

When You Know Better You Do Better: Increasing opportunities for diverse experiences at data science conferences

Latifa Jackson¹, Heriberto Acosta-Maestre², and Hasan Jackson³

¹Department of Pediatrics and Child Health, Howard University

²College of Engineering and Computing, Nova Southeastern University

³QuadGRID, Inc.

latifa.f.jackson@gmail.com

ha221@mynsu.nova.edu

hackson@quadgrid.com

Abstract

Decades have passed since the Internet and consumer technology revolution took hold of our lives. In the past 10 years, millions of dollars have gone into all levels of STEM education from primary school to university and there has been an exponential growth in the number of diversity and inclusion programs. However, the career outlook for STEM minority students has yet to improve; Hispanic and Black people continue being the most underrepresented groups in the tech sector. With recent events, the lack of diversity at all corporate levels of the technology sector has come under the microscope. Current hiring strategies continue to fail minority groups. One major problem we found was that minority populations are invited to participate in technical conferences but mostly as observers and not as subject matter experts who can contribute to the discussion. The goal of this paper is to propose solutions that allow historically marginalized populations to take their equitable place as intellectual contributors in the technology sector and increase the perceived perspectives from a potential idea of diversity window dressing to comply with federal mandates to a true intellectual boon in the economic work space. We propose three solutions: the diversification of available data sets; the creation of equity in the conference marketplace of ideas; and the preparation of conference spaces for the historically marginalized to participation.

KEYWORDS: Broadening Participation, Data Science, Gentrification, Conferences, Mentoring, Retention

Introduction

The past tumultuous year has provided prime opportunities to reflect on the impacts of previous diversity initiatives on the technology workforce. For the past decade, we have seen a growing abundance of diversity and inclusion programs and broadening participation and diversifying the types of workers associated with the tech industry. These efforts have included all kinds of initiatives from funding K12 education to providing research and enrichment opportunities for undergraduate and graduate students on the text side numerous

fellowships have been established that identify diverse populations as the target audience. These initiatives alone have been funded extensively in the millions of dollar range.

Despite these efforts, the constellation of diverse individuals working in technology careers has not significantly changed. Hispanic/Latinx and Black people are the most under-represented in technology relative to their representation in the US with Hispanic/Latinx make up 18 percent of the US population, but only 8 percent of technology employees, while Black people make up 13 percent of the US population but only 5 percent of technology employees (Greet 2020). In the UK the picture is even bleaker, with only 5 percent of technology roles are filled by black and Latino candidates even though they make up 18 percent of computer science graduates (Diversity in Tech UK, 2020) and the same patterns hold in venture capital initiatives (Zhang, 2020).

Recent high profile employment and retention events within the technology sector have brought renewed focus on the need for diverse staffing in technology careers, in management chains and the leadership suite. From an academic perspective we are tasked with training the workforce of the future developing creative and innovative thinkers who bring with them their whole history of experience to the table in solving this country's technology dilemmas. At the same time we know that the focus and utility of technological solutions has not been evenly deployed across the United States. There are some populations who have been saturated with innovation and some populations and markets that have been systematically excluded or under invested. This is why we can find situations where the primary users or the long-term users of a platform are not being included in the decision-making process about what services they will ultimately receive. This is both short-sighted from a resource perspective and unethical from a moral perspective if we as a society believed that all populations of Americans are just clones of the majority population then we lose out on the shared opportunity to build something that works better for all Americans.

There have been numerous anecdotal accounts of how the efforts to recruit diverse Tech employees have been undermined by a hiring structure that does not fundamentally

support broadening participation. All of this leads us to ask ourselves what chance is there to diversify tech spaces if we address diversity as a pipeline or skills issue instead of a systematic exclusion issue. Here we focus on one area tasked with the diversification of the tech Workforce. Conferences represent a major opportunity for industry, academia, non-profits, federal labs, and the federal government to share information about innovation. These conferences also serve as a networking and recruiting opportunity of individuals working within a technology discipline. It is with this justification that many working in the diversity space focus on conferences as a mechanism for historically marginalized students to gain insights into and build networks with technology partners, research institutions, or commercial concerns.

Since 2012, the BPDM has existed with the goal of expanding exposure to data science by bringing historically marginalized voices to technology conferences. In the past eight years it has been formally associated with Society for Industrial and Applied Mathematics-Data Mining and Analytics Conference as well as the Association for Computing Machinery Special Interest Group in Knowledge Discovery and Data Mining Conference (ACM SIGKDD). In 2019 BPDM held a stand alone meeting three day workshop at Howard University, and in 2020, based on our experiences with BPDM, we served as cochairs to the Diversity and Inclusion track for ACM SIGKDD 2020.

One significant problem that we have encountered is that historically marginalized populations have been invited to participate in technology conferences often as external observers and not subject matter experts who can contribute to the ongoing intellectual discourse. We find this a disturbing trend at conferences where the knowledge economy is an intellectual aquarium in which our conference participants are systematically excluded. Furthermore the kinds of data elevated as worthy of conference presentation often are not compelling to historically marginalized peoples who are interested in making intellectual contributions as well as contributing to social good within their communities. In response to this observation we have developed a short list of things that we think can help to ameliorate this occurrences. The goal of this paper is to allow historically marginalized populations to take their equitable place as intellectual contributors in the technology sector and augment the perceived value of diverse perspectives from a potential idea of diversity window dressing to meet federal mandates to a true intellectual boon in the economic workspace.

Idea 1: Diversifying the Data Available

Data science has generated a whole series of compelling technological solutions. These methodologies have challenged the way that we think about analytical Solutions in a whole host of disciplines. But these methodologies do not speak to the underlying motivations that individuals have for becoming subject matter expert. Somewhere in the training of individuals something about the data that they were studying was compelling. That is to say that it reminded them of someone struggling with cancer or identified a business solution that had heretofore eluded Society, or potentially it

reorganize political data in a way that was more meaningful for the communities that the individual was apart of. these motivating factors are often missing in the generic nature of the data sets that historically marginalized populations are presented with. At Howard University, a historically Black university, students are encouraged to find service in their intellectual pursuits. The question then is how do we build and share data sets that allow students to explore the most cutting-edge methodologies and motivate them to provide service to the communities who have made sacrifices so that that individual can go to school. One solution to this, is to build data sets that speak to the social economic health and political needs of communities.

As data science research and applications continue to expand into a variety of fields such as medicine, finance, security, and marketing; the need for talented and diverse individuals is clearly felt. This is particularly the case as 'Big Data' initiatives have taken off in the federal, industry and academic sectors, providing a wealth of opportunities, nationally and internationally. The Broadening Participation in Data Mining (BPDM) program was created with the goal of fostering mentorship, guidance, and connections for minority and underrepresented groups in the data science, machine learning, and computer algorithm communities, while also enriching technical aptitude for a long-term career in data science.

A key obstacle to preparing underrepresented trainees has been the limited ability to expose trainees to opportunities that demonstrate the analytical abilities and sophisticated inferential skills to address the coming data issues that our community faces. This is exacerbated by the lack of true participation of these trainees when they attend meetings at which they are not presenting their own work. We believe true parity exists when these bright underrepresented students have an opportunity to meet their peers as intellectual equals with unique contributions to their fields. One of the major outcomes of the past evaluation of BPDMS has been the need to identify research topics that are of both analytical and social justice interest to help facilitate trainee participation. This past year gentrification was identified by trainees as a pressing social justice issue in the face of housing shortages and the economic displacement that occurs in ethnic minority communities across the nation. It also has the potential to be accelerated by the coming economic crisis, the fight for social justice in the wake of George Floyd's murder, and the global coronavirus/COVID-19 pandemic.

Idea 2: Creating Equity in the Marketplace of Ideas

The stated goal of diversity efforts in technology has been to diversify participation in technology careers. To address this goal, a variety of use cases are leveraged that seek to enhance experience capacity building and economic incentive arguments (Kuhlman et al 2020). Unfortunately, more frequently conference diversity programs are implemented as a spectator event, meant to demonstrate that rewarding technology sector careers remain just out of the reach to historically marginalized populations. This is further reinforced

with technical talks and keynote discussions lacking diverse speakers and panelists, awkwardly scheduled poster sessions that relegate trainees to inopportune times to communicate their expertise, and a strong focus on recruitment to less-desirable (e.g. non-permanent) positions within the technology sector. Being a spectator is a demoralizing reminder that technology spaces were not created with broad, and consistent, inclusion in mind. Given the inherent 'othering' in the marketplace of intellectual ideas, how then are trainees and early career scientists expected to understand broader participation programs?

We believe that technology conferences must make opportunities available for those who do not have the privilege of educational pedigree and societal cache. We suggest that by implementing data themes that are attractive for social good, conference planners can extend the opportunity of historically marginalized to coalesce around topics that both show their intellectual capabilities and address topics of societal good. The limited inclusion results in a lack of presentations from diverse analytical and interpretive perspectives. A key feature of technology conferences has been the extreme selectivity of the submission process that recapitulates the robustness of the technology network as a feature of paper acceptance.

For example, our project to develop a gentrification data space is geared to develop subject matter expertise within a robust analytical framework that participants can use as a calling card to their personalized technological abilities. For this project, we are convening a community and trainee data hackathon on the effects of gentrification to have participants identify up to 12 research questions related to gentrification that employ cutting-edge analytical methods including machine learning approaches, geospatial modeling, neural networks, health bio-informatics, and natural language processing. Each group will be assigned an appropriate subject matter expert to help guide their research development, help to troubleshoot research concerns, and to advise their group on the writing of their findings. Finally, completed manuscripts would be open to a mentor panel peer review process that would ensure high-quality papers with appropriate academic rigor will be sent forward to technical conferences such as ACMs and other conference venues. We ultimately seek to increase the number of underrepresented trainees who are not just attending data science conferences but also to enhance the number of participants who are bringing their unique intellectual contributions to the technology marketplace of ideas. This will facilitate a paradigm shift in the practical utility of diversity initiatives as being more than superficial photo opportunities.

If conferences do not provide equitable opportunities for intellectual exchanges that privilege both innovative analytics with a sophisticated interpretation of the implications of analytical analysis, then they are fundamentally constraining the ability of the technology sector to evolve and innovate.

Idea 3: Preparing Spaces for Broader Participation

In the above section we provided what we think are two ways forward to meaningfully engage historically marginalized trainees as they participate in conference workshops. But this engagement requires conference planners and technology companies to also meet the moment with a renewed commitment to diversity, one that takes identifies with consistent commitment to diverse perspectives, not temporary positions to provide the illusion of inclusion.

Even African American Silicon Valley CEOs have routinely been assumed to not in charge of their companies, constantly challenged regarding their credentials, exposed to debilitating subtle and overt discrimination, and subject to the outrageous suggestion that they hire a white business partner to put 'investors at ease (Anand McBride, 2020).' If this is happening at the highest levels of technology companies, then what can we expect within lower level technology teams. We must identify and track the ability of technology companies to successfully recruit and retain historically marginalized scientists. Historically Black Colleges and Universities/ Hispanic Serving Institutions account for the education of nearly half of all Black and Hispanic/Latinx people in STEM (Cullinane, 2009). Despite this significant contribution to the training of historically marginalized populations, minority serving institutions (MSIs) have been largely excluded from preferential recruitment in favor of a handful of predominantly white research institutions. Meanwhile, our previous work suggests that historically marginalized populations continue to be among the most creative in their analytical solutions because of their lack of access to ready computing resources (Jackson et al. 2019).

Additionally, we have studied the effects of innovative mentoring strategies to improve trainee participation in technology (Jackson Acosta Maestre, 2020). These strategies move the responsibility of mentoring from the mentee to the mentor, whose influence, professional network and life experiences must be leveraged to increase the success of historically marginalized scientists. This mentoring also needs to buffer trainee from the destructive narratives that diversity comes at the cost of intellectual quality and rigor (Jackson et al 2021). Such arguments are not founded in data and are a manifestation of ongoing racialized biases in technology that constrain participation.

Discussion

A commitment to diversity must be infused into technology conference spaces and not an afterthought in response to complex social events. The number of technology conferences that appeared to highlight diversity issues AFTER the tragic death of George Floyd were noted and noticed by those whose participation was sought. The desire to include useful perspectives can not be a form of virtue signaling at times of crisis. The individuals who are being recruited to expand the universe of intellectual perspectives in technology have developed fine tuned survival skill sets in order to protect themselves. This means that conferences do long term damage to their own reputations by only soliciting his-

torically marginalized participation when there is an ascendant social movement such as Black Lives Matter. Diversity is not a topical concern, but should instead be a deeply rooted ethos that permeates the aims of technology conference planning.

The suggestions we propose in this paper are a vital first step to building intellectual parity for historically marginalized population participation at tech conferences. Ultimately, participation in conferences has to incorporate the sociological contexts of analysis in data science, allow for a plurality of perspectives, and open up innovation spaces that truly moves outside the cultural and archetypically constrained solution space. We intend to use our first two suggestions to measure changes in how historically marginalized population experience tech conferences and their long term affinity to technology careers. It would be interesting to explore in the future how these intellectual empowerment modifications moved the perspectives of hiring manager within the technology sector.

Technology conferences must also decide if they are meant to be a curated non-diverse luxury experience or if their purpose is to contribute to the broad marketplace exchange of ideas. Without this clarity we will continue to see diversity programs and diverse perspectives relegated to the fringes of technology conferences and broadening participation never be accepted as central tenet of the tech sector. This will ultimately deprive this vibrant sector of the labor resources needed for true computational innovation.

Bibliography

- Greet, S, June 25, 2020. BeamonJobs: Racial Diversity In Tech By The Numbers, *BeamonJobs*. Accessed online at: <https://www.beamjobs.com/diversity/racial-diversity-in-tech>.
- Diversity in Tech UK, 2020. How to Increase Diversity in Tech. Available online at: <https://www.diversityintech.co.uk/how-to-increase-diversity-in-tech>.
- Zhang, Y., 2020. Discrimination in the Venture Capital Industry: Evidence from Two Randomized Controlled Trials. *Arxiv* arXiv:2010.16084v1
- Kuhlman, C, Jackson, LF, Chunara, R. 2020. No computation without representation: Avoiding data and algorithm biases through diversity. *Arxiv* 2020: <https://arxiv.org/abs/2002.11836v1>.
- Anand P and McBride S. 2020. For Black CEOs in Silicon Valley, Humiliation Is a Part of Doing Business. *Bloomberg* June 16, 2020, Accessed online at: <https://www.bloomberg.com/news/articles/2020-06-16/black-lives-matter-highlights-adversity-facing-black-tech-ceos>
- Cullinane, J. 2009. Diversifying the STEM Pipeline: The Model Replication Institutions Program. Institute for Higher Education Policy.ERIC: ED508104. 2009-Dec. <https://files.eric.ed.gov/fulltext/ED508104.pdf>
- Jackson, LF, Acosta-Maestre, H.The Data Science Mentoring Fire Next Time: Innovative Strategies for Mentoring in Data Science. 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining August 23–27, 2020 Virtual Event, USA. <https://doi.org/10.1145/3394486.3411077>
- Jackson, L, Kuhlman, C, Jackson, FLC, and Fox, K. 2019. Including Vulnerable Populations in the Assessment of Data from Vulnerable Populations. *Frontiers in Big Data* 2:Article 19.
- Jackson, LF, Contreras-Sieck, MA, and Fox, K. 2021. Bending the Arc Towards Genomic Justice: Preventing bio-colonialism by furthering inclusion in genomic science. Forthcoming.

Acknowledgments

LJ and HAAM acknowledge the Broadening Participation in Data Mining Workshop (BPDM, www.dataminingshop.com) with which they have worked for the past several years. LJ (PI)and HJ (CoI) acknowledge the 2020 Award for Inclusion Research grant that supports our ongoing research on Gentrification.