

Reputation Management Systems

This case study on the various uses of reputation management systems on social media was produced as part of [The Justice Collaboratory](#) at Yale Law School's [Social Media Governance Initiative](#) Spring '24 Lab by students [Tanya Aggarwal](#), [Jessica Araujo](#), [Michael Bochkur Dratver](#), and [Julia Gaburo](#) in collaboration with [New_ Public](#). This document collects and documents the Lab's work on the topic.

CONTENTS

Key Insights	P2
From the Field	P5
What We Don't Know	P9

TL;DR

Reputation management, signaling, and verification are increasingly important elements of many online platforms and social media sites.

According to the Pew Research Center, reputation management includes any activity that users take to monitor or curate their identity or learn about others' identities online (Pew). Reputation management systems, which can often take the form of earned badges, qualifiers on a profile, visual logos or insignia, or stated certifications, allow users to differentiate themselves through an established and recognized online identity. Much research has been done on the significance of reputation management systems and how they impact user behavior and platform transactions.

Key Insights

This literature review discusses the role of reputation management systems in online platforms, including how they can improve the platform's functioning and encourage users to behave in a prosocial manner (i.e., exhibiting behavior that benefits one

or many other people). The literature review establishes a psychological theoretical framework and includes recommendations supported by research for explaining and designing a reputation management system.

Theoretical Framework

Online public spaces serve as dynamic platforms where individuals can create connections, form communities, share experiences, and acquire new knowledge. These virtual environments depend on the active participation of their users, whose willingness to contribute is essential to the digital community experience.

Reputation management systems play a crucial role in encouraging and enhancing this contribution by helping to guide decision-making, assessing trustworthiness, organizing social interactions, and fostering a positive communal environment (Lampe, 2012). Moreover, social psychological frameworks, such as the Self-Determination Theory and the Social Identity Theory, can provide valuable insights into the motivations behind why individuals voluntarily engage, share, and contribute in online communities.

Understanding these underlying motivations can help us better design and manage

online spaces that promote meaningful interactions and sustained participation.

Self-Determination Theory (SDT) emphasizes the role of intrinsic motivation, which is driven by an interest or enjoyment in the task itself, and extrinsic motivation, which is driven by external rewards, or pressures, and how fulfilling the basic psychological needs of autonomy, competence, and relatedness can foster more engaged and satisfied community members (Cavusoglu et al., 2021; Deci et al., 1994). For example, in an online learning community, if members feel they have control over their learning (autonomy), are able to effectively contribute to discussions (competence), and feel a sense of belonging (relatedness), they are more likely to participate actively and value the community. Id.

Similarly, Social Identity Theory (SIT) suggests that individuals engage in online communities to affirm their group identities, adhering to group norms and enhancing their social identity (Karasawa, 1991; Tajfel & Turner, 1979; van Kleef et al., 2019). This theory

explains, for instance, why gamers on a forum might actively share advanced gaming strategies and defend their group against critics, reinforcing their identity as expert gamers and deepening their connection to the community.

In the context of public online spaces, prosocial behavior can be significantly influenced by reputation management systems (RMS), which are designed to incentivize and recognize contributions through visible rewards like badges or points (Cavusoglu et al., 2021; Zhang et al., 2020). These systems leverage extrinsic motivation by rewarding users for helpful behavior, such as providing tech support or engaging in community moderation, which benefits the community while simultaneously enhancing the user's reputation. The visibility of such rewards can motivate others to contribute, driven by the desire for similar recognition. Moreover, as extrinsic motivations can transition into intrinsic motivations through the process of internalization, as explained by SDT, users may begin contributing for recognition but continue doing so because the activity aligns with their personal values or becomes a fulfilling part of their identity (Deci & Ryan, 2000). Thus, RMS can transform initial extrinsic motivations into enduring intrinsic motivations, fostering long-term engagement and a vibrant, self-sustaining community.

In an online environment, RMS can effectively meet psychological needs for autonomy, competence, and relatedness, as outlined by SDT. First, clear guidelines and structured levels of achievement are crucial to addressing issues of autonomy within RMS. (Cavusoglu et al., 2021).

Such systems allow users to set well-defined goals and select their degree of involvement, aligning their activities with personal objectives while navigating the system effectively. Second, according to Amichai-Hamburger et al., (2016), one of the key reasons users choose to lurk rather than participate actively in online discussions is a lower level of self-efficacy or feeling incapable of contributing effectively. Furthermore, studies have shown that higher levels of self-efficacy lead to more positive interactions on social networking sites (Gangadharbatla, 2013). Hence, badges and rewards within the RMS can strengthen competence and self-efficacy by visibly acknowledging users' aptitudes and encouraging continued engagement (Cavusoglu et al., 2021). Lastly, the psychological need for relatedness can be fulfilled by creating small, focused communities, as the type of environment is also an important factor in online participation (Amichai-Hamburger et al., 2016). Hwang and Foote (2021) found that users feel their needs are met in smaller communities as they provide specialized spaces that attract members with niche interests and promote a collective identity. According to the same study, small communities enhance user bonding by increasing the likelihood of repeated interactions between the same individuals, thus fostering deeper social connections. Therefore, small communities help cultivate a sense of belonging and shared purpose, strengthening users' connections within the community.

This approach highlights how integrating SDT and SIT can deepen our understanding of online community dynamics, underscoring the significance of autonomy,

competence, and relatedness in driving meaningful participation and building strong, engaged communities.

Impacts on User Behavior & Function of Online Communities

Reputation management systems can help platforms establish and reinforce social norms by awarding reputation-building status to users who exemplify the platform norms in a meaningful way. Having and communicating clear, well-established, reliable norms has been proven to deliver benefits to platforms like the reduction of harassment and the increase of group participation among new members (Matias, 2019). Researchers analyzed conversations on Reddit's r/Science community to observe participation and rule compliance among new users and assess whether making norms more visible to platform users influenced participation rates and rule compliance in online discussions. Id. The study found that posting or otherwise demonstrating clear norms for a platform can increase rule compliance by newcomers by more than 8% and boost newcomer participation by 70% on average. Id. These findings would suggest making community norms visible not only prevents unruly behavior but also encourages more people, especially newcomers, to participate in online discussions. Id. Since reputation signals are a mechanism through which platforms can make their norms more clear and visible, platforms can use them to realize positive change in user behavior. According to experts in the space, it is also

important for platforms to ensure their community norms are positive ones — in some communities, acting “badly” is a recognized norm, and platforms must be sure that their reputation management systems are not helping to further these harmful norms.

Beyond the enforcement of norms, reputation signals can perform a subtle socializing function that clarify what desirable behaviors are and by mitigating harm from anti-social actions. Reputation signals help organize online interactions, socialize new members, sanction normative behavior, and help construct “sociotechnical capital,” in other words, social capital inside a system that involves both humans and their interactions with computers (Kazmer 2006). Conversations with experts illustrate how this socializing function works on the knowledge-sharing platform, Stack Overflow. On that site, reputation signals are provided according to actions users take on the site that are closely tied to the desired core function of the site, which is to encourage engagement with legitimate programming questions and discourage misuse. Providing users reputation benefits for actions like answering questions and upvoting good answers helps to reinforce the kind of social behavior desired on the site.

In addition to benefits to user behavior, reputation signals can also deliver economic benefits to a platform by improving how users transact on an online platform, encouraging higher-value transactions and helping sellers introduce new products based on their existing reputation (Huang et al., 2023). Conversations with experts in

the space familiar with Spotify's reputation management and verification systems pointed out that verifying identity and reputation in online communities is critical for healthy economic transactions on these sites because brands and artists need to know unequivocally that their identities will be represented accurately in order to transact on the platform. Additionally, researchers analyzed the impact of the introduction of the Superhost badge system on the platform's bookings and revenue of Airbnb's property booking platform and found that proper reputation management delivered positive impacts to the platform's overall booking numbers and revenue levels. Id. These positive impacts were achieved in two ways. First, the Superhost badge facilitated bookings of hosts' properties new to the platform that had

new ratings by serving as a positive signal for their latent quality. Id. Essentially, if a property was introduced by a Superhost, it was able to overcome the initial hurdle to booking that resulted from few ratings and reviews leading to more properties available for booking overall. While this result was positive for Superhosts, it could represent a crowding out of properties without a Superhost designation. Second, the Superhost badge enhanced the prominence of the properties of high-quality hosts, improving the chances of buyers finding a suitable match and making a booking. Id. This study illustrates the positive benefits to site transactions that reputation signals can provide, even when the platform already incorporates ratings and reviews.

From the Field

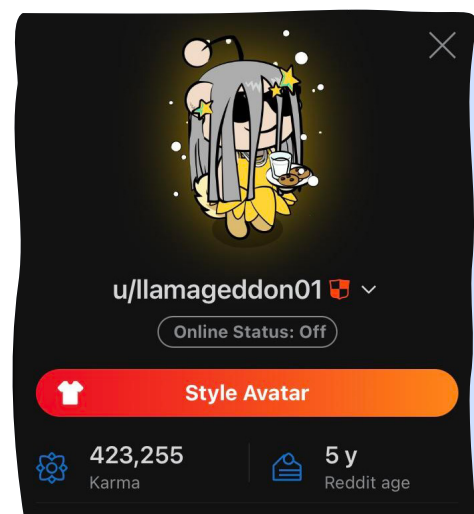
Here, we pull examples from the field of existing ways that platforms have leveraged reputation management systems.

Reddit Karma Points

Reddit is a digital platform where users can post content, participate in discussions, and bond over mutual interests in specific forums known as subreddits. It features a system known as karma points, determined by the total positive votes received from your posts and comments. Each upvote boosts your karma, while each downvote reduces it. Karma serves as a measure of one's contributions to Reddit, enhancing their credibility and impacting

the prominence of their posts and comments, even though it doesn't have substantial value.

THE IMAGE BELOW PRESENTS A HIGHLY ACTIVE REDDIT USER WHO HAS ACCUMULATED OVER 400,000 KARMA POINTS.



Facebook Group Badges

Facebook groups are a component of the Facebook social media platform, providing virtual spaces where users with similar interests can come together to share content and engage in discussions relevant to the group's focus. In 2018, Facebook introduced group badges to enhance engagement and strengthen community ties. These visual indicators next to a member's name indicate various roles or acknowledge significant contributions. Notable badges include the Valued Responder Badge for those who provide helpful answers, Conversation Starter and Booster Badges for initiating or enhancing discussions, the Rising Star Badge for active new members making positive contributions, the Visual Storyteller Badge for members who augment discussions with impactful visual content, the Greeter Badge for those welcoming newcomers, and the Group Expert Badge for members with specialized knowledge relevant to the group's theme.



Valued responder

You and 19 others consistently share answers that people value in Okehampton Information Hub. This badge displays for one month and only group members can see it.



See All Valued Responders

Stack Overflow Participation Badges

Stack Overflow is a Q&A website where users can ask and answer questions related to programming and software development. On Stack Overflow, reputation points are earned when other users upvote your questions, answers, or edits, illustrating the community's trust in your expertise and contributions. Reputation points can grant privileges like commenting on others' posts, voting, and accessing moderation tools. This system incentivizes quality contributions and helps maintain the quality of the site. Badges are awarded for achieving certain milestones or performing specific actions on the site, such as editing posts, asking good questions, or providing helpful answers. There are three types of badges: bronze (for beginners), silver (for more advanced achievements), and gold (for significant contributions). For instance, to earn the bronze "Talkative" participation badge, the user will have to "Post 10 messages, with 1 or more starred, in chat," and to earn the silver "Outspoken" participation badge, one will need to "Post 10 messages in chat starred by 10 different users."

Intro



Member of Growth and Monetisation for Facebook Group Admins since May 19, 2020

Group Badges



Admin



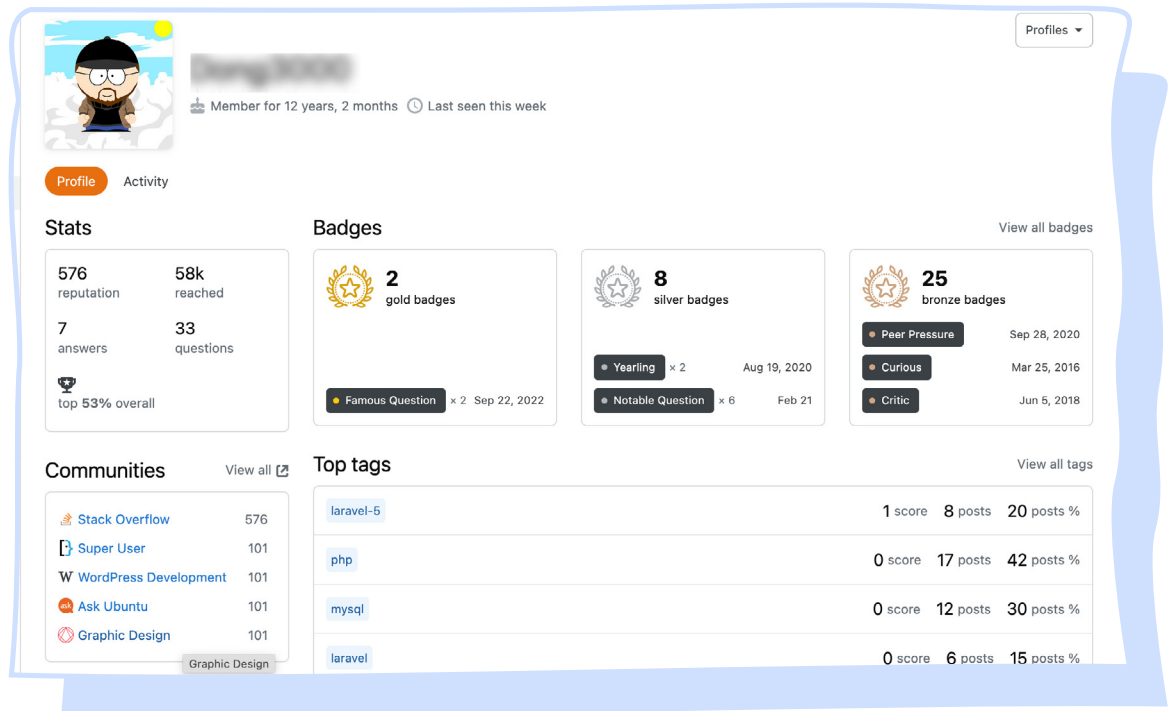
Visual Storyteller



Conversation Starter

THE IMAGES HERE DISPLAY THE VARIOUS BADGES AWARDED TO A MEMBER OF A SPECIFIC FACEBOOK GROUP.

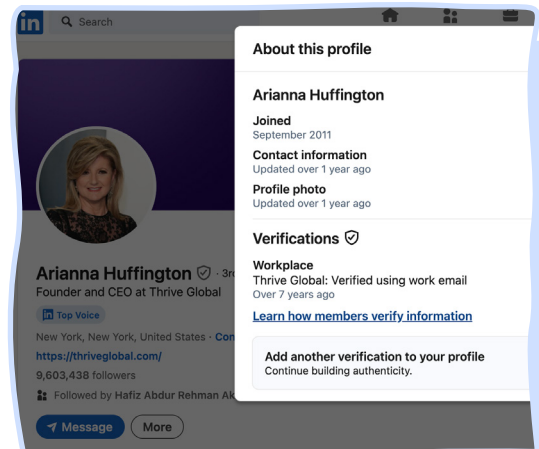
THE IMAGE HERE PRESENTS A USER'S PROFILE ON STACK OVERFLOW, HIGHLIGHTING THEIR BADGES AND REPUTATION POINTS.



LinkedIn Verified Member

LinkedIn is a professional networking platform that allows users to connect with colleagues, industry peers, and potential employers. It serves as a digital resume and networking space where users can share their professional experiences, achievements, and insights and discover job opportunities and business news. LinkedIn employs a verification process that enhances authenticity through several methods: Identity verification (using a government-issued ID and selfie for those in select regions including the U.S., Canada, and Mexico, or an NFC-enabled passport in other countries); workplace verification (via work email, Microsoft Entra Verified ID, or a LinkedIn Learning license for select companies), and Educational Institution verification (using an education email or LinkedIn Learning license). These processes, managed by trusted partners, aim to confirm users' identities, employment, and educational affiliations. Verification is

voluntary and free, and users have control over the visibility of their verifications and can remove them at any time. There is ongoing discussion in the expert community about whether identity verification qualifies as active reputation management.



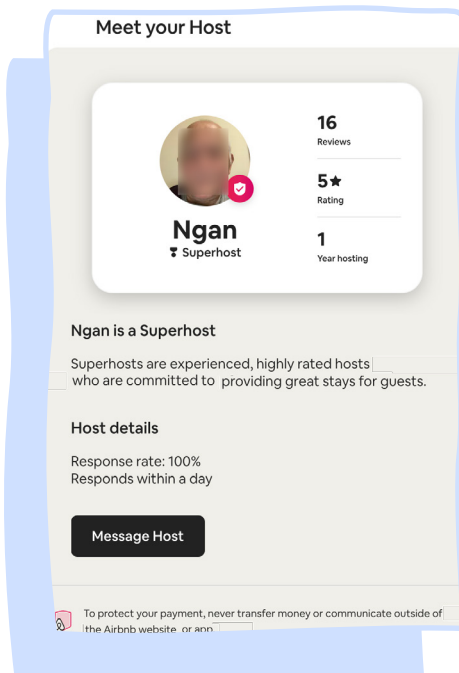
THE IMAGE ABOVE PRESENTS AN EXAMPLE OF A WORKPLACE VERIFICATION SHOWN ON THE USER'S PROFILE.

Airbnb Superhost Program

Airbnb is an online platform that offers an alternative to traditional hotels by allowing people to rent out their properties or spare rooms to travelers. In 2016, Airbnb introduced the Superhost badge to distinguish exceptional hosts who provide outstanding hospitality across their listed properties. While the platform already used ratings and reviews as quality indicators, the Superhost badge offers a more comprehensive evaluation by assessing hosts' portfolios. To earn this badge, which is reviewed every three months, hosts must meet several criteria: a minimum of 10 stays or 100 nights over at least three stays in the past year, an overall rating of 4.8 or higher, a cancellation rate of 1% or lower, and a response rate of 90% or higher. Once earned, the badge is displayed on all of a host's property listings, and any new properties listed during the period they hold the badge will also feature it automatically.

Pinterest Verified Merchant Program:

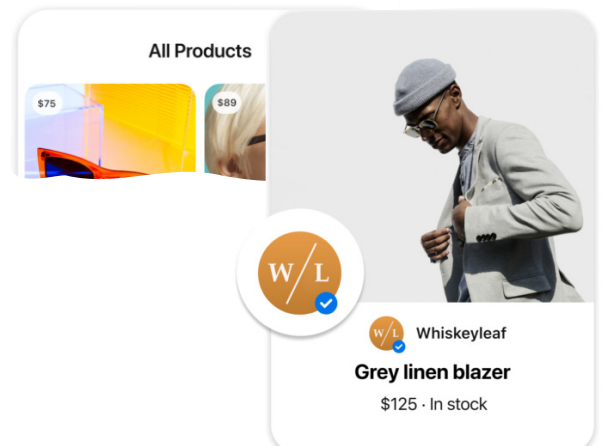
Pinterest's Verified Merchant Program is designed to help users discover and purchase from reputable brands on the platform. Verified merchants receive a blue checkmark, better visibility in shopping experiences, and access to exclusive features. Brands need to meet specific eligibility requirements, including having an established Pinterest account and website and complying with Pinterest's merchant guidelines. Applications are reviewed manually. In addition to the benefits of legitimate reputation and ability to better connect with customers interested in shippable products, the Verified Merchant status allows sellers greater insight into sales conversions and dynamic ad retargeting. The feature was launched on the site in 2020 when Pinterest saw an uptick in searches for at-home solutions during the COVID pandemic lockdown.



THE IMAGE DISPLAYED HERE FEATURES AN AIRBNB HOST WHO HAS EARNED THE "SUPERHOST" BADGE.



THE IMAGES HERE SHOW EXAMPLES OF A VERIFIED MERCHANT PROFILE ON PINTEREST.



What We Don't Know

What is the best way to create meaningful norms for users on a platform?

Reputation management systems benefit from clear, visible, reliable norms and can be an effective way to enforce these norms. Our research made this fact clear, but did not delve deeply into mechanisms for creating norms that reflect community intent and goals for interaction on a platform. Additional information about creating norms would complement this research well.

Reputation management may create and amplify inequalities within the user base:

An important issue an expert highlighted was that reputation management systems may amplify entrenched hierarchies on platforms, and continue over time as users accumulate reputation. The core of the problem lies in the accumulation of reputation over time. For instance, users who participated in 2009 and engaged with basic questions can accumulate a considerable reputation as these questions continue to attract upvotes over the years, even if these users are no longer active. This ongoing accumulation of reputation from old posts can foster a social hierarchy. Although efforts have been made to mitigate this by ranking users based on their

reputation gained in more recent periods (such as the last 3 or 6 months) rather than on a lifetime accumulation, the issue of reputational inertia from past contributions remains unresolved.

There is likely a meaningful intersection between juries and tribunals and reputation management systems that is worth exploring further:

Another [SMGI Case Study](#) looks at the use of juries and tribunals on social platforms. An expert also pointed out that the inherent risk in any system is the potential for individuals to exploit loopholes for undeserved reputation benefits or even to lose reputation benefits unfairly. To counter this, the establishment of a jury to verify reputations and authenticate claims is important. For instance, within Reddit's framework, users accumulate points through comments or answers. However, the system allows for the reduction of points through downvotes if other users disagree with an answer. Introducing a jury system could be beneficial in promoting fairness and addressing potential biases in this process. This jury would enable users to challenge community decisions, providing an avenue for a thorough review of whether point deductions were warranted. Unfortunately, this intersection has not yet been studied, which limits our understanding of such a system.

Should reputation management systems award quality or quantity?

Many reputation management systems prioritize quantity, awarding points or badges based on the frequency of user participation and engagement. However, not all interactions contribute positively — some may involve low-quality contributions like spreading misinformation. This can lead to users being motivated to participate indiscriminately, focusing solely on enhancing their reputation to access specific benefits rather than contributing meaningful content. While numerous studies focus on increasing participation in these systems, there is a noticeable lack of research addressing the quality of such participation and it remains unclear which is the “better” route for a system to take.

—

© 2024 NEW_ PUBLIC. THIS WORK IS LICENSED UNDER A CREATIVE COMMONS ATTRIBUTION NONCOMMERCIAL 4.0 INTERNATIONAL LICENSE. CREDIT MUST BE GIVEN TO BOTH NEW_ PUBLIC AND THE AUTHOR OR AUTHORS OF THIS REPORT.

