Gaming Expertise Metrics: A Sociological Examination of Online Knowledge Creation Platforms

Tanya Osborne[1] · Markus Nivala[1] · Alena Seredko[1] · Thomas Hillman[2]

Abstract

Online environments have the potential to disrupt traditional orderings of expertise and allow a wider audience to engage as experts in the process of knowledge creation. Many online environments use crowdsourced metric-based systems, such as upvotes and reputation scores, to help identify experts. While these online systems provide opportunities to recognize expertise in new ways, they continue to reproduce and reconfigure existing issues around what is considered expertise and who is considered an expert. To explore these issues, we conduct an ethnographic engagement with Stack Overflow, a community knowledge sharing platform for programmers and coders. Our empirical material comprises 14 interviews with Stack Overflow users and documentary material collected from the platform.

We find that while platform owners often present their systems as arbiters of expertise, crowdsourced metric-based systems like Stack Overflow may not locate subject matter experts, but rather may measure community trust and engagement. Reward mechanics may also tempt users to play the game of increasing their metrics rather than developing expertise. Furthermore, our findings suggest that certain conditions in online environments may actively prevent those from under-represented groups from being recognized as experts. This presents an ethical challenge to the premise that online platforms may democratize the sharing of knowledge and recognition of expertise. Our findings suggest that while users may be aware of such issues, metrics are so intertwined with the functioning of the platforms they use that they may accept the premise that those metrics are relevant measures of expertise regardless.

Keywords Expertise · Internet · Ranking · Programming · Democratization · Masculinities

Extended author information available on the last page of the article

Introduction

The advent of online environments has led to significant changes in the dynamics of knowledge creation and expertise. Unlike traditional settings where expertise is often limited to a select few, online platforms have the potential to disrupt these established hierarchies and allow a much wider audience to engage as experts in the process of knowledge creation. However, while online environments may have the potential to disrupt established hierarchies, they may reproduce existing hierarchies as much as they may produce new ones. One common way in which online environments facilitate the recognition of expertise is through the use of crowdsourced, metric-based systems, which might include upvotes and user scores. These systems leverage the collective wisdom of the online community to identify and acknowledge individuals with valuable knowledge within a particular domain.

To delve deeper into these complex dynamics, our study focuses on Stack Overflow, a widely recognized and heavily used community knowledge-sharing platform for programmers and coders. Stack Overflow hosts over 24 million questions and 35 million answers with approximately 2.9 thousand questions asked daily (*All Sites - Stack Exchange*, 2023). As of 2023, there are over 100 million monthly visitors and 22 million registered users. As a large platform with millions of posts, Stack Overflow makes use of metric-based systems, such as upvotes and user scores, to order its content. By conducting an ethnographic engagement with Stack Overflow, we aim to gain a comprehensive understanding of the intricacies and nuances of expertise identification within this online environment. Ethnography, as a research methodology, allows us to immerse ourselves in an online community, observe interactions, and engage with users to capture their experiences, practices, and perspectives. In our ethnographic approach, we combined various data sources, such as participant observation, interviews, and analysis of digital artifacts, to develop a rich and holistic understanding of the complexities surrounding expertise in Stack Overflow. In this paper, we focus on findings from the interviews.

On digital platforms, metric-based ranking systems aim to provide a democratic and inclusive approach to recognizing expertise by relying on the collective evaluation of peers. However, it is important to recognize that while these online systems offer new avenues for recognizing expertise, they also bring forth a set of challenges. Despite the intention to democratize knowledge creation, they may reproduce existing issues surrounding expertise and expert identification. In the case of Stack Overflow, where only approximately 5–7% of users identify as women, a percentage much lower than the number of women working in programming (Nivala et al., 2020), crowdsourced metric-based systems for identifying expertise have clear risks when it comes to favoring certain types of contributions that align with gender norms. If women or other marginalized groups face biases in the evaluation and recognition of their expertise, they may receive fewer upvotes or lower scores than their male counterparts, regardless of the quality of their contributions. As our findings indicate that reputation metrics on Stack Overflow have ramifications understood by members to be consequential both on and off the platform, these kinds of inequalities can be contributing to other social problems in the workforce. This can create a cycle where those who are already recognized as experts continue to receive more visibility and

opportunities, while others struggle to gain recognition, reinforcing the gender gap in expertise recognition, and ultimately disincentivizing contributions from marginalized groups. Such concerns raise the question: what challenge do online platforms present to the promise of democratizing expertise? By exploring this question, we aim to uncover the underlying mechanisms and social processes that shape the construction of expertise within a specific online community. Taking Stack Overflow as a case and critically examining the practices and norms found there, we unpack how online platforms can challenge or reinforce existing notions of expertise and expert authority. The findings shed light on broader debates and discussions on expertise recognition in online environments.

Background

Democratizing Knowledge and Expertise

The premise that online platforms have the potential to democratize access to the production of knowledge has roots in the earliest conceptualizations of the Internet (Hindman, 2008). While the premise is commonly articulated, decades of studies show the promise of democratized knowledge to be highly contested (DiMaggio et al., 2001), not least in relation to intersections with multiple forms of inequality (Graham et al., 2014).

Research examining online platforms suggests that there are multiple ways in which the promise of democratized expertise is complicated. In a poignant example, Carraro and Wissink (2018) examined contributions to the coverage of the city of Jerusalem in Open Street Map, an open wiki world map. They found that submitted contributions for contested areas of East Jerusalem were regularly associated with conflicts in comment threads and related forums. Edits to those areas were often locked in protracted 'edit-wars' where changes were repeatedly reversed. These conflicts were found to often revolve around distinctions in the different geographic knowledges of Arab and Jewish residents.

In their study of the roles users take in discussions of controversial issues on Facebook, Twitter, and Wikipedia, Hara and Sanfilippo (2017) highlight the challenge associated with platforms where users are incentivized to judge the contributions of others to moderate unwanted content. Users who adopt this role may help to maintain an inclusive space but may also act as gate keepers who police certain types of knowledge and expertise. In relation to Wikipedia, König (2013) reveals a recurring tension between laypeople and experts. The study finds that while lay participation is seen as desirable for democratizing expertise, Wikipedia tends to favor elite knowledge and marginalize alternative interpretations. An overload of conflicting contributions is seen to lead to the exclusion and immunization of certain viewpoints. In this way, what König refers to as the 'participatory architecture' of Wikipedia fails to automatically lead to inclusive democratic practices. This point is consistent with Marwick (2013, p. 75), who claims that the outward appearance that anyone can edit Wikipedia is often undermined by existing hierarchies within the community. Such structures may contribute to entrenching existing gender gaps in participation (Reagle, 2013).

One of the key aspects of the architecture of many online platforms is that they make use of measures of user participation and performance to produce metrics (Mau, 2019). These metrics, which might include votes, likes, user scores, view counts and so on, are used to order content, sort users, and incentivize desired practices. For example, in a study of metrics and ranking algorithms on YouTube, Rieder et al. (2018) find that video and user rankings are influenced by factors such as search volumes and the number of videos published with that topic, along with the practices that develop in relation to these metrics or what Gibbs et al. (2015) describe as 'platform vernaculars'. For many platforms, key vernaculars that develop are those that enable users to judge the contributions made by others, such as through up- and down-voting posts. While these practices are occasioned by the material features of a platform, as Graham and Rodriguez (2021) exemplify, taking voting on Reddit as a case, users often disregard the intended purpose of a metric and instead create their own rules and norms around it. In this sense, while metrics such as votes are often framed as measures of the quality of the content with which they are associated, their meaning in relation to the premise that online platforms might democratize knowledge and expertise is complicated by the practices that produce them.

Gaming the System

In addition to complicating the way in which metrics on digital platforms contribute to the project of democratizing knowledge, metrics are also open to being gamed or manipulated. Search algorithms are one such place where gaming or manipulation has been normalized as part of how these systems are used, even spawning the profession of Search Engine Optimization (SEO) where consultants guide companies and individuals in how best to manipulate rankings and matchings to best serve their objectives. In an example of such manipulation, Gillespie (2017) follows the gaming of Google's PageRank algorithm that sees a cluster of political satire websites become top ranked when the name of a politician is searched for, in effect drowning out the official web presence of that politician. In response to this, Google changed their search algorithm. Such changes to algorithms are presented as trying to align better with an objective reality. While platforms like Google tend to present their results as impartial, they are calculated in ways that are open to manipulation. As Gillespie points out, if information providers want their information to be seen and used, they are compelled to become 'algorithmically recognizable', and to an extent need to play the game of the algorithm. There is a complex web between the various platforms and institutions that operate algorithmically, that reciprocally affects and produces effects on how we rank and judge the worthiness of information.

Petre et al. (2019), in their examination of how news articles deploy search engine optimization to game algorithms, reflect on how the line between legitimate strategy and gaming the system is very blurry. Their study examines three situations of gaming: Google search engine optimization, so called 'clickbait' or sensational content that provokes engagement on Facebook, and Instagram bots. Common to differentiating which strategies for success are legitimate is for the platform to position them as either organic or authentic, using a metaphor that good search results reflect something that is objective and natural. Petre et al. (2019) point out that what is considered

unwanted gaming is often discursively positioned as that by the platform despite using the same techniques as legitimate strategies. For example, Facebook simultaneously labels clickbait strategies as 'spammy', while actively advising people to use calls for action – a common feature of clickbait – to increase their engagement. It seems, then, that there is an intractable issue around identifying and ranking information versus the value judgement of whether this information is good.

Gaming algorithms does not only apply to manipulating visibility of information, but it also applies to manipulating the rankings and standings of individuals. These issues are particularly pertinent in Higher Education contexts. For example, the h-index is a measure of citation impact, allocating a score that represents the success of publications, where a h index of X means a researcher has X papers that have X or more citations. This h-index can be manipulated by social means, such as through informal citing arrangements and self-citation (Bartneck & Kokkelmans, 2011). These arrangements work by informally agreeing to cite work reciprocally in order to artificially inflate the number of citations that a published piece has. Measures of h-index can also be manipulated by gaming the platforms that are used to calculate h-index (van Bevern et al., 2016). For example, a person might use the merge feature on Google Scholar, which is ordinarily used to combine the citations for pre-prints and reissues with their original publication, to merge unrelated papers. In doing this, one might be able to falsely create a higher h-index by combining papers with lower amounts of citations. Pressure to perform on metrics like the h-index presents issues of academic integrity (Oravec, 2019). The work of pursuing the metrics can detract from doing meaningful research, while failure to participate in the game of metrics can have negative impacts on career and ability to attain funding.

These pressures and tensions drive people to play the game of improving their metrics, but ultimately call into question just what these metrics really measure. The best-ranked hit found via a search algorithm is often the most optimized rather than the most reliable. Similarly, an academic performing well on citation metrics may be savvy about playing the game, rather than producing higher quality research.

Programming Expertise

As a prominent arenas in programming culture, Stack Overflow and GitHub are often used to study programming and coding expertise, to identify expert programmers. While Stack Overflow is a community questions and answers platform where users seek and provide solutions to programming problems, GitHub is a version control and sharing hub for coding projects where many programmers coordinate their work and make it available to others. With their different use areas, the two platforms can be seen as complementary and significant figures in the work of many programmers.

A substantial amount of research exists that tests and evaluates different measures of expertise on Stack Overflow to judge the most reliable expert finding methods. For environments like Stack Overflow, the reason for wanting to identify experts from their activity is so that the vast amounts of posts and information that appear on the platform can be more easily ranked, sorted, and recommended to software developers when they search for similar questions. In some cases, this information is wanted in order to help job recruitment or to facilitate people finding someone that can answer

their specific questions (Huang et al., 2017, 2020; Procaci et al., 2016). These expert finding measures draw on content analysis, platform metrics, user behaviors, and network connectivity (Faisal et al., 2019). While this paper does not delve into the validity of these different approaches to producing metrics and identifying experts, it is relevant to note that expert identification on Stack Overflow is a topic that attracts significant research interest.

In addition to identifying experts, Stack Overflow is also used to study the behavior of expert users. Early research on expertise on Stack Overflow estimated that those who begin their tenure on the platform posting high reputation answers will continue to create high reputation answers, while those who produce low scoring answers will not substantially improve, suggesting that users join as experts rather than develop expertise on the platform (Posnett et al., 2012). Paralleling this assertion, Vadlamani and Baysal (2020) report that Stack Overflow users who participated in their study were wary of posting questions or of contributing to discussions where they did not consider themselves qualified, due to the perceived cruelty of the community toward novices.

However, what constitutes expertise in the domain is also contested. In search of a definition for what makes a programming expert, Baltes and Diehl (2018) developed a theory of software development expertise based on a survey of 355 Stack Overflow and GitHub users, focused on developers who use the Java programming language. Rather than focus on knowledge of specific coding tasks, their model suggests that software development experts have a balance of general and task specific programming knowledge, are able to write code that is easy to maintain, are open, analytical, and engage in peer review and mentoring activities.

While studies of programming expertise based on Stack Overflow suggests that openness is characteristic of experts, programming expertise in general has been repeatedly characterized as suffering from gender-based discrimination. Ethnographers have observed that software engineers tend of make a dichotomy between the technical and the social (Faulkner, 2000). This dichotomy has tended to be gendered, rendering femininity and the social together, while maintaining a relationship between masculinity and the technical. In engineering culture, this has also been seen as a hierarchy, where the technological and the masculine have a higher importance than the feminine and the social (Faulkner, 2000). In this respect, technical expertise is seen as masculine expertise.

Empirical research to some extent supports this understanding. Research by Joshi (2014) suggests that in male-dominated engineering and science settings, men are more favorable towards the expertise of other men and are less likely to recognize the expertise of women. This bias toward perceiving the technical as masculine persists even in online settings. Ford and Wajcman (2017) argue that emerging online knowledge infrastructures, such as Wikipedia, are reconfiguring expertise in a way that remains coded as masculine. This reconfiguring works by creating a power structure that depends on a strong technical understanding of the policy governance of the platform, as well as a mastery of the software and programming infrastructure underpinning the platform, which are required to participate at the highest levels. Reflecting this, a number of studies have suggested that women are more likely to doubt their expertise in online programming settings and are less likely to contribute

as a result. In an interview study with 22 women, Ford et al. (2016) locate a number of reasons why women do not contribute to Stack Overflow, among them a fear that their expertise isn't enough to make a meaningful contribution, and fears of negative feedback received on the platform.

In summary, the aspiration to democratize knowledge and expertise through online platforms is fraught with complexities and contradictions. While the Internet and platforms like Stack Overflow offer unprecedented access to information and collaborative opportunities, they also reflect and sometimes exacerbate existing inequalities and biases. Studies have shown that the participatory architecture of these platforms does not guarantee inclusive democratic practices; rather, they can become battlegrounds for 'edit-wars,' gatekeeping, and the marginalization of non-elite knowledge. Moreover, the use of metrics and algorithms to rank content and users can be manipulated, calling into question the objectivity and fairness of these systems. Gaming the system, whether through SEO tactics or citation arrangements, can distort the representation of knowledge and expertise, privileging optimization over reliability. Additionally, the domain of programming expertise, while benefiting from the collaborative nature of platforms like Stack Overflow, is not immune to gender-based discrimination, reinforcing the masculine coding of technical expertise. Thus, while online platforms have the potential to democratize knowledge and expertise, realizing this potential requires critical examination of these systems and their uses.

Method

This study follows an ethnography for the internet (Hine, 2015) approach, considering the internet to be an embodied, embedded, everyday phenomenon. Our approach is enhanced with a digital methods focus (Caliandro, 2018), following the medium of Stack Overflow as a platform and paying particular attention to the way that communication is structured on the platform through the use of mechanics such as tags. The study has been approved by the Swedish Ethical Review Authority and has produced a dataset that combines interviews, observational material, and information from the Stack Overflow data viewer. The data viewer allows access to the user information database that underlies the regular graphical user interface for the platform. This database contains all posts and meta-data, and the viewer allows users to create novel ways of displaying that information. While this paper concentrates on interview data, it also draws on analysis of platform documentation relating to the reputation mechanic on Stack Overflow such as company blog posts and official guidelines, and analysis of Stack Overflow's user information database through the data viewer. This additional material sensitized our engagement with the interviews.

This study uses 14 semi-structured interviews with high-reputation members of Stack Overflow. Recruitment was based on those who appeared in the top 2% of reputation on the platform. Interviewees were selected if they had an email contact associated with their user profile. The interviews were conducted between 2019 and 2021. Interviews lasted between 33 and 108 min. Interviews were conducted in English, although those interviewed represented a range of nationalities: 3 from North America, 3 from the United Kingdom, 6 from Europe, 1 from the Middle East, and 1

from Asia. Of those interviewed, 13 were male and 1 was female. The interviewees ranged in age from 21 years old to 59 years old. While this is reasonably representative of the demographics of the platform, our interviews significantly underrepresent India and over-represent the UK (Stack Overflow Developer Survey, 2022, 2022). Interview informants are referred to in the paper by pseudonyms, which have been fabricated to reflect the general style of usernames on the platform. Analysis of the interviews is based on an open coding process, whereby all the authors contributed toward coding and annotating the interview dataset. Presented in this paper is an analysis of interview materials coded as related to expertise.

Findings

Why do Users Game Expertise Metrics?

There are two main metrics that measure expertise on Stack Overflow: reputation and badges. Both of these metrics are displayed prominently on user profiles and next to usernames when a person posts a message. These metrics also have a material effect on the kinds of powers a user has on the platform, and for some of our interviewees even have real life advantages.

In this section of the findings, we explore why people might be motivated to game the expertise metrics, and how these metrics are understood to represent expertise.

Reputation

Reputation is a number associated with an individual user that is displayed next to their name on every post. Reputation is also displayed on the user profile along with other contextualizing metrics, including how many views their posts have had, how many posts they have made, and their reputation rank, for example, "top 15% this year". Stack Overflow also maintains reputation rankings that organize and list users based on the reputation they accumulated within the last week, month, quarter, year, or based on total reputation scores (Users, 2023). There are many mechanisms by which a user can gain reputation (detailed in Table 1). The most visible way that users earn reputation is through upvotes on questions and answer posts. Users can continue to earn reputation from upvotes on posts as long as the post is still live and viewable on the platform. Our informants recognize that this can lead to unintended problematic side effects; older posts may accumulate reputation based on knowledge that no longer reflects the expertise of the user:

> So, it's tricky that I gained 20,000 reputation points on [programming language], but ten years ago. And those questions keep getting their upvotes, even now, because people are finding that answer and so on (GhostByte, 36-year-old male).

GhostByte points out that things in which one was an expert ten years ago may not represent one's current expertise. Since older answers may still earn reputation

despite time gaps, reputation can be deceptive when it comes to locating a current expert in a particular programming language.

Reputation can be seen as carrying symbolic meaning. According to Stack Overflow guidelines, "Reputation is a rough measurement of how much the community trusts you; it is earned by convincing your peers that you know what you're talking about" (*What Is Reputation?*, 2022). Importantly, and a distinct feature in the case of Stack Overflow, reputation levels also act as milestones that allow users to access more tools and powers on the platform. At 15 reputation points users can upvote posts, at 125 reputation points they may downvote posts, and by 2,000 reputation points they may edit the questions and answers of other users without needing their revisions to be approved. Users with a reputation over 20,000 are considered "trusted users," and have significant moderation powers, such as the ability to vote for deleting and undeleting posts, and more (Privileges, 2023). In this regard, reputation on Stack Overflow has real and material implications for how users engage with the platform. The process of accumulating reputation also makes a user more 'algorithmically recognizable' (Gillespie, 2017), not only because the metric is so visible, but because it enables a user to have much more impact on the way information is sorted. This entwining of responsibility, expertise, and trust has an important discursive function in positioning those who succeed on the platform as more credible.

However, reputation is distributed unevenly on the platform, meaning that relatively few people have access to all these privileges, and only a slim percentage of users are "trusted". In fact, the majority of Stack Overflow accounts are not even able to upvote posts. Using the Stack Overflow data explorer, which allows us to run queries on a database containing platform statistics, we were able to see the distribution of reputation across accounts. Based on our analysis of platform data in November 2023, analyzing 21,564,349 Stack Overflow accounts, we found that the average account reputation is just 90, more than 70% of the user accounts have only 1 reputation point, and the maximum amount of reputation held is around 1.4 million. From this data we calculate that only 15% of registered users, representing 3,206,152 accounts, have enough reputation to upvote posts, and 5% of registered users, representing 1,055,557 accounts, have enough reputation to downvote posts. Only 0.05%, or 11,745 accounts, have enough reputation to be considered trusted. This gives a sense of just how concentrated reputation is in a relatively small section of the population of the platform. In this respect, while the basic principles of the platform are open and democratic, with reputation being earned on the basis of decentralized peer nomination, users still need to earn enough trust to participate even on a very basic level. Thus, there is a certain circularity in the way that the platform deems people trusted or expert; these evaluations are made by people who are already to some extent trusted.

It is also possible, to an extent, for users to trade reputation through the "bounty" mechanism (see Table 1). A bounty is a reward offered to someone for answering a question. Anyone with enough reputation can nominally stake between 50 and 500 of their own reputation on a question, deducting it from their own reputation and awarding it to someone who can answer that question (*What Is a Bounty?*, 2023). Using a bounty can increase the likelihood that a question is answered in most situations (Zhou et al., 2020).

Table 1 Events leading to reputation change (What Is Reputation?, 2022)

Reputation score increases		Reputation score decreases	
Event	Score change	Event	Score change
Question is voted up	+10	Question is voted down	−2
Answer is voted up	+10	Answer is voted down	−2
Answer is **marked** "accepted"	+15 (+2 to accepting user)	A user votes down an answer	−1
Suggested edit is accepted	+2 (up to +1000 per user)	A user places a bounty on a question	− full bounty amount
Bounty awarded to a user's answer	+ full bounty amount	A user receives 6 spam or offensive flags on their posts	−100
Bounty is awarded automatically	+ half of the bounty amount		
Site association bonus	+100 on each site		

In addition to sanctioned ways to trade reputation, there is also evidence that users coordinate to manipulate their reputation in other ways. Early in the life of the platform, the Stack Overflow blog reveals that there is substantial unusual upvoting activity (Atwood, 2008, 2009), suggesting that users may mutually upvote other users posts systematically, or use 'sockpuppet' accounts to increase their own reputation. Sockpuppet accounts are accounts that are owned by a user in addition to their main account, they may have a deceptive identity, and they are likely only used for the specific purpose of exploiting the system.

With such a framing, it can be argued that reputation score is positioned as a measure of trust by the platform, a measure that is operationalized by material features such as voting on posts and based on the judgement of peers. Reputation score can be seen as a means of gaining recognition within the platform, as a gateway to additional privileges, and as a form of currency to incentivize answering questions.

Badges

Badges, much like reputation, are displayed prominently on user profiles and next to posts. Unlike reputation they are displayed alongside small graphics. They are divided into three categories: bronze, silver, and gold. The color category of badges, much like medals in sport, are ordered in terms of the effort they take to attain, with gold badges representing the most challenging to obtain. Unlike reputation, badges are not linked to privileges in any way and are more like 'achievements' in video games. While some badges are awarded for general activities on the platform, such as the "necromancer" badge which is awarded for answering older questions, tag badges are awarded for activity within a particular programming topic which is demarcated on the platform by a tag, for example, python or java. A user earns a golden tag badge if they have provided more than 200 answers with a total score of 1000 or more; a silver badge if they have a total score of 400 for at least 80 answers; and a bronze badge if they have for a total of 100 for at least 20 answers. In some cases, badges can motivate behaviors. Even though users cannot earn reputation for making edits after they

have reached 2,000 reputation, they may continue to earn badges from editing, and this can lead users toward making low-value edits to earn badges (Wang et al., 2020).

Our interviewees felt that badges were often a helpful indicator of someone's engagement with a topic. Janet talks about the difference between bronze level badges and gold tag badges with regard to judging subject knowledge:

> *You can get bronze badges by just being a plotter, but you're not going to get a gold tag badge unless you actually know your subject. (Janet, 59-year-old female)*

Janet's assertion is the lower value tag badges are substantially less meaningful. Wang et al. (2020) confirm this to a degree in their quantitative study of badge-holders, estimating that 77% of users who have a badge only have one badge (i.e. a bronze badge), and that 33% of users who have one badge cease doing actions that would allow them to earn further badges. Obtaining a bronze badge may be a simple vanity project, but having multiple badges might indicate a deeper engagement, both with the platform and with the subject matter. It is therefore more indicative of skill if someone has obtained a higher order badge. This helps users to convey depth of programming knowledge, which is consistent with Baltes and Diehl's (2018) model of programming expertise.

However, some of our interviewees were aware that their tag-related activity and badges are not a good representation of their own areas of expertise:

> *if you look at my profile you could say: "Oh, he must know Java. He must be able to program in Java." I am very much a noob in Java actually, but since I was answering a large number of Eclipse questions in the beginning, most of those Eclipse questions were also tagged Java, because they were about Java programs made in Eclipse (iEagle, 49-year-old male).*

In this excerpt, iEagle talks about identifying in which programming languages one might be considered an expert. Eclipse is an integrated development environment (IDE) that can be used to program a number of languages, but is most frequently associated with Java, a language commonly used for web development. iEagle calls himself a "noob" at the Java programming language, which is internet slang for a newbie, often used in a derogatory way to imply a lack of skill. The problem alluded to here is that because of the wide range of tools and languages within programming, one requires a certain level of expertise in order to determine what someone else is an expert in based on their tag or badge activity. Someone with sufficient knowledge of programming tools would see a profile that has tags for both Eclipse and Java and may be able to determine that iEagle is proficient in Eclipse; someone without the relevant knowledge may assume that iEagle is adept in both areas.

This perhaps problematizes the assertion that a gold badge would be a good indicator of knowledge; the badge might not be sufficient information alone to determine what kind or depth of knowledge is held. The tagging system is an important part of the platform vernacular (Gibbs et al., 2015); however, unlike hashtags in other platforms, tags on Stack Overflow can be edited by others and their specific use changes

over time. For example, as new technologies and versions of technology emerge, older tags may be retrospectively changed to fit with current understandings, meaning that posts that were previously tagged as one topic area at the time they were written may later be reclassified. Knowing how to navigate and interpret these tags as they develop over time is a skill that demonstrates mastery of the platform, and to an extent, an understanding of the history and policies of the platform.

Both of these comments in summation support the idea that reputation is not as important a metric for determining expertise as badges, which also supply contextual information. Though programming may seem like a niche topic to outsiders, there is a great breadth and depth of specialization within coding and programming that, in these quotes, is implicitly treated as not transferrable. To an extent, badges help to narrow and define the possible nature of someone's expertise far more than reputation alone. Where reputation may measure trust, badges supply the context in which someone holds expertise. However, badges may fail to be reliable indicators because of shifting vernaculars.

This Game is Real Life

As a prominent, global platform for sharing programming knowledge, Stack Overflow has noticeable effects and impacts on the day-to-day work lives of programmers. In our interview set, many acknowledge that the visibility that their reputation on the platform gave them has offered opportunities to further and advance their careers.

> It [my reputation] helped me get a job at [well-known company]. [...] I like it because of that, but otherwise it's fake money (Brian, 59 year old male).
> This game is actually real life and real job opportunities (Datenrausch,21 year old male).

Brian shows that he is dismissive to reputation having any benefit outside of advancing his career. While Stack Overflow is a relatively niche platform, it benefits from being well-known in the programming industry, and well rooted in programming culture. Describing how his reputation score led to getting direct job offers from recruiters, Datenrausch acknowledges the real-world effects of being visible on the platform. He refers to playing the platform mechanics as a "game".

"Game" in this context hints at an orientation toward a contest, or something that can be won. In studies of masculinity, games and contest are noted as a way in which men both bond and build hierarchies (Meuser, 2007). Positioning something as 'a game' can be a way in which masculine cultures create a buffer of 'instrumental rationality' (Almog & Kaplan, 2017), allowing individuals to master the game while maintaining a distance from their emotional investment. In the examples from Brian and Datenrausch, their ambivalence of calling reputation a 'game' or 'fake money' distances them in some way from having an emotional investment in the act of acquiring reputation while simultaneously acknowledging the importance of its dividends. This allows the player to focus on the act of playing the game well without buying in to the principles of the game.

We can understand the game of Stack Overflow as one wherein a hierarchy is built of people who perform programming expertise. The winners of this game attain benefits and recognition in their everyday work lives and increase their power on the platform. For some, the impact of succeeding in the game is transformative:

> It helped me to actually have a voice in the crowd when I'm introduced [...] when I'm introduced, I see that a couple of big voices in the room, they lower their attitude, they try to recognize me, and they act differently (GhostByte, 36-year-old male).

For GhostByte, status on the platform has converted into real-world respect and recognizability. GhostByte is from a smaller, developing country, and does not work for one of the large, well-known tech firms. He clarifies in his interview the impact of his Stack Overflow reputation:

> Maybe I'm from let's say … no named country, like [redacted]. If I go to a conference, I'm not from USA, I'm not from Google, I'm not from Microsoft. I'm just from [country redacted], but I'm advertised as a top Stack Overflow user and people accept that and they know that this is a value. This is how it's used. I'm fine with this. (GhostByte, 36-year-old male)

Being associated with having a high reputation on Stack Overflow helps GhostByte to feel on an equal footing when sharing a stage with people from well-known technology firms. In this respect, some of the promise of platforms disrupting traditional orderings of expertise are realized. Stack Overflow reputation can allow individuals opportunities to be recognized as experts beyond their local geographies, potentially allowing anyone to be seen as an expert on a global stage.

However, this kind of disruption cannot be considered democratic, nor can it be considered meritocratic in terms of elevating domain knowledge. While acknowledging that systems like Stack Overflow could have the power to effect global shifts in how we locate experts, we must also trouble how expertise is represented, and also acknowledge how the game of reputation operates.

How to Game Expertise

Knowing the importance of the two main markers of expertise, and the potential benefits of performing well on those measures, many users find ways to game the system. Some of these ways are passive: being in the right place and the right time, being 'in the know', and having a good command of English bring a natural advantage that many are able to leverage to their benefit. Some users take a more active stance and are able to strategically perform simple and repetitive tasks in order to boost their activity.

Be a Good Communicator

Over and above programming expertise, many of the people in our interview sample felt that they had needed to develop and improve their communication skills to thrive on Stack Overflow. This in turn had opened up their career opportunities.

> *[…]But what has changed also is the international nature of the collaboration of [companies], which now have not tools, but have […] facilities in foreign countries, and who have contractors typically in India as well. And so, for all that kind of professional community you have to improve your English. (iEagle, 49-year-old male)*

In this excerpt, iEagle talks about how the changing nature of his work as a programmer means that he has a greater demand to speak English in professional settings. This means being able to be understood by a globalized English-speaking audience and requires that one can be understood as a technical communicator. In some ways, the global nature of platforms like Stack Overflow provides a test bed where a user can practice technical communication skills to an audience with a mixture of different language backgrounds.

> *I'm not a native speaker of English, and speaking English is still [difficult] […]. But writing lots of answers just trains you, and you also get feedback. That's the important thing. You can see whether people understand what you're writing because they will ask questions if it is confusing, and I think that is the most important thing I learned. (carrotpie, 43-year-old male)*

Carrotpie talks about feedback from other users being a helpful tool in developing his skills in English communication. In other places, this kind of feedback may be undesirable or promote hostility. Ford et al. (2016) note that fear of negative feedback on posts is in part a reason why women may not contribute to Stack Overflow. Similarly, in studies on Wikipedia, commentors note that women who contribute to Wikipedia may not want to invest the time and emotional resources required to stand their ground in the face of negative feedback (Reagle, 2013).

In some respects, Stack is famous for having a rigid communication style. Previous studies have linked this communication style to the low participation of women on the platform (Brooke, 2021; D. Ford et al., 2017), and other studies have certainly shown evidence that perceived hostility on the platform, linked to this communication style, deters women from participating (Ford et al., 2016). There is a complicated relationship between communicating in clear, technical English that can be understood by a general audience, and communicating in a way that makes the platform welcoming to others. Succeeding at the game of communication on Stack Overflow requires a familiarity with the platform policies governing communication conventions, and a willingness to eschew pleasantries.

Speaking about communication on Stack Overflow, RANA says:

If you recognize that communication skills are important, then Stack Overflow gives you a great place to improve them, [.] I would hope there are engineers whose managers are advising them in performance reviews and in one-to-ones: "Hey, you could do with working on your technical communication skills", and maybe linking that to Stack Overflow as a good place to do that. (RANA, 44-year-old male)

Indeed, it seems that it is important to communicate in line with a technical language, perhaps cognizant of Connell's exploration masculine expertise, in which technical language is raised as one way in which men form social groups (Connell, 1995/2005, p. 171). Other interviewees talk procedurally about techniques for communication and talk about the tensions in conveying quite abstract concepts in a language that can be understood in an international community. On Stack Overflow, what makes an expert stand out is almost certainly how successfully they can communicate technical information.

Speaking about the importance of communication skills generally in programming, Janet says:

Any job above entry level, at some point you have to persuade other people to do things your way. You have to be forgiven for something you've done that was wrong, you have to ask for help and ask for instruction in a way that will inspire someone to help or instruct you. (Janet, 59-year-old female)

While expressing her frustration at how the importance 'soft skills' is frequently downplayed in programming, Janet makes a very clear case that these skills are important for eventual career success. Interesting about Janet's position is that she talks particularly about skills in persuasion, rather than the more technical communication that the men we interviewed brought up. It may be that for some this distinction is not so apparent, but the task of having one's answer accepted and upvoted on Stack Overflow requires both that one can solve the problem, and also that one can persuade others that the solution is worth upvoting.

Be Early and in the Know

Because of the structure of reputation gain on Stack Overflow, there is a significant benefit to being an early adopter, or to having insight into technology trends. If someone joined the platform at the very beginning, they were able to earn reputation from trivial but popular questions, which persist and continue to gain reputation today. Similarly, if someone is a developer working on new systems, they may be able to predict which programming languages are going to be in demand and grow a portfolio of easy to answer but popular questions under that language.

In technology communities, being an early adopter may signal greater technological expertise, and therefore a higher social status. For example, users on X (formally known as Twitter) with usernames containing only one or two letters are usually early adopters who were able to get a desirable username early (Marwick, 2013, p. 77). In cases like these, status is conveyed by signaling that one is astute to technological

developments and able to identify an up-and-coming platform, demonstrating expertise on technology trends more generally. In the case of Stack Overflow, being one of those 'in the know' during the early inception of the platform was crucial in reaping benefits.

In some respects, being an early adopter of the platform resulted in having a higher reputation simply by being able to answer early questions:

> *So, these users, they were kind of lucky. I think of them in this way, they are lucky, and they joined Stack Overflow in 2008, and people just have really simple questions and he answered this (Virendor, 27-year-old male).*

Virendor notes a suspicion that questions available earlier on the platform were simpler and therefore easier to answer. Since the platform discourages duplicate questions, older questions with less complex content are referred to often when newcomers have issues. The typical practice is to use a comment to point the question asker toward the older question and close the duplicate. If a user has a few questions that they have answered that are frequently asked, those questions continue to attract views and upvotes even 14 years later. This is contrary to Posnett et al. (2012), who suggested that early adopting users having higher reputation was due to these members joining the community as experts. Instead, we find that early adopting users were able to accumulate a portfolio of answers while there were fewer competing answerers. While Virendor attributes this fortune to luck, it may instead point toward a different type of expertise and technical knowledge that exists outside of the platform. Those who already had engagement with the programming community were better able to leverage their position to secure high status on the Stack Overflow, a task that may feel very difficult for newcomers to the platform. In this respect, old hierarchies are reinforced by the system of metrics, and those already in the know at the inception of the platform had a substantial advantage.

Farm Easy Questions

Gaining reputation from answering questions on Stack Overflow is deceptively difficult. The platform is known for having a very fast pace, with very low average response times. Overall, the platform discourages repetitive questions, seeing its purpose more as building a repository of programming knowledge (Hillman et al., 2021). In practice this means that repetitive or duplicated questions are often deleted.

> *[...]those are the sort of people that have ... you know, they've just built up reputation for answering, you know, mundane questions over and over again, that aren't ... that don't really contribute that well, because they've been answered before [...](warpwiz, 31 year old male).*

In this excerpt, warpwiz discusses how some high reputation users have engaged in a cherry-picking process, whereby they seek out simple and repetitive questions to answer, in order to maximize reputation gain. Other studies have also noted that

users feel there is greater reward in contributing to easy topics (Vadlamani & Baysal, 2020).

This observation was made by other interviewees:

sometimes with the gamification part of the site, people prefer to get some 20, 30, 40 points by answering the same answer five times a day (Octave_M, 36-year-old male).

Octave_M makes a similar observation to warpwiz, and attributes this to the gamification mechanics. By 'gamification' here, Octave_M is referring to the system of reputation and badges. While the reputation system cannot filter between someone using the same knowledge in different places, and indeed novice interlocutors will award upvotes for repeated information if it is useful to them, expert peers like Octave_M and warpwiz do not judge this behavior to be indicative of expertise. Instead, they value contributors' ability to solve novel problems and transfer experience between situations. This maps well to the theory put forward by Baltes and Diehl (2018) that software developers having expertise is understood as depth of knowledge in their language area as well as a broader knowledge that encompasses general strategies and concepts in software.

When talking to our informant, Janet, about why she does not spend as much time answering questions on Stack Overflow as she once did, she told us:

[...] the speed of answers on Stack Overflow specifically is kind of insane. Someone could ask a question and if it's answerable it will have an answer within half an hour. So, if you're looking at something from an hour ago someone else has answered it (Janet, 59-year-old female).

To be successful at the question farming strategy, one must be able to respond very quickly, and have time to invest in monitoring for questions that are relevant. Earning reputation from answering questions is in general quite difficult. Stack Overflow has a reputation for speed, with many questions answered exceptionally quickly; one study estimated that the median answer time for a question on Stack Overflow was just 11 minutes (Mamykina et al., 2011). This means that taking the time and effort to answer a more difficult question makes it likely that someone else may be working on an answer at the same time, and if that person answers sooner, your own effort goes to waste. In part this may contextualize the claim by Posnett et al. (2012) that users do not seem to grow expertise over time; if answers are 'sniped' very easily, this would discourage users from taking on questions that stretch their own programming abilities as there would potentially be no reward for doing so.

Edit, Edit, Edit

Question and answer posts on Stack Overflow can be edited by anyone, including by people with no account on Stack Overflow. Users under 2,000 reputation, and people editing without an account, must have their edits approved by another user with sufficient reputation. Users are able to gain reputation by having suggested edits accepted

(Stack Overflow, 2011). There are limits on how much reputation a user can earn from editing, and once a user has 1,000 reputation points earned through accepted edits they may no longer earn reputation this way. Once a user has gained 2,000 reputation points, they no longer require their edits to be accepted to be applied, but they also can no longer gain reputation from editing.

For new users, editing can be an accessible way to quickly progress to a reputation level where they have more privileges. Our interview informants noted this as a common strategy:

> there are people who will go for reputation through edit and through other odd ways of getting points (Brian, 59 year old male).

For some, that might involve using apps or extensions to help locate suitable posts for editing in the large myriad of posts that grow daily. There are many such apps designed to semi-automate certain kinds of edits, for example, users can use predefined queries on the Stack data explorer to search for common grammatical errors. This kind of mastery of the materiality of the platform is reminiscent of the behavior of Wikipedians, who also need to grapple with complex systems to engage in knowledge making processes (H. Ford & Wajcman, 2017).

However, many who do achieve reputation points through edits are not using their programming knowledge to contribute. While Stack Overflow discourages trivial editing, it does encourage editing to fix grammatical errors and other issues relating to formatting and presentation (*Privileges - Edit Posts*, 2023). In theory this means that someone can know very little about a programming topic but still acquire reputation, and related editing badges, in that topic. This is reminiscent of ways in which other platforms simultaneously encourage and discourage the same activity, discursively positioning some instances of the same action as gaming while other instances are seen as legitimate (Petre et al., 2019). The line between a trivial edit and a meaningful edit are very much open to interpretation. Other research has noted that these kinds of editing behaviors can introduce technical errors or inaccuracies over time. Mondal et al. (2021) discuss the issues that arise when a poster rejects an edit (termed a 'rollback inconsistency'), noting that rollbacks which revert many versions can accidentally restore a the technical content of an answer to an incorrect state. Mondal et al. (2021) identify that tensions around particular conventions of writing can cause these kinds of rollbacks, for example, differences of opinion in how code snippets are to be formatted.

In summary, the fastest way to earn reputation and to grow a user account does not require any programming knowledge at all. Instead, a user gains their reputation through acculturation to the practices and policies of the platform, gained by learning enough about the infrastructure to become a successful editor. This is like Wikipedia, where being able to navigate the platform hierarchies and policies are key to successful participation in editing (Ford & Wajcman, 2017; Marwick, 2013). While editing is not the only way to earn reputation, it is certainly a faster and lower-risk way of growing a reputation score than writing answers. In this respect, it is very difficult to claim that a reputation score alone could be a proxy for expertise.

Conclusion

In this paper, we have explored the complexities surrounding the recognition of expertise on Stack Overflow in relation to its ability to reconfigure existing hierarchies of expertise. We have found that, while Stack Overflow may help to create expertise recognition opportunities that disrupt western-centric orderings, it fails to disrupt existing hierarchies of recognizing technical expertise. In part, these metrics fail to reconfigure expertise because there are substantial rewards for engaging in the game of performing well on the metrics, and for leveraging knowledge of the policies and infrastructures that govern the platform. Such observations have applications for other measures of expertise, such as h-indexes which are often used as a proxy for researcher effectiveness.

Stack Overflow uses two different metrics that are discursively positioned as representing expertise: badges and reputation. However, these metrics arguably do more to represent community trust and engagement rather than knowledge. Since badges are more linked to specific tags and topics, they are a better reflection of subject knowledge, but they still require a degree of interpretation. While these metrics may superficially point to users who have knowledge in particular programming areas, it requires substantial knowledge of platform vernaculars and customs to determine if the activity of the user really aligns with the behavior of an expert in that topic. Equally, high reputation may not even be the result of programming knowledge. Users can exploit other means of attaining reputation that do not rely on having any subject specific knowledge at all, like through contributing grammatical and formatting edits to answers. While metrics superficially legitimize certain users as experts, or as trusted, the extent to which these metrics are accurate must be subjected to scrutiny. The same level of scrutiny is required for many ranking metrics; incentives within the system and tactics for gaming metrics can easily undermine their reliability.

Although reputation scores can be seen as peer-assessments of professional performances, their quantified, abstract, and detached nature can obscure the specifics of what is being evaluated, who is doing the evaluation, and the criteria used. In this regard, expertise is falsely rendered objective through metrics. However, these very characteristics allow scores to travel easily across contexts, making Stack Overflow reputation scores a hands-on tool to convey expertise in professional contexts beyond Stack Overflow. Stack Overflow facilitates the participation of people in sharing programming knowledge from across the globe, significantly lowering barriers to access information that has a very real and material impact on everyday work life. For some of our informants, this has had a substantial impact on career trajectory and professional opportunities.

However, it remains to be seen whether these mechanics can live up to their claim of being democratic or of representing expertise. Accumulating reputation functions as an investment; a good portfolio of simple but common answers allows one to have a bank of posts that will passively generate reputation over time. Astute investment early in the lifecycle of the platform has allowed some users to reap substantial reward. The side effect of this early adopter benefit is a reinforcement of existing norms and hierarchies. Reputation gain awards privileges, and greater privileges on the platform allow a user more control and influence over content. As a result, reputa-

tion translates directly into greater power within the structure of the platform. This gives those early adopters a greater ability to gatekeep the expertise of others. Similarly to Wikipedia, the power structure makes it difficult to disrupt masculine-coded ways of working on the platform (H. Ford & Wajcman, 2017). With many complex layers building up the platform infrastructure, gendered logics underpinning the distribution of power and reputation can become obfuscated under a veneer of openness.

Programming expertise is uniquely entwined with its materiality. The exact code that a user might place in a question or answer can be replicated and run by another person by copying the code into their own programming environment. Users can judge whether an answer was correct not by their own knowledge, but by testing the code and seeing if it produces the desired outcome. This potentially makes communication the most important tool at the disposal of a high-performing member of Stack Overflow; being able to communicate how and why one's solution works helps to translate this knowledge to a non-expert audience, making it more likely that reputation can be earned. However, advanced programming knowledge, such as how to write easy to maintain code (Baltes & Diehl, 2018) can only be recognized by another expert. For an outsider using the metrics to identify someone with advanced knowledge, there is no way to know – an upvote is an upvote.

Similar problems exist in other ranking metrics; it is necessary to understand how something has been measured to determine if the measurement is useful in context. Like many metric-based measures of expertise, interpreting the metrics provided by Stack Overflow requires contextual sensitivity. Similarly to how an expert peer can examine a researcher's oeuvre to see if the given h-index is truly reflective of their work, an expert peer on Stack Overflow is able to use contextual information, like tags and user activity, to determine to what extent that user has programming expertise. The issue remains that these metrics always demand an interpretation, and a naïve reading of the metrics is open to manipulation. This calls into question the legitimacy of using such metrics to allocate reward and recognition.

Data Availability Not Applicable.

Code Availability Custom script used to collect user information is available on Github, at https://github.com/paddlelab/SOCDEX.

Declarations

Conflict of Interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Ethics Approval This study has been approved by the Swedish Ethical Review Authority.

Consent to Participate Interview informants signed an informed consent sheet and had the opportunity to ask questions or withdraw their data. All personal data has been handled in accordance with the EU General Data Protection Regulation (GDPR).

Consent for Publication Consent for publication is given.

References

All Sites—Stack Exchange (2023). Stack Exchange. https://stackexchange.com/sites?view=list#questions.

Almog, R., & Kaplan, D. (2017). The nerd and his discontent: The Seduction Community and the logic of the game as a Geeky Solution to the challenges of Young Masculinity. *Men and Masculinities*, *20*(1), 27–48. https://doi.org/10.1177/1097184X15613831.

Atwood, J. (2008, December 23). Vote Fraud and You. Stack Overflow Blog. https://stackoverflow.blog/2008/12/23/vote-fraud-and-you/.

Atwood, J. (2009, March 21). More Voting Anomalies. Stack Overflow Blog. https://stackoverflow.blog/.

Baltes, S., & Diehl, S. (2018). Towards a theory of software development expertise. Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, 187–200. https://doi.org/10.1145/3236024.3236061.

Bartneck, C., & Kokkelmans, S. (2011). Detecting h-index manipulation through self-citation analysis. *Scientometrics*, *87*(1), 85–98. https://doi.org/10.1007/s11192-010-0306-5.

Brooke, S. (2021). Trouble in programmer's paradise: Gender-biases in sharing and recognising technical knowledge on Stack Overflow. *Information Communication & Society*, *0*(0), 1–22. https://doi.org/10.1080/1369118X.2021.1962943.

Caliandro, A. (2018). Digital methods for Ethnography: Analytical concepts for ethnographers exploring Social Media environments. *Journal of Contemporary Ethnography*, *47*(5), 551–578. https://doi.org/10.1177/0891241617702960.

Carraro, V., & Wissink, B. (2018). Participation and marginality on the geoweb: The politics of non-mapping on OpenStreetMap Jerusalem. *Geoforum*, *90*, 64–73. https://doi.org/10.1016/j.geoforum.2018.02.001.

Connell, R. (2005). *Masculinities* (2nd ed.). Routledge. (Original work published 1995).

DiMaggio, P., Hargittai, E., Neuman, W. R., & Robinson, J. P. (2001). Social implications of the internet. *Annual Review of Sociology*, *27*, 307–336.

Faisal, M. S., Daud, A., Akram, A. U., Abbasi, R. A., Aljohani, N. R., & Mehmood, I. (2019). Expert ranking techniques for online rated forums. *Computers in Human Behavior*, *100*, 168–176. https://doi.org/10.1016/j.chb.2018.06.013.

Faulkner, W. (2000). Dualisms, hierarchies and gender in Engineering. *Social Studies of Science*, *30*(5), 759–792. https://doi.org/10.1177/030631200030005005.

Ford, H., & Wajcman, J. (2017). Anyone can edit, not everyone does: Wikipedia's infrastructure and the gender gap. *Social Studies of Science*, *47*(4), 511–527.

Ford, D., Smith, J., Guo, P. J., & Parnin, C. (2016). Paradise unplugged: Identifying barriers for female participation on stack overflow. Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering, 846–857. https://doi.org/10.1145/2950290.2950331.

Ford, D., Harkins, A., & Parnin, C. (2017, November 20). Someone Like Me: How Does Peer Parity Influence Participation of Women on Stack Overflow? 2017 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC). IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC). https://doi.org/10.1109/VLHCC.2017.8103473.

Gibbs, M., Meese, J., Arnold, M., Nansen, B., & Carter, M. (2015). #Funeral and Instagram: Death, social media, and platform vernacular. *Information Communication & Society*, *18*(3), 255–268. https://doi.org/10.1080/1369118X.2014.987152.

Gillespie, T., & Information (2017). *Communication & Society*, 20(1), 63–80. https://doi.org/10.1080/13 69118X.2016.1199721.

Graham, T., & Rodriguez, A. (2021). The sociomateriality of Rating and Ranking devices on Social Media: A Case Study of Reddit's Voting practices. *Social Media + Society*, 7(3), 20563051211047667. https:// doi.org/10.1177/20563051211047667.

Graham, M., Hogan, B., Straumann, R. K., & Medhat, A. (2014). Uneven geographies of user-generated information: Patterns of increasing informational poverty. *Annals of the Association of American Geographers*, *104*(4), 746–764. https://doi.org/10.1080/00045608.2014.910087.

Hara, N., & Sanfilippo, M. R. (2017). Analysis of roles in engaging contentious online discussions in science. *Journal of the Association for Information Science and Technology*, *68*(8), 1953–1966. https:// doi.org/10.1002/asi.23850.

Hillman, T., Seredko, A., Nivala, M., & Osborne, T. (2021). Knowledge Sharing in Tension: Interacting and Documenting on Stack Overflow. Proceedings of the Eighth ACM Conference on Learning @ Scale, 331–334. https://doi.org/10.1145/3430895.3460981.

Hindman, M. (2008). *The myth of digital Democracy*. Princeton University Press.

Hine, C. (2015). Ethnography for the internet: Embedded, Embodied and Everyday. *Ethnography for the internet*. Taylor & Francis Group.

Huang, C., Yao, L., Wang, X., Benatallah, B., & Sheng, Q. Z. (2017). Expert as a Service: Software Expert Recommendation via Knowledge Domain Embeddings in Stack Overflow. 2017 IEEE International Conference on Web Services (ICWS), 317–324. https://doi.org/10.1109/ICWS.2017.122.

Huang, C., Yao, L., Wang, X., Benatallah, B., & Zhang, X. (2020). Software expert discovery via knowledge domain embeddings in a collaborative network. *Pattern Recognition Letters*, *130*, 46–53. https://doi.org/10.1016/j.patrec.2018.10.030.

Joshi, A. (2014). By whom and when is women's expertise recognized? The interactive effects of gender and education in Science and Engineering teams. *Administrative Science Quarterly*, *59*(2), 202–239.

König, R. (2013). Wikipedia. Information, Communication & Society, 16(2), 160–177. https://doi.org/10 .1080/1369118X.2012.734319.

Mamykina, L., Manoim, B., Mittal, M., Hripcsak, G., & Hartmann, B. (2011). Design lessons from the fastest Q&A site in the west. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2857–2866. https://doi.org/10.1145/1978942.1979366.

Marwick, A. E. (2013). Status update: Celebrity, publicity, and branding in the social media age.

Mau, S. (2019). The metric society: On the quantification of the social.

Meuser, M. (2007). Serious games – competition and the homosocial construction of masculinity. *NORMA*, *2*(1), 39–51. https://doi.org/10.18261/ISSN1890-2146-2007-01-04.

Mondal, S., Uddin, G., & Roy, C. K. (2021). Rollback Edit Inconsistencies in Developer Forum. 2021 IEEE/ACM 18th International Conference on Mining Software Repositories (MSR), 380–391. https://doi.org/10.1109/MSR52588.2021.00050.

Nivala, M., Seredko, A., Osborne, T., & Hillman, T. (2020). Stack Overflow – Informal learning and the global expansion of professional development and opportunities in programming? 2020 IEEE Global Engineering Education Conference (EDUCON), 402–408. https://doi.org/10.1109/EDUCON45650.2020.9125165.

Oravec, J. A. (2019). The Dark side of academics? Emerging issues in the Gaming and Manipulation of Metrics in Higher Education. *Review of Higher Education*, *42*(3), 859–877. https://doi.org/10.1353/rhe.2019.0022.

Petre, C., Duffy, B. E., & Hund, E. (2019). Gaming the System: Platform paternalism and the politics of algorithmic visibility. *Social Media + Society*, *5*(4), 2056305119879995. https://doi.org/10.1177/2056305119879995.

Posnett, D., Warburg, E., Devanbu, P., & Filkov, V. (2012). Mining Stack Exchange: Expertise Is Evident from Initial Contributions. 2012 International Conference on Social Informatics, 199–204. https://doi.org/10.1109/SocialInformatics.2012.67.

Privileges (2023). Stack Overflow. https://stackoverflow.com/help/privileges.

Privileges—Edit posts (2023). Stack Overflow. https://stackoverflow.com/help/privileges/edit.

Procaci, T. B., Nunes, B. P., Nurmikko-Fuller, T., & Siqueira, S. W. M. (2016). Finding Topical Experts in Question & Answer Communities. 2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT), 407–411. https://doi.org/10.1109/ICALT.2016.68.

Reagle, J. (2013). Free as in sexist? Free culture and the gender gap. *First Monday*. https://doi.org/10.5210/fm.v18i1.4291.

Rieder, B., Matamoros-Fernández, A., & Coromina, Ò. (2018). From ranking algorithms to 'ranking cultures': Investigating the modulation of visibility in YouTube search results. *Convergence*, *24*(1), 50–68. https://doi.org/10.1177/1354856517736982.

Stack Overflow Developer Survey (2022). (2022). Stack Overflow. https://survey.stackoverflow.co/2022/?utm_source=social-share&utm_medium=social&utm_campaign=dev-survey-2022.

Stack Overflow (2011, January 25). Answer to "How do suggested edits work?" Meta Stack Exchange. https://meta.stackexchange.com/a/76284.

Users (2023). Stack Overflow. https://stackoverflow.com/users.

Vadlamani, S. L., & Baysal, O. (2020). Studying Software Developer Expertise and Contributions in Stack Overflow and GitHub. 2020 IEEE International Conference on Software Maintenance and Evolution (ICSME), 312–323. https://doi.org/10.1109/ICSME46990.2020.00038.

van Bevern, R., Komusiewicz, C., Niedermeier, R., Sorge, M., & Walsh, T. (2016). H-index manipulation by merging articles: Models, theory, and experiments. *Artificial Intelligence*, *240*, 19–35. https://doi.org/10.1016/j.artint.2016.08.001.

Wang, S., Chen, T. H., & Hassan, A. E. (2020). How do users revise answers on Technical Q&A websites? A case study on Stack Overflow. *IEEE Transactions on Software Engineering*, *46*(9), 1024–1038. https://doi.org/10.1109/TSE.2018.2874470.

What is a bounty? How can I start one? - Help Center. (2023). Stack Overflow. https://stackoverflow.com/help/bounty.

What is reputation? How do I earn (and lose) it? - Help Center. (2022). Stack Overflow. https://stackoverflow.com/help/whats-reputation.

Zhou, J., Wang, S., Bezemer, C. P., & Hassan, A. E. (2020). Bounties on technical Q&A sites: A case study of Stack Overflow bounties. *Empirical Software Engineering*, *25*(1), 139–177. https://doi.org/10.1007/s10664-019-09744-3.

Authors and Affiliations

Tanya Osborne[1] · Markus Nivala[1] · Alena Seredko[1] · Thomas Hillman[2]

✉ Tanya Osborne
tanya.osborne@gu.se; tanya.theory@pm.me

[1] Department for Education, Communication and Learning, Gothenburg University, Gothenburg, Sweden

[2] Department of Applied Information Technology, Gothenburg University, Gothenburg, Sweden