RESEARCH STATEMENT ASHTON ANDERSON

The Web is transforming our lives: it is increasingly where people communicate with their friends, find employment, make purchases, look for romantic partners, and stay in touch with the world. As the new platforms for daily life, digital social systems deliver billions of dollars of value to the economy and a huge amount of social capital to society. Despite their importance, however, these systems are still in their infancy, and rigorous principles for their design are lacking.

My long-term research goal is to develop *principled foundations for social systems*. In my work, I aim to achieve this by enacting an end-to-end research methodology: I collect massive social datasets from these systems, then analyze them to generate new knowledge about the human behavior they contain, from which I formulate rigorous design principles for the next generation of social systems—which can then restart the process anew.

I have applied this framework to many different real-world social systems. For example, I analyzed how LinkedIn's member invitation features affected its spread to over 330 million people; my analysis of Stack Overflow question-and-answering behavior directly led to predictive models of a piece of content's long-term value; and I developed a formal model of badges, which I then used in my implementation of a large-scale badge system on Coursera, which strongly increased their student engagement.

An integral part of the process is developing our social understanding. As the microscope and telescope opened up the worlds of the very small and very large to us, online social data is a "socioscope" that is opening up the social world to observation. To this end, I have collaborated with computational linguists, sociologists, educators, economists, and political scientists to use this powerful new instrument to answer old and new social questions.

Throughout my work, I apply a computational perspective to accomplish my research objectives. As life becomes increasingly digital, social data becomes increasingly computational—behaviors, opinions, and decisions are represented as bits, sets, and time series. Thus not only are computational techniques required to efficiently study the huge volumes of data we possess, but a computational framework is necessary to properly analyze and understand it, and to apply the resulting insights in the real world.

Social systems are made up of two fundamental components: people and information. In what follows, I will describe examples how my research develops rigorous principles for both of these components.

MOTIVATING PEOPLE WITH BADGES

People have long been motivating others by awarding them tokens of recognition: teachers reward their students with gold stars, militaries reward their soldiers with medals, and awards can be won in every discipline imaginable. Recently, a wide range web sites have followed suit by offering *badges*. However, the incentive structures induced by badges are not well understood, making it difficult to deploy them in a principled way. Poorly designed badges mean less motivation, which means less gets done.

We developed the first formal model of user behavior in the presence of badges. Our model is inherently computational: a web site defines an *action space* that users travel through, and badges define boundaries in this space that users can aim towards (Figure 1(Left)). Given a set of badges and a user's preferences over actions, we show how to solve for optimal user behavior at a point in action space by transforming the user's non-convex optimization problem into an efficiently-solvable form, then use dynamic programming to solve for the entire space. To evaluate the predictions of our model, we conducted a large-scale empirical analysis of badges and their effects on the widely used Stack Overflow question-answering site. We found that their badges steer behavior in ways closely consistent with the predictions of our model: as users get closer to fulfilling the conditions set to win a badge, they elevate their overall activity levels and increasingly focus their actions on winning the badge (Figure 1(Right)).

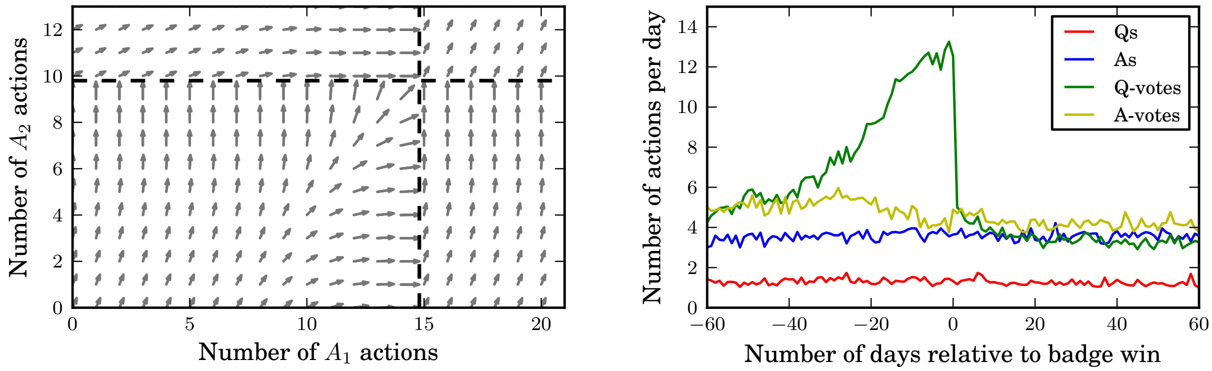

Figure 1: (Left) Optimal user behavior in the presence of two badges (whose boundaries are indicated by dashed lines). Arrows indicate the direction in action space users will move. Note how arrows both change direction (reflecting changes in their mix of on-site actions) and get longer (reflecting higher overall site engagement) as users approach badge boundaries. **(Right)** Empirical data showing both increased levels of activity and steering towards the badge-incentivized action (green curve) as users approach winning the Electorate badge on Stack Overflow.

In our framework, a badge system corresponds to a set of boundaries in an action space, and changing the boundaries changes the resulting incentives, and thus also the induced user behavior. Without this computational view, it would be very difficult to design badges effectively, but with our model it becomes easy. Using it, I deduced several rigorous *principles of optimal badge system design,* bringing rigor to a previously ad hoc practice. Moreover, my framework is equally applicable offline and online, meaning these principles can be used everywhere from loyalty programs to encouraging students to learn. This work is the 2nd-highest cited paper from WWW 2013, and has been taught at Columbia, Cornell, Harvard, Stanford, and University of Michigan, among other institutions [5][1].

Design and implementation of a large-scale badge system. To put my ideas into practice, I initiated and led a collaboration with Coursera, a leading MOOC platform, to apply the principles we formulated to the design and implementation of a real badge system. In April 2013, over 100,000 Machine Learning students started earning badges in my badge system, which I designed to incentivize particular actions in the class. As a result, user engagement on these targeted actions increased *fivefold* compared to previous offerings of the course, while engagement on "un-badgified" actions remained constant—evidence of my badge system's effects on student behavior.

At the same time, I took the unique opportunity to run a large-scale randomized experiment. To shed light on the mechanism by which badges derive their incentive power, I designed the badge system to be presented in several subtly different ways to different subsets of students, with each presentation accentuating different aspects of badges. Each method of making the badges more salient increased engagement, with the strongest effect coming from a design that made a student's own progress towards badges more explicit. It was very fulfilling to successfully apply my ideas in a real-world domain, and especially one as societally important as free online education. This work is the 2nd-highest cited paper from WWW 2014, where it was awarded Best Paper runner-up [8].

My work on badges exemplifies my end-to-end approach: motivated by a fundamental problem in social systems (how to motivate people), I applied ideas and techniques from both computer science and a relevant other discipline (behavioral economics) to formulate a model, then empirically validated it with real online social data, derived rigorous insights with it, implemented this new knowledge in practice to great effect, and finally ran a randomized experiment to generate further understanding that will inform future real-world systems.

Online reputation systems. Badge systems are part of a larger effort to develop the community trust that supports long-term interaction online. A well-established insight from game theory is that settings with repeated interactions can foster cooperation and lead to good social outcomes, largely because *reputation* matters. In light

[1] Citation numbers reference the numbered publication list in my CV.

of this, many online platforms have been instituting reputation and voting systems, which provide mechanisms for online communities to identify trusted members and signal high-quality content. Using data from these systems, I studied how *relative status* and *similarity* between two people affect how they evaluate each other. For example, I found that user–user evaluations are heavily status-driven between people who do not have much in common, but are less status-driven when they are similar. Insights like this were key to understanding how groups of individual evaluations are synthesized into an aggregate group opinion: they enabled us to predict the outcomes of adminship elections on Wikipedia using *only* the attributes of early voters (and not their expressed opinions) [2].

Question answering (Q&A) websites are a concrete example of the online shift towards long-term community: initially they were aimed at providing useful answers to the question asker, but now they function as group knowledge-creation sites whose end product is of enduring value to a broad audience. I analyzed the community dynamics of Stack Overflow and showed that reputational, temporal, and social signals surrounding questions are useful for the prediction of important outcomes, such as the long-term value of a question, and whether a question needs further attention from experts [4].

THE STRUCTURAL VIRALITY OF DIFFUSION

In addition to formulating principled foundations to support the *people* in social systems, I do the same for the *information* that is exchanged on them—the ideas, content, and products that spread through interpersonal networks. For decades, it was conjectured that this diffusion process worked analogously to the spread of an infectious disease, but until recently it has been prohibitively difficult to directly observe purportedly viral events, and thus determine the mechanisms underlying how cultural, professional, and personal "content" is disseminated. Although it is plausible that popular content spreads from person to person like a virus, it could just as easily be that it instead reaches large audiences through broadcasts from well-connected hubs.

Formalizing and empirically measuring virality. To settle this question, we introduced a rigorous measure of *structural virality* that quantifies the intuitive distinction between broadcast and viral diffusion, and allows for interpolation between the two (Figure 2). We then conducted an empirical analysis of more than 1 billion news stories, videos, pictures, and petitions on Twitter—to our knowledge, the most comprehensive diffusion study to date. With our measure and our dataset, we were able to establish the statistical frequency of structurally viral cascades for the first time, as well as make several other important discoveries; for example, we found that for any given cascade size, structural virality is surprisingly diverse: cascades can range from pure broadcasts to highly viral, multigenerational structures. This work illustrates how online data enables better social understanding, since the extreme size and granularity of our data was necessary for this work: individual-level adoption information is required for a structural analysis, and the "viral hits" occurred at a rate of about one in a million, thus requiring at least around 1 billion events to estimate their frequency reliably. This work is forthcoming in *Management Science*, the top journal in its field [9]. A team at Microsoft Research implemented these ideas in a product demo called ViralSearch, which was widely discussed in the press (including a video segment on The Economist and being spotlighted by Bill Gates on Twitter).

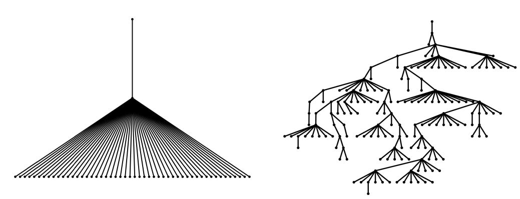

Figure 2: Two real cascades depicting broadcast versus viral diffusion, where nodes represent individual adoptions and edges indicate who adopted from whom. Our measure of structural virality interpolates between these two extremes.

How do massive web sites spread? Many of the world's largest web sites spread via an invitation mechanism—what are the characteristics of this spread? What are the implications of a web site's diffusion mechanism on its resulting user population? Does the adoption of social systems spread like the pieces of information that are exchanged on them? Previous work examined the spread of a few modestly successful "products", but how the most successful web sites spread—those we typically use every day—was unknown. To answer these questions, I collaborated with LinkedIn to study how registrations on their web site diffused around the world. With over 330 million members, a significant fraction of whom registered through an invitation from a friend, this was by far the largest product diffusion study to date. I found that LinkedIn's spread is significantly more viral than all previously studied diffusion events, and that it operates on vastly different timescales than information diffusion, providing a first and surprising look at the structure of an extremely successful product's diffusion. Using the incredibly detailed user-level attribute information available on LinkedIn, I also examined the interplay between the diffusion structure and user traits like country of residence and professional industry, and identified the traits through which LinkedIn spreads most readily (under review [10]).

RESEARCH AGENDA

My long-term research aim is to develop rigorous foundations for the design of social systems by using web-scale data to advance our understanding of human behavior, then implement this knowledge to vastly improve the digital systems that now support our world. Having only recently acquired the wealth of massive human data we now possess, as well as the computational ability to handle it, we are clearly only at the beginning of this road. In the following I describe several future directions I am excited to pursue.

Developing a computational understanding of reputation and incentives. My work on badges [5, 8] showed that thinking about incentive structures in a computational framework leads to conceptual clarity and direct practical benefits. Formalizing the incentives and resulting behavior introduced by other important online social systems, such as reputation systems and voting systems, would be similarly valuable.

Incorporating richer forms of content. Much of my work on improving social systems has so far exploited the structured social signals available online: binary signals (positive/negative votes, possession of a badge), series and sets of interest, and metadata information like timestamps. But there is a trove of knowledge to be discovered in the *content* that people author, such as text. I have already collaborated with computational linguists to initiate the development of a people-centric computational history of science through the text of scholarly articles academics write over their lives [3], and look forward to incorporating these techniques into the study of the wealth of content present in online social systems.

Running experiments. My experience running a large-scale online randomized experiment to establish the causal effects of badge systems [8] convinced me that online experimentation will be an ever-increasing part of the computational social science toolkit. I want to continue collaborating with industry partners to run experiments that are both scientifically illuminating and practically useful. There is a host of interesting and difficult problems to deal with when running randomized experiments online, and I also plan to contribute methodological tools and best practices in working with online participants to help advance the modern practice of scientific experimentation.

Analyzing and designing the sharing economy. One of the most promising socio-technological developments of recent times has been the rapid growth and adoption of the sharing economy, or *collaborative consumption*. Online social systems now enable people to provide and stay in private accommodations (Airbnb), enlist the help of others in doing various tasks (TaskRabbit), and contribute funds to each other's causes and projects (Kickstarter). All of these applications dramatically increase societal efficiency, and demand computational perspectives and techniques to improve them. Long-term user reputation is absolutely critical in these systems; many people wouldn't stay in someone's house, for example, if they didn't have good reason to trust them. My work in analyzing and designing trust-based systems for long-term interactions [2, 4, 5, 8] analyzes how these community status processes work and provides a framework for understanding and designing them, and I am excited to apply this knowledge in the context of the sharing economy.