntu projects as well as regular editing in Wikipedia where I have made over 6,000 edits. I often integrate participation in Wikipedia into my classes.

In terms of my second focus on data science learning, I have run a number of outreach efforts focused on teaching data science to the general public. Most importantly, these have involved a series of project-based Community Data Science Workshops that I have organized six times over four days each [BC6]. These workshops have engaged more than 100 volunteers and reached nearly 500 learners from the general public in the Seattle area to build basic programming and data science skills. As part of this work, I have developed curriculum that has been adopted by others.

Finally, my service to the public involves translating scholarly work and research for nonacademic audiences. Toward this end, my research group maintains an active blog and Twitter account which have received hundreds of thousands of impressions. Additionally, I have given an annual talk at Wikimania (the annual Wikipedia community conference) on the "State of Wikimedia Research" and I co-organized a track of the conference on academic research in 2019. I have also given more than 100 talks, including many keynote addresses, to nonacademic audiences at conferences and events for peer production practitioners in FLOSS, wikis, and similar. My effort to reach broad audiences on my own has been complemented by coverage in the mainstream and technical press. For example, my work has been described in the *Washington Post*, *Financial Times*, *Wired*, *Die Ziet*, *Ars Technica*, and more.

REFERENCES

All citation keys correspond to items in the bibliography included at the end of my *curriculum vitæ* in Interfolio.