

External engagement and outreach

I strongly believe in outreach to practitioners in industry, both to ground my own research and to increase the adoption of research in practice. As a direct result of this outreach, I am currently working with engineers at Facebook to address replication lag in the MySQL database system, where it has the potential to significantly simplify database maintenance and the broader ecosystem of systems built on top of MySQL. We plan to open-source this work, and expect it to impact other production users of MySQL. (Several of whom, including Booking.com¹ and Percona², have reported replication lag to be a long-standing issue with the system.) I have also given talks about my research onsite at several organizations and at two major practitioner focussed conferences, StrangeLoop in 2017 and Ricon in 2015. These talks provided an excellent platform to make my research accessible to a broader audience. For instance, my talk on addressing replication lag at StrangeLoop 2017³ resonated with several practitioners at large-scale organizations (e.g., LinkedIn and Uber). Finally, my ideas for future research are significantly motivated that by conversations I have had with practitioners. For instance, one direction I am interested in is scaling out program analysis over distributed and multi-core systems. In conversations with engineers at Facebook and Google, I found that restricted parallelism and limited memory resources on a single machine are among the biggest barriers to the analysis of large codebases on the order of millions or more lines of code. I have found engaging with practitioners extremely rewarding during the course of my PhD, and will continue to do so in the future as a professor.

I am also interested in the role technology plays, for better or worse, in developing communities. For instance, in India, there is a government mandated push towards dragnet-like collection of citizen information via the Aadhaar project. Under this project, sensitive information such as bank account details, income tax documents, and so forth, are linked to each individual, and stored in a centralized government authority controlled database. While the ostensible justification for this project is the reduction of bureaucracy and corruption in citizen access to government services, it raises questions with respect to privacy, potential for abuse, and accountability. More generally, given the increasingly widespread collection of user information on the Internet, I am especially interested in understanding and communicating privacy risks to developing populations with an aspirational view of technology.

Finally, I am also interested in the role of Big Data in society. Big Data is often touted as an unbiased tool to reduce human bias and inefficiency in decision making. However, Big Data has its own limitations due to the data itself reflecting existing biases. More importantly, the current generation of tools lack any form of transparency. As a reaction, there has been recent interest in *fairness* in machine learning, and making data and algorithms public domain. Although these are encouraging first steps, they are ultimately limited because they are insufficient to understand *why* a machine learning model makes a particular prediction; they lack the ability to provide provenance and explainability of results. I believe that explainability can be one of the key underlying technical mechanisms to make Big Data more accountable; explainability could provide individuals with a path to recourse in the face of black-box algorithmic decision making. In addition to the technical challenges involved in explainability, I am interested in the broader issue of explainability and Big Data in the real-world. For instance, how does explainability affect existing business models and the various stakeholders? (E.g., in the context of online recommendation systems; users, tech companies, and advertisers.) How do we educate non-technical users on the pitfalls of using non-explainable Big Data? What are the legal implications of using non-explainable Big Data to filter job, loan, and student applications? Given their ability to profoundly impact the lives of people, not just their online personas, I strongly believe that these questions deserve at least as much attention as the technical issues around Big Data and explainability.

Diversity

My outlook on diversity is influenced by my personal experiences growing up in India. I grew up in the state of Goa, a small former Portuguese colony on the western coast of India. Under Portuguese occupation, Goa evolved its own unique sub-culture and demographics, quite different from the rest of India. When I moved out of Goa to pursue my undergraduate education, I was one out of just three Goan students in a

¹<https://tinyurl.com/booking-replication-lag>

²<https://tinyurl.com/percona-replication-lag>

³Addressing the Parallelism Gap in Replicated Server Systems. https://www.youtube.com/watch?v=0BC_FOIxy4

total incoming class of over 700. Despite the overwhelming odds, my undergraduate experience was, for the most part, very inclusive. I was fortunate to attend an institution that attracted a diverse body of students from all over the country. This experience made me empathetic to individuals who were different from me.

The lack of diversity in several computer science disciplines, including my own field of database systems, is often attributed to the *pipeline problem*; a lack of diversity in individuals studying computer science at the school and undergraduate level. This is a complicated problem with a number of underlying causes that requires a pro-active approach to ameliorate and resolve. I believe that professors and admissions committees can play an active role in recruiting a diverse graduate student body. For instance, admissions committees should themselves consist of a diverse group of faculty in order to get an unbiased take on applications from underrepresented communities. Professors can play an even more direct role by recruiting a diverse set of undergraduate student researchers. These student researchers can seed the pool of graduate students and industry practitioners, and also serve as role models for younger undergraduate students. Finally, I strongly believe in the power of role models to inspire students to take up computer science. Professors from underrepresented communities can play a unique role in engaging with their local schools and undergraduate populations via talks and lectures. However, outreach takes effort and time, and I believe that professors from well represented communities can support their peers by, for instance, reaching out to school administrations or student bodies to set up speaking engagements. I would be delighted to support diversity by helping colleagues with outreach efforts.

As an international student, I am acutely aware of the challenges in adjusting to life as an undergraduate or graduate student in the United States. For instance, international students often come from backgrounds in which classroom participation is less vigorous than that in the United States. As a consequence, some international students can find it intimidating to engage in discussions with peers and professors. Furthermore, there is also the challenge of managing important practical issues such as personal finance, immigration, and healthcare. Although information on these issues is generally available via international student centers, practical advice on how to manage them is often lacking. I had very little guidance on these issues as a graduate student, and although I have learnt how to deal with them over time, I would have benefited immensely from a better support structure. In this regard, professors who were themselves international students are in a unique position to offer relevant advice to new students. As a professor, I will encourage establishing a mentoring program to help international students adjust to life in the United States.

Finally, I believe that the best way to contribute to diversity is to actually encourage students to unapologetically be themselves. My academic advisor gave me the freedom to define and explore my own research path. I valued this freedom, and with time realized its importance in giving me the confidence to develop a working style and philosophy I can call my own. As a professor, I will encourage students to be themselves, and avoid burdening them with an expectation to conform to a strict set of cultural norms. Our community will benefit immensely from a vibrant set of researchers, each with their unique approach to mentoring, communication, and problem solving.