Timeliness, Consensus, and Composition of the Crowd: Community Notes on X

Olesya Razuvayevskaya
The University of Sheffield
Sheffield, United Kingdom
o.razuvayevskaya@sheffield.ac.uk

Adel Tayebi
CY Cergy Paris University
NOVA University Lisbon
Paris, France
adel.tayebi@cyu.fr

Ulrikke Dybdal Sørensen
Aalborg Universitet
Aalborg, Denmark
ulrikkeds@ikp.aau.dk

Kalina Bontcheva
The University of Sheffield
Sheffield, United Kingdom
K.Bontcheva@sheffield.ac.uk

Richard Rogers
University of Amsterdam
Amsterdam, Netherlands
R.A.Rogers@uva.nl

Abstract

This study presents the first large-scale quantitative analysis of the efficiency of X's Community Notes, a crowdsourced moderation system for identifying and contextualizing potentially misleading content. Drawing on over 1.8 million notes, we examine three key dimensions of crowdsourced moderation: participation inequality, consensus formation, and timeliness. Despite the system's goal of collective moderation, we find substantial concentration effect, with the top 10% of contributors producing 58% of all notes (Gini Coefficient = 0.68). The observed consensus is rare—only 11.5% of notes reach agreement on publication, while 69% of posts receive conflicting classifications. A majority of noted posts (≈68%) are annotated as "Note Not Needed", reflecting the repurposing of the platform for debate rather than moderation. We found that such posts are paradoxically more likely to yield published notes (OR = 3.12). Temporal analyses show that the notes, on average, are published 65.7 hours after the original post, with longer delays significantly reducing the likelihood of consensus. These results portray Community Notes as a stratified, deliberative system dominated by a small contributor elite, marked by persistent dissensus, and constrained by timeliness. We conclude this study by outlining design strategies to promote equity, faster consensus, and epistemic reliability in community-based moderation.

CCS Concepts

• **Information systems** → **Crowdsourcing**; **Web applications**; **Social networks**; **Answer ranking**; **Trust**; **Reputation systems**; **Crowdsourcing**.

Keywords

fact-checking, community moderation, misinformation

1 Introduction

Online content moderation of misinformation typically relies on either automated detection or a manual approach, be it expert fact-checking or crowd-sourced moderation. As the volume of false information continues to grow, manual expert fact-checking is becoming an increasingly intractable task, requiring expertise across multiple languages, topics, and region-specific domains [31, 47]. Moreover, it carries risks of partisanship, bias and censorship, stemming from the judgments of a small group of experts [24]. In an effort to address scalability and bias concerns, social media platforms have increasingly turned to crowdsourcing as an alternative to expert fact-checking. This shift, started by Twitter (now X) with the introduction of Birdwatch in January 2021 to the users in the United States, was later extended worldwide and renamed Community Notes (CNs). This was followed by Meta's announcement[1] in early 2025 that it would discontinue its fact-checking program and adopt CNs.

The move toward crowd-sourced moderation signals a change in the philosophy of moderation from expert adjudication to one based on consensus building [18], despite the fact that CNs still rely on fact-checking sources, with around 33% of notes reported to cite at least one such source [6]. This change in philosophy has prompted a series of observations concerning potential issues with crowdsourced moderation. One concerns the prospect that it will facilitate biased or politically "asymmetric interventions" against misinformation [36]. Another is that CNs can be vulnerable to coordinated attacks, potentially suppressing helpful notes by downvoting them [3]. Along those lines, the polarised nature of online communities could make crowdsourcing itself difficult to achieve[51]. These concerns raise a central question: How effective are CNs in achieving the goal of crowdsourced content moderation?

To address this question, in this work, we undertake three complementary analyses. First, we examine the composition of the crowd in terms of potential concentration of note production by a select group of users. Second, we assess the extent to which CNs promote consensus or dissensus—both at the level of individual posts and within the crowd-based evaluation of notes themselves. Within this analysis, we also look into the case of the misuse of CNs as a debating platform. Third, we examine certain conditions of consensus-making, particularly the dependence on the timeliness of producing CNs. By addressing these questions, we can evaluate the rate of consensus building in crowdsourced moderation, specific conditions that can affect this goal and whether CNs are "crowdsourced" in practice or function more as a specialised moderation activity dominated by a narrow set of contributors.

[1] https://www.nytimes.com/2025/01/07/business/meta-community-notes-x.html

2 Background

2.1 Community Notes on X

Similar to Wikipedia and Reddit's models of community-driven moderation, the CN system on X relies on a community of users identifying misleading content or adding important context to the shared information, with the primary goal of surfacing high-quality information [14, 52]. To become part of the community, known as a *contributor*, there is a set of eligibility criteria to be met, such as a six-month user history and verified phone number to prevent multiple accounts [15]. Upon becoming a contributor, a user accrues 'influence' by rating notes to eventually earn the privilege of a *note writer*. To maintain writer status, a contributor must ensure that the majority of the notes they produce are rated as helpful by the community.

Eligible users write notes about the posts they consider misinformation or potentially misleading, with the option to add explanation tags concerning why the post should be 'noted'. They can also create notes to mark a post as non-misinformation, which some note writers do to counter another contributor's misinformation-flagging note. A note is then rated by the community as helpful or not, which is decisive for the note to be displayed. Only helpful notes that flagged misinformation are published under the post on the main platform. A note has two weeks to achieve a status, and within that period a note's status can change at any point. After that, the note's status is 'locked'.

All the notes are assigned the status 'needs more ratings' (NMR) when first created and must be rated at least five times before being considered for display. A note must not only be rated by enough contributors who consider it helpful; the contributors must also represent a 'diversity of perspectives'. This is X's method of reaching consensus through a so-called bridging algorithm [37, 54] rather than a simple majority of votes. Crucially, X does not incorporate any explicit attributes of contributors when inferring perspectives; 'diversity' is derived entirely from statistical patterns in how contributors have rated other notes. This process can be described as computational consensus defined not by direct deliberation or explicit representation, but by algorithmic classification.

Previous studies have argued that this computational approach to consensus building introduces complications, since coming to an agreement is not only a technical determination, but also a social task [19, 32]. The social aspect of consensus formation became apparent after X added a new rating category in 2023: NotHelpful_NoteNotNeeded (NNN). CN contributors use this feature to indicate that the post represents an opinion rather than a checkworthy claim, and therefore does not require a note. Despite the presence of this rating feature, some prefer to create a new note under the post, with the note text mentioning that the note is not needed. This behaviour is typically motivated by the desire to counter the other notes under the post that mark it as potentially misleading.

2.2 Crowdsourced content moderation

Despite the ongoing questions about the effectiveness of community-driven models for content moderation, such models are increasingly adopted by social media platforms in the form of an additional context alongside posts [29]. The adoption of such models is based on the idea of leveraging the crowd and harnessing its aggregated insights [54]. For the platforms, leveraging the crowd can be a beneficial approach even in noisy and inefficient user environments, as it relies on accumulated accuracy [55]. It is also considered scalable in addressing a large volume of information, with broader geographical coverage and faster detection [4, 29].

Community-based content moderation is often regarded as a less intrusive alternative to traditional, top-down fact-checking systems. It represents a bottom-up mechanism for monitoring information and shaping collective judgments about content credibility[4]. A growing body of literature that examines crowd moderation in comparison to expert fact-checking has recently emerged in relation to CNs. Within this area of research, there is no agreement on whether crowdsourcing can address some of the limitations inherent in expert fact-checking [7]. While some studies have shown that crowd moderation can effectively identify low-quality news sources and that community ratings often align closely with professional fact-checkers' assessments [31], others found differences in how the crowd and fact-checkers select and evaluate content [44].

Certain operational realities of crowdsourced approaches have also become evident. These systems are not necessarily working as intended, indicating that efforts to harness the 'wisdom of the crowd' may rest on overly optimistic assumptions about the integrity, diversity, and efficacy of user collaboration [4]. Applying the community-model to social media rests on the belief that the crowd has the ability to effectively moderate content; however, this perspective may overlook the difference of being able to moderate and choosing to do so [40]. Indeed, community-driven approaches arguably need to address the challenges of the willingness of its users to participate, as user contributions may be driven by political agendas or by purposive, alternative interpretations of facts [27, 30].

2.3 Effectiveness of Community Notes on X

There is a growing academic interest in evaluating the effectiveness of community-based moderation approaches, with particular attention to X's CNs, whose open-source data offer an opportunity for empirical analysis [50]. Key research questions concern the relationship between community-based moderation and professional fact-checking, the impact of CNs on the misleading posts they annotate, the capacity of the community to reach consensus as well as the system's overall capacity to mitigate politically motivated polarization.

With respect to the timeliness of CNs, previous studies have shown that CN moderation frequently lags during the initial and often most viral phase of content dissemination, implying that the system's current speed may not be sufficient to combat the circulation of misleading content [41]. This observation further highlights concerns about scalability of the system, suggesting that the effectiveness of moderation may not increase proportionally as the user base grows.

The effectiveness of the system also depends heavily on the quality and characteristics of sources, since CNs frequently draw upon fact-checked materials [7, 22, 40]. Solovev et al. have found that references to external sources significantly increase the odds of the note being perceived as helpful [46]. These findings are

consistent with prior research on note credibility, especially in the context of Covid-19 vaccines, which reported that notes addressing misleading posts were typically of good quality [2]. Other studies have found that the sources referenced in the notes are highly factual which in turn contributes to greater rater consensus [28].

Another research direction focuses on the impact of published notes on content perception and dissemination. Prior studies have found a remarkable reduction in the spread of misleading information as a result of the post being 'noted' [11]. Evidence suggests that when misinformation is annotated, the authors of such posts show concern for their reputation, frequently leading to broad retractions [25]. Moreover, posts with notes indicating misleading information have seen a decline in their visibility through a reduction in reposting compared to the posts determined to be not misleading [21]. Notes also affect user engagement, with the number of shares experiencing decrease by approximately 50%, while replies and views experience smaller, though still noticeable, declines [45].

Consensus formation within the CN system represents a key bottleneck, constraining the overall effectiveness of moderation. Empirical evidence underscores this challenge: only a small fraction of all created notes become visible, and a limited proportion of contributors produce notes that achieve consensus [9]. Contributors have been found to be more likely to write negative notes on posts by counter-partisans and to rate counter-partisans' notes as unhelpful [1]. Furthermore, Augenstein et al. highlight contributor behaviour and assessment dynamics as a barrier to the system's effectiveness, including its vulnerability to coordinated behaviour [3, 23]. Aligned with this observation, some researchers contend that the level of polarization on X is such that the bridging algorithm appears unable, perhaps even by design, to moderate the most highly polarized content [8, 53]. Given that the effectiveness of CNs depends on the ability to achieve consensus, failure to bridge the partisan divide can result in misleading content remaining unaddressed [19, 40].

In an effort to increase the effectiveness of the system, X continually updates and introduces new features. One such feature is the option for users to request notes on posts, where the goal is to increase the system's scale. Chuai et al. found that such requests prompt contributors to prioritise posts they characterise as greatly misleading, concurrently de-prioritising political content [12]. Another direction of research concerns the use of LLMs to generate "supernotes", thereby overcoming the prevalent dissensus phenomenon [17, 35].

As can be seen, prior research on CNs has largely focused on system-level outcomes, such as overall reductions in the spread of misleading content, without systematically investigating how the composition of the contributor crowd or the distribution of note production affects moderation dynamics. Additionally, the role of CNs as a platform for debate, rather than strictly as a moderation tool, has received limited attention. Our study addresses these gaps by combining analyses of contributor behaviour, note-level agreement, and temporal dynamics. This provides a more granular understanding of the mechanisms underlying CN effectiveness, offering novel insights into the social and operational factors that shape crowd-based moderation outcomes.

3 Methodology

3.1 Data

This study draws upon the publicly available daily updated dataset of CNs provided by X. The dataset comprises five principal categories, three of which are utilized in our analysis: (i) Community Notes data, containing the full text of each note, the associated post identifier, and the categorical labels assigned when flagging a post; (ii) Note rating data, which, when aggregated and evaluated in terms of the rater diversity, enables the determination of a note's perceived helpfulness; (iii) Note status history, documenting the sequence of status changes from initial creation to a permanently locked state after two weeks. For our analysis, we retrieved all available records on July 1, 2025, resulting in a corpus of approximately 1.8 million notes. For the research question addressing the time to note success, we defined the observation window with a start date of January 1, 2023, and an end date of May 31, 2025. The selection of this observation window was informed by X's announcement on January 23, 2023, instituting a policy that locks the status of Community Notes (CNs) two weeks following their creation[2]. For notes that had already been active for more than two weeks at the time of the announcement, X applied this locking rule retroactively. To ensure that this change does not affect the natural time to success, we therefore restricted the dataset to the notes created after this policy came into effect. Additionally, we limited the end date to May 31, 2025, to allow sufficient time for each note to reach its final status. This restriction ensures that all notes analyzed for this research question had attained their finite state.

3.2 How concentrated is the "crowd" in Community Notes?

To estimate the extent of concentration in note production among a small group of authors, we conducted an analysis drawing on methods from econometrics, information theory, and competition analysis to quantify 'dominance' or 'inequality' in contributions.

First, we calculate a metric of inequality estimation called the *Gini Coefficient* [20]. Despite being traditionally used in economic welfare domain for income inequality estimation, it can theoretically be applied to a variety of problems, from participation inequality estimation in health-oriented online communities [48] to citation distribution per author [39]. Since the cumulative number of CNs per user follows a discrete and continuous distribution, the Gini coefficient can be adopted to our task (Equation 1) in order to estimate inequality of contribution among unique authors using.

$$G = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \left| x_i - x_j \right|}{2n^2 \bar{x}} \tag{1}$$

where:

- n is the total number of individuals (or authors),
- x_i is the contribution (e.g., number of notes) of author i,
- $\bar{x}$ is the mean contribution across all authors.

Gini coefficient ranges from 0 to 1, representing perfect equality and perfect inequality respectively. It is directly derived from the Lorenz curve [26], a visual representation of the distribution of

[2] https://x.com/CommunityNotes/status/1616641919173685254

income or wealth in a country, illustrating the cumulative share of income held by the cumulative share of the population.

The second inequality metric we explore is Herfindahl-Hirschman Index (HHI) [42], which serves as a statistical measure of concentration. HHI is used to compute concentration in a variety of contexts, including the output of companies in banking or market sectors [43]. We adapt this metric to our task using Equation 2.

$$HHI = \sum_{i=1}^{n} s_i^2 \tag{2}$$

where

- $s_i = \frac{x_i}{\sum_{j=1}^{n} x_j}$ is the share of total activity (number of notes) contributed by author i,
- x_i is the contribution of author i,
- and n is the total number of authors.

To look into inequality of contribution from the entropy-based perspective, we additionally estimate the Theil Index (T) [13]. T a metric is derived from information theory to quantify the divergence of the observed distribution of contributions from a perfectly equal distribution. A Theil Index of 0 indicates complete equality (all authors contribute equally), while higher values reflect greater inequality, with a small subset of authors producing a disproportionate share of total notes.

$$T = \frac{1}{N} \sum_{i=1}^{N} \frac{x_i}{\bar{x}} \ln\left(\frac{x_i}{\bar{x}}\right) \tag{3}$$

where

- x_i is the contribution (e.g., number of notes written) by author i,
- $\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$ is the mean contribution across all N authors,
- and N is the total number of authors.

To measure the diversity in note authorship, we estimate Shannon Entropy (H) [10]. It quantifies the unpredictability of selecting a random contribution's author, with higher entropy indicating a more uniform distribution of contributions. Together with the Theil Index, entropy provides a complementary perspective on the concentration and equality of participation.

$$H = -\sum_{i=1}^{N} p_i \ln(p_i) \tag{4}$$

where

- $p_i = \frac{x_i}{\sum_{j=1}^{N} x_j}$ is the proportion of total contributions made by author i,
- x_i is the number of notes (or total activity) from author i,
- and N is the total number of authors.

Because entropy depends on the number of contributors N, it is often normalized to range between 0 and 1. The normalized form (H^*) rescales entropy by its theoretical maximum, allowing comparisons across datasets with different numbers of authors.

$$H^* = \frac{H}{\ln(N)} \tag{5}$$

Here, $H^* = 1$ indicates maximum diversity / perfect equality when all the authors contribute equally, while $H^* = 0$ indicates complete concentration (a single author produces all contributions).

Finally, we calculate the top-k share (S_k) to measure the proportion of total contributions made by the most active k authors. For instance, a Top-1% Share of 0.28 indicates that the top 1% of contributors produce 28% of all notes. This metric provides an intuitive view of concentration that complements formal inequality indices such as the Gini coefficient or Theil index. Higher values of S_k signify stronger dominance by a small elite of contributors. In our analysis, we focused on top-1% and top-10% measures.

For each of the metrics described above, we additionally explore how concentration (inequality) changes over time. This provides further insight into whether the note-writing activity is becoming more dominated by a few people during certain periods of time. To perform such an analysis, we used the timestamp of note creation to estimate the monthly contributions per note author. We then computed the inequality metrics over the monthly period of the CN dataset.

3.3 Fostering consensus or dissensus?

To further analyse the types of engagement produced by the crowd, we examine the patterns of consensus building. The extent of the consensus building through CNs can be operationalised through two metrics. First, for the posts that received more than one CN, the notes can agree or disagree on whether a particular post is *misinformation* or *not misinformation*. We refer to this dimension as "classification consensus". The second method of defining consensus is based on the agreement of the crowd regarding the helpfulness of each individual note. Under this definition, consensus may be reached for both published and unpublished notes. We refer to it as "rating consensus".

There are three potential outcomes for the *classification consensus*, and each concerns posts receiving more than one note.

- *Dissensus*: There is no agreement among the note posters on whether a post on X is misinformation or not misinformation. By the lack of agreement, we mean that there is a mixture of notes marking the post as misinformation or not.
- *Consensus — not misleading*: All the notes that are created for a particular post mark it as not misinformation.
- *Consensus — misleading*: All the CNs that are flagging the post marked it as misinformation.

To estimate the extent of classification consensus, we analysed the temporal distribution of consensus and dissensus by grouping the notes related to the same unique post and the same "classification" field.

For the rating consensus, we define consensus as the presence of agreement on whether the note should be published. To estimate the extent of the rating consensus, we selected the notes within the "Note Status History" dataset that have the status of "needs more ratings" (NMR). As per X's definition, this status serves as an indicator that the note has not yet met the algorithmic consensus criterion of a "diversity of perspectives" and therefore has not been agreed upon as either helpful or unhelpful.

While the CN system is designed to render a judgement, dissensus may occur potentially owing to user misuse, trolling behaviour, or strategic non-cooperation. Thus, dissensus is not necessarily organic. In certain cases, it may be deliberately evoked, promoted, or sustained as part of ongoing contestation over a note's validity.

One of such misuse cases is the NNN phenomenon described in Section 2.1. Rather than flagging potentially misleading content, participants use the platform to engage in arguments by providing counterpoints and meta-commentary.

To examine the role of Note Not Needed (NNN) in shaping patterns of dissensus, we conducted a complementary analysis by identifying all posts that either contain the term "NNN" or "Note Not Needed" in their note text, or have at least one rating marking at least one of its notes as not helpful for the reason "Note Not Needed". We then examine the temporal distribution of this engagement, focusing on posts associated with notes flagged as NNN to assess how the crowd misuses the system over time.

We also performed a statistical analysis using Odds Ratios (OR) to assess whether the *Note Not Needed* (NNN) phenomenon is associated with the level of helpfulness consensus among notes attached to a given post [33]. Our assumption is that if at least one person indicated that the post needs no notes, we can assume that the post represents an opinion rather than a false claim. One would therefore expect the successful crowdsourcing community to agree on non-helpfulness of any notes under such posts. Under our analysis, we define *NNN posts* as posts that have at least one note flagged as NNN, either through the note taking or note rating crowd engagement. Similarly, a *non-NNN posts* are the posts that had no notes flagged as NNN. This leads to the following 2×2 contingency table:

	Published (True)	Not Published (False)
NNN-flagged (True)	a	b
Not NNN (False)	c	d

where:

- a = number of NNN posts that contain at least one published note,
- b = number of NNN posts that have no any published notes,
- c = number of non-NNN posts that have at least one published note,
- d = number of non-NNN posts that have no NNN-flagged notes and that have no notes published.

We then test whether the NNN posts are less likely to have any notes published compared to non-NNN posts under the following hypotheses:

$$H_0 : P(\text{Published} \mid \text{NNN}) = P(\text{Published} \mid \text{Not NNN})$$
$$H_1 : P(\text{Published} \mid \text{NNN}) < P(\text{Published} \mid \text{Not NNN})$$

Based on these hypotheses, we estimate OR as a quantifier of the association between NNN flagging and people refusing to accept the notes under such posts:

$$\text{OR} = \frac{a/b}{c/d} = \frac{ad}{bc} \tag{6}$$

According to the OR interpretation, OR < 1 indicates that NNN-flagged notes are less likely to be published compared to other notes.

3.4 Is consensus more likely the outcome of the timeliness of notes?

As discussed in Section 2.3, prior research has emphasized that the timeliness of notes plays a critical role: the inherently slower pace of crowdsourced moderation can limit its effectiveness in preventing the spread of misinformation before it gains traction. Because the CN bridging algorithm requires a sufficient diversity of perspectives for a note to become visible, the time needed to meet this threshold may extend beyond the period in which most misinformation is spread. We therefore aim to examine more closely whether the timeliness of a note itself can enhance the effectiveness of consensus building.

To answer this question, we first retrieve the exact time of each post associated with a community note data using X's Snowflake library for Python[3]. We then calculate the time lag between the post on X and note publication time, subtracting the timestamp of the post from the earliest note publication timestamp. To calculate the lag, we utilize the Note Status History dataset, which records several relevant timestamps:

- The first instance when the note's status transitioned from NMR to either "Helpful" or "Not Helpful".
- The timestamp corresponding to the change to the current status (either "Helpful" or "Not Helpful").
- The time of the most recent non-NMR status update.
- The time when the note was locked, i.e., when its status became permanent as either "Helpful", NMR or "Not Helpful".

To identify this earliest time, we utilize a two-step filtering procedure:

- Eligibility check: Filter in only the notes with a locked status of Currently Rated Helpful or Currently Rated Not Helpful.
- Earliest status timestamp: For each eligible note, search all status transitions to identify the first moment the note entered its final state.

We then group the calculated time lags into 3-hour intervals and stratify them by note status to estimate the temporal window during which note creation is most likely to maximize publication likelihood.

4 Results

4.1 Crowd composition analysis

We examined whether note creation is broadly distributed across contributors or is concentrated among a small subset of highly active authors. To quantify this, we calculated multiple inequality measures — the Gini coefficient, Herfindahl–Hirschman Index (HHI), Theil index, and normalized Shannon entropy — alongside the cumulative contribution shares of the top 1%, 10% of authors. The results of the cumulative analysis are provided in Table 1.

[3] https://github.com/twitter-archive/snowflake

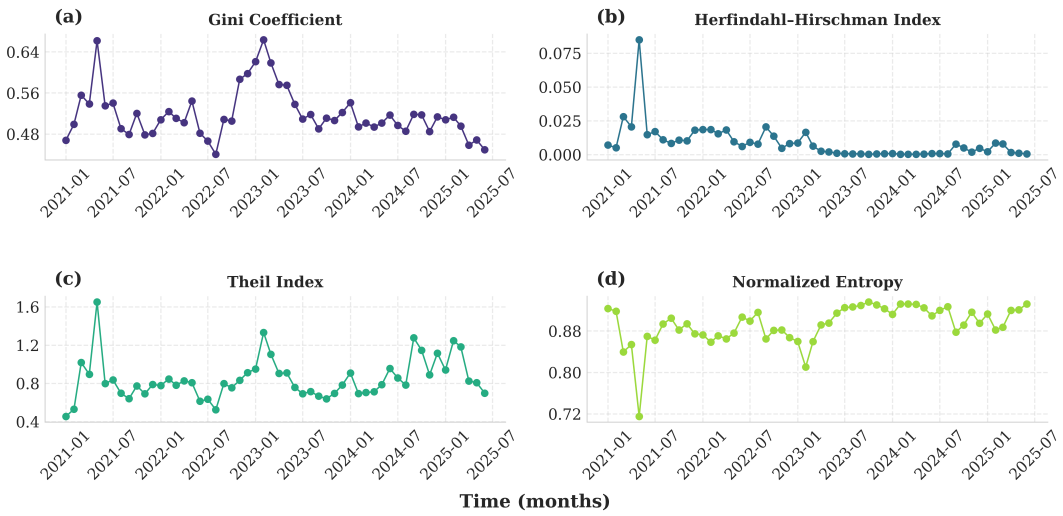

Figure 1: Inequality of Note Contributions Over Time.

Table 1: Summary of inequality metrics for note authorship distribution.

Metric	Value
Gini coefficient	0.679
Herfindahl–Hirschman Index (HHI)	0.0009
Theil index	1.442
Normalized entropy	0.884
Top 1% share	0.282
Top 10% share	0.584

Overall, our analysis reveals that note authorship is strongly skewed, with a Gini coefficient ≈ 0.68, indicating that contributions are far from evenly distributed. In many social systems or collaborative content platforms, Gini values in the 0.5–0.8 range are consistent with "heavy-tailed" contribution patterns (i.e., a few contributors are responsible for a large share of activity). For instance, in studies of scientific publishing, Gini values of 0.73 for citation counts reflect extreme skew in impact among authors [39]. The observed value of ≈ 0.68 for CN data therefore indicates a high degree of inequality consistent with a platform where a more committed core of users dominates contribution.

Although the HHI metric is numerically small (≈ 0.000878), in a system with many authors, even small deviations from uniformity imply meaningful concentration. Such levels of inequality are comparable to skew observed in other high-activity platforms [48].

While normalized entropy (≈ 0.88) suggests that the system maintains a broad base of participants, the top 1% of authors account for ≈28% of all notes, and the top 10% comprise ≈58%, underscoring the dominance of a relatively small elite. In cumulative participation literature, such "skewed top-share" patterns are well-known: for instance, in health-oriented platforms or scientific citations, a small fraction of participants or scholars account for a disproportionately large fractions of impact [39, 48].

The Theil index (≈ 1.44) further confirms substantial divergence from an ideal uniform distribution. This pattern suggests that the system's productivity is centralized: while many users contribute, a few exert disproportionate influence.

The temporal analysis of the concentration metrics is shown in Figure 1. It reveals distinct evolutionary phases. During 2021–2022, participation was highly concentrated, with Gini values exceeding 0.6 and Theil indices above 1.5, reflecting dominance by a small cohort of early contributors. As the platform expanded globally through 2023, inequality initially intensified — consistent with cumulative advantage dynamics observed in open collaboration systems [5, 16]. However, from late 2023 onward, the Gini coefficient declined to ≈ 0.5 and normalized entropy rose above 0.93, indicating increasing diversification of contributors. This suggests that the system has matured toward a more balanced state. Such evolution aligns with participation cycle theories [38], where initial concentration transitions to broader engagement as communities scale. Overall, the results depict a platform moving from elite-dominated origins toward a more inclusive participation regime, while retaining residual concentration characteristic of online moderation ecosystems.

4.2 Do CNs tend to build consensus?

Our analysis suggests that CNs are ineffective in building consensus. In particular, we found that the prevalent proportion of notes ($\approx 69\%$) are in a dissensus about the post being misleading or not, with only $\approx 29\%$ of notes having a consensus regarding a potentially misleading post and a small proportion of notes ($\approx 2\%$) in a consensus regarding a post being not misinformation. The temporal analysis (Figure 2) further reveals a strong trend towards the prevalence of dissensus over time as more people author the notes. While the early system, with its few note authors, produced a consensus about a substantial majority of posts being misleading, with the global enrolment of CNs in early 2023, the proportion of cases of disagreement increased to 69%. Given the change in

crowd diversification discussed in Section 4.1, this finding can be attributed to a variety of opinions and a higher percentage of active contributors.

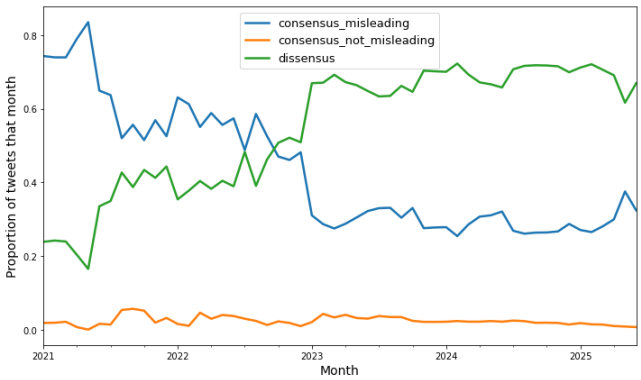

Figure 2: Note classification agreement per month.

With regard to the rating consensus, we found that raters reached consensus on 11.5% of all notes, with the remaining 88.5% of the notes remaining in the NMR status. This could be attributed to two factors. First, some of the notes never receive 5 ratings to move out of the initial NMR state. Second, the helpfulness estimation model may never reach the required diversity of perspective in the ratings, thereby not satisfying the algorithmic criteria.

The "NNN" phenomenon. The distribution of the post proportion that exhibits the NNN phenomenon is illustrated in Figure 3. As can be seen, the prevalence in dissensus notes reported in Figure 2 is aligned with a spike in the NNN-noted posts toward the end of 2022–beginning of 2023, suggesting that this phenomenon can be one driver of dissensus. After this period, the proportion of NNN-noted posts stabilises at a high level of $\approx$ 68% of all posts. This observation highlights the persistence of what we term"noting as debating", a dynamic that can seen as a form of system repurposing, where participants use a tool intended for content evaluation as a forum for interaction and contestation. While this constitutes a "misuse" of the system from the perspective of platform design and policy, it may also indicate an unmet user demand for structured discussion spaces within the platform ecosystem.

Table 2: Contingency table showing the relationship between NNN presence and note publication status.

NNN Presence	No Published Note	Has Published Note
Absent	456,727	23,599
Present	684,282	110,461

Contrary to our expectation, we found no significant association between the NNN posts and decreased likelihood of publication of any notes under that post, with an Odds Ratio= 3.124 and $p < 10^{-300}$. Table 2 represents the contingency table for Fisher's exact test. This finding indicates that the use of Community Notes for debate or opinion-driven exchanges is not merely incidental but appears to be reinforced by community feedback, as such notes are disproportionately marked as helpful.

4.3 The timeliness of CN posting

Figure 4 illustrates the success rate of Community Notes (CRH) in the context of the timeliness of writing a CN after the publication of the post on X. CRH exhibits dependency on the delay in note creation relative to the time of the post, which is associated with both longer times and lower percentages of reaching to CRH status. We calculated Spearman's rank correlation to assess the monotonic relationships between the creation delay and two outcome variables: mean time to CRH and percentage of CRH. The analysis reveals a statistically significant positive monotonic relationship between the creation delay and the mean time to CRH (rs=0.8633, p<0.001), indicating that longer delays are associated with increased time to become helpful. In contrast, we observed a significant negative monotonic relationship between the creation delay and the percentage of helpful notes (rs=-0.5118, p<0.001), showing that as the delays increase, the proportion of helpful notes decreases. These findings suggest that longer creation delays negatively impact the timeliness and helpfulness of notes.

On average, notes reach a *Currently Rated Helpful (CRH)* status 21.7 hours after creation, with a median time of 6.2 hours. In comparison, notes classified as *Currently Rated Not Helpful (CRNH)* take longer to stabilize, averaging 43.7 hours with a median of 5.0 hours. When considering the full timeline from post publication to note evaluation, the temporal gap becomes more pronounced. Posts receive their first helpful note after an average of 65.7 hours ($\approx$ 2.7 days) and a median of 15.3 hours, while notes ultimately deemed not helpful appeared later—after an average of 94.3 hours ($\approx$ 3.9 days) and a median of 18.6 hours. These findings suggest that helpful notes tend to emerge more quickly, whereas unhelpful notes are both slower to appear and slower to converge on their final status.

5 Discussion

Our findings highlight a fundamental tension between the design aspirations of CNs and their empirical outcomes. The platform aims to crowdsource content moderation by harnessing diversity to reach consensus; however, our analyses show that this diversity often amplifies disagreement instead. In particular, we observed a correlation between the time when the concentration of the crowd became less skewed and the time when disagreement among the note writers became prevalent, with nearly 69% of notes being in disagreement with each other regarding the presence of misleading content. Our analysis of the note moderation consensus also revealed a high state of disagreement, with 90% of notes remaining unresolved. This persistent dissensus may stem from either an insufficient number of contributor ratings or a lack of the "diverse perspectives" required by the platform's bridging algorithm.

A key factor underlying this dissensus is the widespread occurrence of the notes under the posts that "need no notes" (NNN posts). We observed not only the persistent prevalence of notes attached to opinion-based posts but also that such notes frequently achieve successful publication. Several factors may explain this pattern. First, such posts may potentially combine factual claims

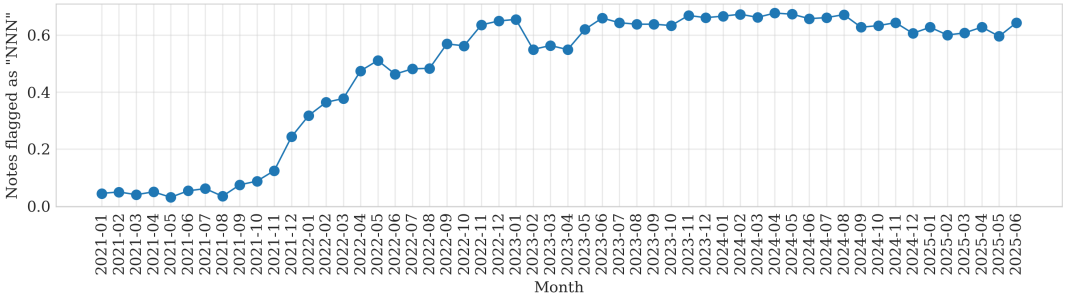

Figure 3: Monthly proportion of posts flagged as NNN.

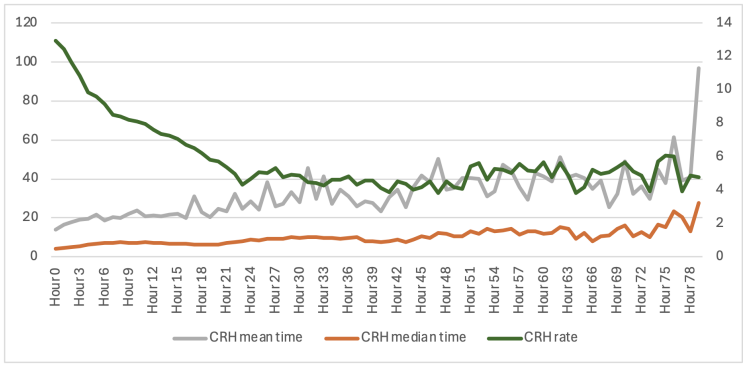

Figure 4: Effect of Creation Delay on Ratio and Time to Reach CRH in the First 80 Hours

with subjective commentary, leading contributors to disagree on whether they warrant annotation. In such cases, some notes may legitimately address misleading elements, while others invoke the NNN label to signal that opinionated content requires no correction. Second, some notes may have been created and published before subsequent contributors marked the post as NNN. Finally, highly controversial or widely circulated opinionated posts attract more overall engagement, increasing the likelihood that at least one note will meet the publication criteria despite the underlying disagreement.

Timeliness further constrains CN effectiveness. On average, notes that are timely in terms of combating potentially misleading content are significantly more likely to be published. However, we still observed an over 65 hour delay in publishing the note after the emergence of a misleading post. Therefore, contrary to the belief that community-based moderation offers a more timely reaction to misinformation, this delay is comparable to that of professional fact-checkers that is reported to take up to four days on average [34, 49]. As a result, CNs primarily function as post-hoc correctives rather than preventative interventions.

Finally, participation inequality remains a defining characteristic of CNs. We found that a small minority (approximately 10%) of contributors account for 58% of all notes, while the majority remain largely inactive. Such concentration mirrors the epistemic asymmetries seen in traditional fact-checking, centralizing influence in a narrow elite. If unchecked, this concentration may limit diversity of perspectives, increase fragility to contributor churn, and entrench power among the most active authors. Our temporal analysis suggests that the platform seems to be slowly moving toward the more diverse and truly crowdsource-based state.

Together, these results challenge the foundational assumption of CNs as a form of effective collective moderation. Instead of aggregating distributed expertise, CNs operate as a stratified and temporally constrained system shaped by a small subset of users—an algorithmic crowd whose diversity is more formal than substantive.

6 Conclusions and Future Work

This study demonstrates that Community Notes, while designed to democratize moderation, often reproduce the very asymmetries they were meant to overcome. The system's reliance on algorithmic diversity produces persistent dissensus, its timeliness fails to match the velocity of misinformation, and its participation dynamics concentrate epistemic authority within a small core of contributors.

To realise the democratic and epistemic potential of crowdsourced moderation, future designs must integrate deliberative affordances— structured opportunities for transparent dialogue—alongside algorithmic scaling. Introducing features that separate fact annotation from meta-discussion could preserve CNs' corrective value while accommodating legitimate debate. Additionally, quality-weighted diversity metrics and balanced participation incentives could mitigate contributor inequality. Going forward, decomposing inequality across content categories and tracking its evolution longitudinally will be critical to assessing whether the

platform is becoming increasingly concentrated and to inform interventions aimed at broadening participation.

Future research should examine the longitudinal evolution of Community Notes participation, particularly during major political events, to better understand how patterns of dissensus and contribution concentration shift over time. Moreover, persistent asymmetries in note production raise critical questions about the potential for coordinated or strategic behaviours, including brigade-style interventions in both note writing and rating [12, 40]. Investigating these phenomena through network or temporal anomaly detection methods could help distinguish organic crowd participation from orchestrated collective action. Such analyses would advance our understanding of the social and algorithmic mechanisms underlying crowdsourced moderation, offering a foundation for more resilient, transparent, and democratically accountable systems.

Acknowledgments

This work has been co-funded by the UK's innovation agency (Innovate UK) grant 10039055 (approved under the Horizon Europe Programme as vera.ai, EU grant agreement 101070093) under action number 2020-EU-IA-0282. The contribution by Richard Rogers is funded by the Horizon Europe Programme project, SoMe4Dem, grant nr. 101094752. We would like to thank Bumju Jung, Maricarmen Rodríguez Guillen, and Wilma Ewerhart for their contributions to this research during 2025 Digital Methods Summer School, Media Studies, University of Amsterdam. The contribution by Adel Tayebi is financed by CY Cergy Paris University as part of the EUTOPIA co-tutelle scheme.

References

[1] Jennifer Allen, Cameron Martel, and David G Rand. 2022. Birds of a feather don't fact-check each other: Partisanship and the evaluation of news in Twitter's Birdwatch crowdsourced fact-checking program. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–19. doi:10.1145/3491102.3502040

[2] Matthew R. Allen, Nimit Desai, Aiden Namazi, Eric Leas, Mark Dredze, Davey M. Smith, and John W. Ayers. 2024. Characteristics of X (Formerly Twitter) Community Notes Addressing COVID-19 Vaccine Misinformation. *JAMA* 331, 19 (May 2024), 1670. doi:10.1001/jama.2024.4800

[3] Isabelle Augenstein, Michiel Bakker, Tanmoy Chakraborty, David Corney, Emilio Ferrara, Iryna Gurevych, Scott Hale, Eduard Hovy, Heng Ji, Irene Larraz, et al. 2025. Community Moderation and the New Epistemology of Fact Checking on Social Media. *arXiv preprint arXiv:2505.20067* (2025).

[4] Isabelle Augenstein, Michiel Bakker, Tanmoy Chakraborty, David Corney, Emilio Ferrara, Iryna Gurevych, Scott Hale, Eduard Hovy, Heng Ji, Irene Larraz, Filippo Menczer, Preslav Nakov, Paolo Papotti, Dhruv Sahnan, Greta Warren, and Giovanni Zagni. 2025. Community Moderation and the New Epistemology of Fact Checking on Social Media. doi:10.48550/ARXIV.2505.20067 Version Number: 1.

[5] Albert-László Barabási and Réka Albert. 1999. Emergence of Scaling in Random Networks. *Science* 286, 5439 (1999), 509–512. arXiv:https://www.science.org/doi/pdf/10.1126/science.286.5439.509 doi:10.1126/science.286.5439.509

[6] Nadav Borenstein, Greta Warren, Desmond Elliott, and Isabelle Augenstein. 2025. Can Community Notes Replace Professional Fact-Checkers?. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar (Eds.). Association for Computational Linguistics, Vienna, Austria, 535–552. doi:10.18653/v1/2025.acl-short.42

[7] Nadav Borenstein, Greta Warren, Desmond Elliott, and Isabelle Augenstein. 2025. Can Community Notes Replace Professional Fact-Checkers?. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Wanxiang Che, Joyce Nabende, Ekaterina Shutova, and Mohammad Taher Pilehvar (Eds.). Association for Computational Linguistics, Vienna, Austria, 535–552. doi:10.18653/v1/2025.acl-short.42

[8] Paul Bouchaud and Pedro Ramaciotti. 2025. Algorithmic resolution of crowdsourced moderation on X in polarized settings across countries. doi:10.48550/arXiv.2506.15168 arXiv:2506.15168 [cs].

[9] Roberta Braga, Cristina Tardáguila, , and Marcelo Soares. 2025. A Deep Dive into X's Community Notes: An Analysis of English and Spanish Contributions Between 2021 and 2025. Digital Democracy Institute of the Americas, Ljubljana Slovenia. https://ddia.org/en/a-deep-dive-into-xs-community-notes-report

[10] PA Bromiley, NA Thacker, and E Bouhova-Thacker. 2004. Shannon entropy, Renyi entropy, and information. *Statistics and Inf. Series (2004-004)* 9, 2004 (2004), 2–8.

[11] Yuwei Chuai, Moritz Pilarski, Thomas Renault, David Restrepo-Amariles, Aurore Troussel-Clément, Gabriele Lenzini, and Nicolas Pröllochs. 2024. Community-based fact-checking reduces the spread of misleading posts on social media. doi:10.48550/arXiv.2409.08781 arXiv:2409.08781 [cs].

[12] Yuwei Chuai, Anastasia Sergeeva, Gabriele Lenzini, and Nicolas Pröllochs. 2025. Community Fact-Checks Trigger Moral Outrage in Replies to Misleading Posts on Social Media. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–23. doi:10.1145/3706598.3713909 arXiv:2409.08829 [cs].

[13] Pedro Conceição and Pedro Ferreira. 2000. The young person's guide to the Theil index: Suggesting intuitive interpretations and exploring analytical applications. (2000).

[14] Giulio Corsi, Elizabeth Seger, and Sean Ó hÉigeartaigh. 2024. Crowdsourcing the Mitigation of disinformation and misinformation: The case of spontaneous community-based moderation on Reddit. *Online Social Networks and Media* 43-44 (Nov. 2024), 100291. doi:10.1016/j.osnem.2024.100291

[15] Anamaria Crisan and Andrew M McNutt. 2025. Linting is People! Exploring the Potential of Human Computation as a Sociotechnical Linter of Data Visualizations. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, Article 634, 10 pages. doi:10.1145/3706599.3716230

[16] Kevin Crowston, Kangning Wei, James Howison, and Andrea Wiggins. 2012. Free/Libre Open Source Software Development: What We Know and What We Do Not Know. *Journal of Information Technology & Software Engineering* 2, 7 (2012), 1–21. https://jisajournal.springeropen.com/articles/10.1186/s13174-017-0061-4

[17] Soham De, Michiel A. Bakker, Jay Baxter, and Martin Saveski. 2024. Supernotes: Driving Consensus in Crowd-Sourced Fact-Checking. doi:10.48550/arXiv.2411.06116 arXiv:2411.06116 [cs].

[18] Emillie de Keulenaar. 2025. From Twitter to X: demotion, community notes and the apparent shift from adjudication to consensus-building. *Available at SSRN 5165083* (2025).

[19] Emillie de Keulenaar and Ivan Kisjes. 2025. Content moderation from Twitter to X: policies, enforcement and community notes. doi:10.17632/D9GGKZP7BC.2

[20] Robert Dorfman. 1979. A formula for the Gini coefficient. *The review of economics and statistics* (1979), 146–149.

[21] Chiara Drolsbach and Nicolas Pröllochs. 2023. Diffusion of Community Fact-Checked Misinformation on Twitter. doi:10.48550/arXiv.2205.13673 arXiv:2205.13673 [cs].

[22] Chiara Patricia Drolsbach, Kirill Solovev, and Nicolas Pröllochs. 2024. Community notes increase trust in fact-checking on social media. *PNAS Nexus* 3, 7 (June 2024), pgae217. doi:10.1093/pnasnexus/pgae217

[23] Vittoria Elliott. [n. d.]. Elon Musk's Main Tool for Fighting Disinformation on X Is Making the Problem Worse, Insiders Claim. *Wired* ([n. d.]). https://www.wired.com/story/x-community-notes-disinformation/ Section: tags.

[24] Daniela Flamini. 2019. *Fact-Checking in Buenos Aires & the Modern Journalistic Struggle Over Knowledge*. Ph. D. Dissertation. Trinity College.

[25] Yang Gao, Maggie Zhang, and Huaxia Rui. 2024. Can Crowdchecking Curb Misinformation? Evidence from Community Notes. doi:10.2139/ssrn.4992470

[26] Joseph L Gastwirth. 1971. A general definition of the Lorenz curve. *Econometrica: Journal of the Econometric Society* (1971), 1037–1039.

[27] Dan M. Kahan. 2017. Misconceptions, Misinformation, and the Logic of Identity-Protective Cognition. *SSRN Electronic Journal* (2017). doi:10.2139/ssrn.2973067

[28] Uku Kangur, Roshni Chakraborty, and Rajesh Sharma. 2024. Who Checks the Checkers? Exploring Source Credibility in Twitter's Community Notes. doi:10.48550/arXiv.2406.12444 arXiv:2406.12444 [cs].

[29] Travis Lloyd, Tung Nguyen, Karen Levy, and Mor Naaman. 2025. Beyond Community Notes: A Framework for Understanding and Building Crowdsourced Context Systems. doi:10.48550/arXiv.2509.15434 arXiv:2509.15434 [cs].

[30] Jacqueline M. Otala, Gillian Kurtic, Isabella Grasso, Yu Liu, Jeanna Matthews, and Golshan Madraki. 2021. Political Polarization and Platform Migration: A Study of Parler and Twitter Usage by United States of America Congress Members. In *Companion Proceedings of the Web Conference 2021*. ACM, Ljubljana Slovenia, 224–231. doi:10.1145/3442442.3452305

[31] Cameron Martel, Jennifer Allen, Gordon Pennycook, and David G Rand. 2024. Crowds can effectively identify misinformation at scale. *Perspectives on Psychological Science* 19, 2 (2024), 477–488.

[32] Ariadna Matamoros-Fernández and Nadia Jude. 2025. The importance of centering harm in data infrastructures for 'soft moderation': X's Community Notes as a case study. *New Media & Society* 27, 4 (April 2025), 1986–2011. doi:10.1177/14614448251314399 Publisher: SAGE Publications.

[33] Mary L McHugh. 2009. The odds ratio: calculation, usage, and interpretation. *Biochemia medica* 19, 2 (2009), 120–126.

[34] Nicholas Micallef, Vivienne Armacost, Nasir Memon, and Sameer Patil. 2022. True or False: Studying the Work Practices of Professional Fact-Checkers. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–44. doi:10.1145/3512974 Interviews with 21 professional fact-checkers across 19 countries about their practices and constraints..

[35] Saeedeh Mohammadi and Taha Yasseri. 2025. AI Feedback Enhances Community-Based Content Moderation through Engagement with Counterarguments. doi:10.48550/arXiv.2507.08110 arXiv:2507.08110 [cs].

[36] Mohsen Mosleh, Qi Yang, Tauhid Zaman, Gordon Pennycook, and David G Rand. 2024. Differences in misinformation sharing can lead to politically asymmetric sanctions. *Nature* 634, 8034 (2024), 609–616.

[37] Aviv Ovadya and Luke Thorburn. 2023. Bridging Systems: Open Problems for Countering Destructive Divisiveness across Ranking, Recommenders, and Governance. doi:10.48550/arXiv.2301.09976 arXiv:2301.09976 [cs].

[38] Katherine Panciera, Aaron Halfaker, and Loren Terveen. 2009. Wikipedians Are Born, Not Made: A Study of Power Editors and Their Motivational Trajectories. In *Proceedings of the ACM 2009 International Conference on Supporting Group Work (GROUP '09)*. ACM, 51–60. doi:10.1145/1531674.1531682

[39] Alexander M Petersen and Orion Penner. 2014. Inequality and cumulative advantage in science careers: a case study of high-impact journals. *EPJ Data Science* 3, 1 (Oct. 2014). doi:10.1140/epjds/s13688-014-0024-y

[40] Nicolas Pröllochs. 2022. Community-Based Fact-Checking on Twitter's Birdwatch Platform. *Proceedings of the International AAAI Conference on Web and Social Media* 16 (May 2022), 794–805. doi:10.1609/icwsm.v16i1.19335

[41] Thomas Renault, David Restrepo Amariles, and Aurore Troussel. 2024. Collaboratively adding context to social media posts reduces the sharing of false news. doi:10.48550/arXiv.2404.02803 arXiv:2404.02803 [econ].

[42] Stephen A Rhoades. 1993. The herfindahl-hirschman index. *Fed. Res. Bull.* 79 (1993), 188.

[43] Stephen A. Rhoades. 1993. The Herfindahl-Hirschman index. *Federal Reserve Bulletin* (1993), 188–189. https://api.semanticscholar.org/CorpusID:153018440

[44] Mohammed Saeed, Nicolas Traub, Maelle Nicolas, Gianluca Demartini, and Paolo Papotti. 2022. Crowdsourced Fact-Checking at Twitter: How Does the Crowd Compare With Experts?. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*. 1736–1746. doi:10.1145/3511808.3557279 arXiv:2208.09214 [cs].

[45] Isaac Slaughter, Axel Peytavin, Johan Ugander, and Martin Saveski. 2025. Community Notes Moderate Engagement With and Diffusion of False Information Online. doi:10.48550/arXiv.2502.13322 arXiv:2502.13322 [cs].

[46] Kirill Solovev and Nicolas Pröllochs. 2025. References to unbiased sources increase the helpfulness of community fact-checks. doi:10.48550/arXiv.2503.10560 arXiv:2503.10560 [cs].

[47] Mark Stencel, Erica Ryan, and Joel Luther. 2024. With half the planet going to the polls in 2024, fact-checking sputters. *Duke Reporters' Lab* (2024).

[48] Trevor van Mierlo, Douglas Hyatt, and Andrew T Ching. 2016. Employing the Gini coefficient to measure participation inequality in treatment-focused Digital Health Social Networks. *Network Modeling Analysis in Health Informatics and Bioinformatics* 5, 1 (2016), 32.

[49] Morgan Wack, Kayla Duskin, and Damian Hodel. 2024. Political Fact-Checking Efforts are Constrained by Deficiencies in Coverage, Speed, and Reach. *arXiv preprint arXiv:2412.13280* (2024). A full assessment during the 2022 U.S. midterm elections finds the median fact-check release time is 4 days after the narrative's first appearance..

[50] Chenlong Wang and Pablo Lucas. 2024. Efficiency of Community-Based Content Moderation Mechanisms: A Discussion Focused on Birdwatch. *Group Decision and Negotiation* 33, 3 (June 2024), 673–709. doi:10.1007/s10726-024-09881-1

[51] Valerie Wirtschafter and Sharanya Majumder. 2023. Future challenges for online, crowdsourced content moderation: evidence from Twitter's community notes. *Journal of Online Trust and Safety* 2, 1 (2023).

[52] Valerie Wirtschafter and Sharanya Majumder. 2023. Future Challenges for Online, Crowdsourced Content Moderation: Evidence from Twitter's Community Notes. *Journal of Online Trust and Safety* 2, 1 (Sep. 2023). doi:10.54501/jots.v2i1.139

[53] Valerie Wirtschafter and Sharanya Majumder. 2023. Future Challenges for Online, Crowdsourced Content Moderation: Evidence from Twitter's Community Notes. *Journal of Online Trust and Safety* 2, 1 (Sept. 2023). doi:10.54501/jots.v2i1.139

[54] Stefan Wojcik, Sophie Hilgard, Nick Judd, Delia Mocanu, Stephen Ragain, M. B. Fallin Hunzaker, Keith Coleman, and Jay Baxter. 2022. Birdwatch: Crowd Wisdom and Bridging Algorithms can Inform Understanding and Reduce the Spread of Misinformation. doi:10.48550/arXiv.2210.15723 arXiv:2210.15723 [cs].

[55] Anita Williams Woolley, Christopher F. Chabris, Alex Pentland, Nada Hashmi, and Thomas W. Malone. 2010. Evidence for a Collective Intelligence Factor in the Performance of Human Groups. *Science* 330, 6004 (Oct. 2010), 686–688. doi:10.1126/science.1193147

Received 7 October 2025