JISTaP https://www.jistap.org
Journal of Information Science Theory and Practice

RESEARCH PAPER
J Inf Sci Theory Pract 13(1): 1-14, 2025

eISSN : 2287-4577 pISSN : 2287-9099
https://doi.org/10.1633/JISTaP.2025.13.1.1

Do Retweets Undermine Twitter Mentions of Journal Articles as Indicators of Research Significance?

Tint Hla Hla Htoo*
Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore
E-mail: tint0010@e.ntu.edu.sg

Na Jin-Cheon
Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore
E-mail: TJCNa@ntu.edu.sg

Michael Thewall
Information School, University of Sheffield, Sheffield, United Kingdom
E-mail: m.a.thelwall@sheffield.ac.uk

ABSTRACT

This study aims to investigate whether the prevalence of retweets challenges the validity of Twitter mentions (tweet counts and tweeter counts) as indicators of research significance. The study first examines whether the tweeting of research papers signifies either the wisdom of crowd effect or herd-like behavior for a dataset of COVID-19 papers from *The Lancet*. The study then uses the Modality, Agency, Interactivity, and Navigability (MAIN) model to examine the nature of the influence involved in retweeting. The Mann-Whitney U test and multiple linear regression were used. Findings show that there was extensive evidence of herd-like behaviour, rather than the wisdom of crowd effect, in tweeting research papers, challenging the validity of Twitter mentions as indicators of research significance. Credibility heuristic cues of the original tweeters were associated with their retweet rates, suggesting that retweets are more likely influenced by perceptions of the original tweeter's credibility rather than the quality of research papers.

Keywords: altmetrics, Twitter mentions, MAIN model, heuristic cues

Received: January 14, 2024 **Revised:** August 15, 2024
Accepted: August 16, 2024 **Published:** March 30, 2025

***Corresponding Author:** Tint Hla Hla Htoo
https://orcid.org/0000-0001-9331-5260
E-mail: tint0010@e.ntu.edu.sg

1. INTRODUCTION

As part of the "Impact Agenda," government-led research policies in many countries aim to increase the impact of research on society (Gunn & Mintrom, 2016). In various nations, policy interventions promoting the Impact Agenda have fostered a culture in which academics are expected to consider societal benefits. For example, in the UK, university departments are required to describe the non-academic impacts of their research under the UK's Research Excellence Framework (Thelwall, 2021). Such efforts have contributed to the development of altmetrics, which report how often a research paper has been tweeted, shared, or saved on social media, reference management tools, and other online platforms, because traditional citation-based indicators can only reflect impacts within the academic publishing system.

Altmetrics were originally proposed as a form of crowd-sourced peer review to help researchers quickly identify significant scholarly works (Priem et al., 2011). In this sense, altmetric indicators, such as Twitter mentions, are potentially crowd-sourced indicators of research significance. According to Lamont (2009), the concept of significance in research evaluation can be defined in terms of three dimensions: scholarly, political, and social. Therefore, in this study, research significance is understood as the extent to which a study has relevance and impact in one or more of these domains. According to the "wisdom of crowds" concept, in some contexts a diverse group of individuals can make better decisions and predictions than a few individuals or even an expert (Galton, 1907; Hertwig, 2012; Lorenz et al., 2011; Surowiecki, 2005). But not all crowds are "wise." One of the criteria required to form a wise crowd is independence (Surowiecki, 2005). Without independence, much of or at least part of the information selected can occur by crowd following rather than through separate judgments, and in this case the crowd selection process may not lead to better outcomes. Banerjee (1992) referred to sequential decision making behaviour involving crowd following as "herd behaviour," which could lead to a cascading failure, called the "stupidity of herds" by Yu et al. (2018).

Herd behaviour is common in many social and economic situations (Banerjee, 1992). In the two-step flow of communication model, Katz (1957) proposed that, in general, ideas first flow from mass media to opinion leaders, who in turn spread them to a segment of the population that they influence. Even on social media, where ideas seem to flow directly between peers, the influence of opinion leaders has been found in several areas. Numerous studies consider retweets on social media as indicators of social influence (Cano et al., 2014; Yin & Zhang, 2012; Zhang et al., 2015). If retweets signify social influence, Twitter data greatly violates the independence assumption above because there are more retweets than original tweets (Firdaus et al., 2018; Shioda & Minamikawa, 2019). In addition, the distribution of retweets follows a power law (Lu et al., 2014), meaning that not all the original tweets or tweeters are equally retweeted. In particular, a few influential tweeters generate most retweets, and content popularity on Twitter can therefore be due to these influencers.

Power law and similar distributions can be created by cumulative advantage mechanisms, also known as the Matthew effect (Perc, 2014). This has been hypothesised to occur for academic citations because people are more likely to cite highly cited papers because they are better known, or for other reasons not related to the quality of the paper. The Matthew effect in citation counts has led to criticism of the use of citation counts as indicators of research significance, as they unfairly reward researchers with greater social capital or fame rather than reflecting the actual quality or impact of their work. However, Wang (2014) argued that while there are components in the Matthew effect that challenge the validity of citation counts as indicators of research importance, the usefulness of a paper to others, such as to provide foundations, does not undermine the validity of citation counts as valid indicators of research importance. While the Matthew effect and the underlying influences, and therefore the validity of citation counts as indicators of research significance, are well understood in citation counts, they remain largely unexplored in altmetric research.

This study contributes to this area by investigating whether social influence and the Matthew effect, manifested in retweets, in Twitter mentions of research papers challenges the validity of Twitter mentions (tweet counts and tweeter counts) as crowd-sourced altmetric indicators of research significance.

2. LITERATURE REVIEW

2.1. Wisdom of Crowds or Stupidity of Herds

Social media offers unprecedented opportunities for the development of crowd-sourced recommendation solutions. Research on the wisdom of crowd effect indicates that when certain conditions are met, the collective judgments of a group can be more accurate than the judgments of individuals (Galton, 1907; Hertwig, 2012; Lorenz et al., 2011; Surowiecki, 2005). Azar and Lo (2016) showed the wisdom of Twitter crowds by using tweet data to successfully predict stock market returns. One crucial requirement for the wisdom of crowd effect is independence (Surowiecki, 2005). A study by Lorenz et al. (2011) found that social influence can reduce the wisdom of crowd effect over time. This occurs because social influence tends to cause individuals' estimates to converge, decreasing group opinion diversity. Consequently, the group's overall accuracy decreases due to the loss of independent judgments.

Golub and Jackson (2010) argued that a society is "wise," meaning all opinions in a society converge to the truth, only when the influence of the most influential agent vanishes as the society grows. For example, if individuals listen to only one particular agent, their beliefs will then converge to that agent's initial information and have a substantial probability of deviating substantially from the truth. When individuals follow the actions or decisions of others, regardless of their private information, it is known as herd behavior, as identified by Banerjee (1992). Yu and Chen (2021) explored the impact of big data on the quality of society-level decision-making. The study revealed that when big data technology reduces the attention cost only for shallow information, but not for deep information, it leads to an increase in herd behaviour among individuals. This results in inadequate exploitation of available information and ultimately leads to a decline in the quality of society-level decisions.

Social influence on social media platforms has been extensively studied, with retweets often used as indicators of influence (Venkatesan et al., 2021; Yin & Zhang, 2012). According to Venkatesan et al. (2021), a retweet reflects the influence of the original tweet as well as the original tweet author, as the author was cited along with the content. Prior research on the dynamics of tweeters suggests that retweeting is the core driver of content popularity on Twitter, and there are more retweets than original tweets about real-world phenomena (Firdaus et al., 2018; Shioda & Minamikawa, 2019). Shioda and Minamikawa (2019) also found that retweets accounted for 65% to 86% of all tweets discussing real-world news events, such as "earthquake" and "Miss Universe." For health-related topics, Albalawi et al. (2019) collected two datasets of tweets containing health-related Arabic keywords, each of which included around 200,000 tweets. In their datasets, original tweets comprised 20% to 27% of all tweets.

A similar trend was found in an early dataset of COVID-19 tweets. Out of 74,758 tweets, only about 12% were original tweets and the remaining 88% were retweets (Prager et al., 2021). A frequency analysis of tweets relating to COVID-19 "intubation box(es)," "aerosol box(es)," or "intubation barrier(s)" indicated that, for 3,899 original tweets, there were 18,831 retweets (Mariano et al., 2021). But in a rare case where original tweet counts were slightly higher than retweet counts, Yoon et al. (2019) reported that 52% of all tweets were original in their study of the Zika virus. For altmetrics, Pandian et al. (2019) studied Twitter mentions of psychology research articles and found that retweeting increases the number of Twitter mentions received by articles. Although previous studies have suggested that the popularity of content on Twitter depends on retweets and most tweets are retweets, the proportions of original tweets/tweeters and retweets/retweeters in the overall tweet or tweeter counts of research papers are largely unexplored in general, and in the medical subject area in particular. The following research question addresses this gap:

RQ1: To what extent is the tweeting of COVID-19 medical research papers an independent process in terms of the proportions of original tweets/tweeters and retweets/retweeters?

2.2. The Matthew Effect in Citation Counts and Twitter Mentions

In addition to the findings that most tweets are retweets, the distribution of retweet counts tends to be highly skewed, following a power law (Mathews et al., 2017): Not all original tweets or tweeters are equally retweeted, but retweeting is concentrated on a few. In altmetrics research, Pandian et al. (2019) found that the popularity of psychology research articles on Twitter was primarily determined by their most popular tweet. To understand the challenge imposed by the power-law distribution of retweet counts on the validity of Twitter mentions as indicators of research significance, we first need to understand the similar phenomena in citation counts. Merton (1973, p. 446) defined the Matthew effect in science as "the accruing of greater increments of recognition for particular scientific contributions to scientists of considerable repute and the withholding of such recognition from scientists who have not yet made their mark."

Wang (2014) identified three components in the Matthew effect in citations: Networking, prestige, and appropriateness. For the networking effect, Blashfield (1982, p. 3) found that a highly cited paper was mostly cited by closely connected authors, known as the "invisible college," which is "a collection of intellectuals who had a sense of allegiance to each other and who frequently interacted both professionally and socially." Several studies have also provided strong evidence of the influence of the author, institution, and journal prestige on citation counts (Larivière & Gingras, 2010; Medoff, 2006; Tanner-Smith & Polanin, 2016). Due to the networking and prestige components of the Matthew effect, research evaluation based on citations has been criticized as rewarding fame or social capital rather than the intrinsic value of a paper, and so "citation-based metrics should not be used as fair tools for research evaluations" (Wang, 2014, p. 329). Appropriateness, in the sense of the usefulness of a paper to others, such as providing foundations that others can build their career on, is a third component of the Matthew effect (Wang, 2014). The appropriateness effect may also be cumulative, as the more people pursue the problems raised by the paper, the more opportunities for further enquiries and the higher the appropriateness of the paper. When discussing the Matthew effect in citations, it is the networking and prestige components that challenge the validity of citation metrics, not the appropriateness component (Wang, 2014).

While the distribution of altmetric data is known to follow power law (Banshal et al., 2022), the distribution of retweets in altmetric data is largely unexplored in many areas in altmetrics research, including for medical research. This leads to the next research question:

RQ2: Does the distribution of retweets suggest the existence of skewing in the data, in that some original tweeters are retweeted disproportionately more than others?

2.3. Credibility Heuristic Cues of Original Tweeters and Retweeting

One way to find conditions that lead to a power law or similar highly skewed distributions of retweet counts is to identify the characteristics of original tweeters (i.e., the authors of the original tweets) that are associated with retweet rates. One Twitter user characteristic known to be linked to retweeting and information sharing on Twitter is source credibility (Boehmer & Tandoc, 2015; Son et al., 2020). In the traditional sense, the source of a research article is the journal in which it is published. But on social media and the Internet, multiple layers of sources are usually involved in the spread of information. On Twitter, in addition to the credibility of the tweeter and the tweet content, studies based on dual process theories suggest that some peripheral cues (e.g., academic credentials) of a visible source (e.g. the user tweeting the content) also influence audience perceptions and decisions to retweet (Son et al., 2020).

Dual process theories, such as the heuristic-systematic model (HSM), suggest that people process information in a dual fashion, in systematic and heuristic modes (Chaiken, 1980). In the systematic mode, people expend considerable cognitive effort to analyse content to form judgments. In contrast, while in heuristic mode, they rely on easily available peripheral "cues" about the content to form quick judgements. Here, "heuristics" are generalizations of knowledge developed from previous experience (Chen & Chaiken, 1999), and "cues" serve to trigger them. For example, "experts' statements can be trusted" is an expertise heuristic that can be invoked by the expert source cue in the presentation of the message (Sundar, 2008).

Taking the heuristic approach, the MAIN model of Sundar (2008) argues that technology can assimilate information about multiple sources in the communication process that are influential in shaping users' perceptions and credibility assessments of sources and the underlying content. In the MAIN model, Sundar (2008) proposed four broad technology affordances: Modality, Agency, Interactivity, and Navigability, each of which is a repository of cues that have been demonstrated to trigger cognitive heuristics for credibility judgements. Agency affordances of technology are cues attributed to sources involved in the chain of online communication that trigger corresponding heuristics, which in turn affect audience perceptions about the credibility of the information provided by them. For example, users make quick judgments about the credibility of a social media post or an online website via the expertise heuristic, cued when they see an expert as the source of information.

On Twitter, account authoring tweets were investigated as sources of information, and the authority, identity, and bandwagon cues of those sources were found to be linked to credibility perceptions of the audience, supporting the MAIN model (Lin et al., 2016). Son et al. (2020) empirically examined the influence of reputation, social presence, and recency cues of sources on retweet decisions, finding strong evidence for the importance of tweeter profile information in retweet decisions during disastrous situations. More interestingly, Remy et al. (2013) found that the parameters of the power-law distribution of retweets can be estimated by the follower count of Twitter users, which is a bandwagon cue for tweeters that invokes bandwagon heuristics associated with credibility perception, according to the MAIN model. Their study suggested that the more followers a user has, the higher the probability of their tweets being highly retweeted, triggering power-law behaviour in retweets.

Thus, to better understand how original tweeters influence the retweeting of research papers, which ultimately affects the overall tweet count or tweeter counts discussing the papers, this research investigates the relationship between the credibility heuristic cues of the original tweeters and their retweet counts. Specifically, this study examined the following cues of original tweeters in relation to the retweet counts they receive.

2.3.1. Authority Cues

The influence of source authority cues on audience credibility perceptions has been extensively demonstrated in the literature. In one study, Lin et al. (2016) found that authority cues affected audience source credibility perceptions most strongly in the case of tweets containing health risk information related to the rise of drug-resistant gonorrhoea. As for COVID-19 information sharing intentions, Lu et al. (2021) found that source trust was also positively associated with information sharing intention, and health professionals were the most-trusted source, followed by academic institutions, government agencies, and news media. This leads to the following hypothesis:

H1: Authority cues of original tweeters positively associate with their retweet counts.

2.3.2. Bandwagon Cues

This cue is based on the psychology of "other users" and collective endorsement. Sundar (2008) claimed that people tend to have a favourable attitude toward products or content endorsed by many people. Thus, bandwagon cues are mainly popularity indicators, such as the number of followers, through which credibility is established. Lee and Sundar (2013) tested the number of followers that a source has as a bandwagon cue associated with credibility assessments, and found that it was associated with the perceived credibility of health content. In a separate study on a different topic, more-followed tweeters were found to be more likely to be retweeted (Son et al., 2020). This leads to the following hypothesis:

H2: Bandwagon cues of original tweeters positively associate with their retweet counts.

2.3.3. Social Presence Cues

In times of crisis, such as the global outbreak of a disease, social media serves not only as an information-sharing tool but also as a place to get moral support (Mukkamala & Beck, 2017). Consistent posting is known to be favoured by social media algorithms (Teah Hopper Consulting, 2018), and the more people see someone's tweets, the stronger the sense of togetherness everyone feels (Son et al., 2020). For example, Jahng and Littau (2016) found that highly interactive journalists were perceived as more credible than their less interactive counterparts on Twitter. Son et al. (2020) found significant relationships between social presence cues (status count and affiliation length) and retweet rates on Twitter. This leads to the following hypothesis:

H3: Social presence cues of original tweeters positively associate with retweet counts.

In addition, previous studies have indicated that there is an interaction effect between bandwagon cues and authority cues on the credibility of messages (Kim & Sundar, 2016; Lee & Sundar, 2013). Lee and Sundar (2013) studied the perceived credibility of health messages on Twitter and found that the presence of authority cues (professional accounts) with large numbers of followers (bandwagon cues) was more credible. In contrast, Kamiński et al. (2021) found that when it comes to tweets on COVID-19, those from celebrities with the largest number of followers were retweeted the most, outperforming those from health and scientific organizations, indicating that authority cues alone may not lead to high retweet rates. Therefore, interaction effects between authority cues (present vs. absent) and bandwagon cues (follower counts) on retweet counts were also tested in the present study. It is expected that among the original tweeters with authority cues, those with high bandwagon cues (high follower counts) will be retweeted more.

H4: Bandwagon cues moderate the relationship between authority cues and retweet counts.

3. METHODS

COVID-19 articles from a single major medical journal, *The Lancet*, were selected to control for journal and topic-related influences on the number of tweets and retweets received by the articles. Due to the importance of rapidly disseminating medical research about COVID-19 on Twitter, this topic should generate high Twitter mentions, giving higher statistical power to the study, although it will not generalise. Previous literature has indicated that the availability of data, and thus the potential usefulness of the altmetrics, varies across subject areas and journals (Htoo & Na, 2017; Thelwall et al., 2013). More Twitter interactions occur for articles from medical subject areas (Thelwall et al., 2013) and for articles from higher ranking journals (Htoo & Na, 2017), supporting the likely high volume of data. *The Lancet* is a well-known and important general medical journal. It was the highest-ranking journal in the medicine (general & internal) category, according to the 2021 Journal Impact Factor and so was selected in preference to other similarly important general medical journals.

3.1. Data Collection

The Scopus keyword query "COVID-19," matching the title, abstract or keyword fields of articles from 2020 with the results filtered to *The Lancet,* identified 53 articles (document type=article). All tweets (tweet IDs) mentioning these 53 articles were collected from Altmetric.com via its applications programming interface (API) by matching article DOIs. There were 246,840 tweet IDs collected from Altmetric.com for the 53 articles. For each tweet ID, the corresponding tweet and user information was retrieved from Twitter via its API. This included the retweet count, retweeted status (to differentiate between original tweets and retweets), and for retweets, the original tweet ID and original tweeter details. For tweet IDs that were no longer valid, no information was collected. Details for 238,339 tweets (177,011 retweets and 61,328 original tweets) for 53 articles were successfully collected from Twitter.

Unfortunately, Altmetric.com's coverage was found to be incomplete in the data collected via its API. For instance, one original tweet (tweet ID: 1285207186591887360) had 9,863 retweets according to Twitter, but only 2,050 were associated with it in the data collected from Altmetric.com's API, revealing a significant gap in its coverage. Additionally, some tweets included in the data from Altmetric.com were no longer available on Twitter, further limiting its coverage. These limitations are not unique to this study, but apply to all studies using data from Altmetric.com and likely other altmetric data providers as well.

3.2. Sampling Original Tweeters

To address RQ3 and hypotheses H1, H2, and H3 concerning the nature of the influence of original tweeters, a sample of original tweeters was selected for a manual content analysis to help find the relationship between their retweet counts and their credibility cues.

Among 177,011 retweets in the dataset, 867 (0.49%) were self-retweets. Although this is a very small number and unlikely to affect the results, self-retweets were filtered out before sampling original tweeters. After filtering out self-retweets, there were 176,145 retweets in the dataset. For each retweet in the dataset, Twitter provides a corresponding original tweet ID, retweet count, and original tweeter information. There were 332 different retweet counts, giving 332 groups of original tweeters. The group with the highest retweet count (9,863) had one member (*The Lancet*). In the groups with the lowest retweet counts (1 and 0), there were 4,494 members with 1 retweet and 4 members with no retweets. When sampling original tweeters, one tweeter from each group was selected. In groups with more than one user, one user was selected randomly. The same user was selected only once. For users appearing in more than one group, their average retweet count was calculated and used for analysis. This process produced 265 original tweeters for analysis.

3.3. Measures of Credibility Heuristics Cues

3.3.1. Authority Cues

Perceived credibility is often investigated with heuristic cues based on official, professional, and academic statuses. Lee and Sundar (2013) explored the effects of three authority cues on the perceived credibility of tweeted health messages. They used the affiliation "doctor" (professional status) as a high-authority cue. Similarly, Zhang et al. (2021) explored how users judge information quality in academic social networking sites and used a doctorate or master's degree (academic status) as a high authority cue. Likewise, Lin et al. (2016) evaluated the effects of three agency cues (authority, identity, and bandwagon) on the perceived credibility of tweets about drug-resistant gonorrhoea. They operationalized The Center for Disease Control and Prevention (CDC) as the expert authority cue, given that it is an "official" Twitter account (official status) associated with the topic.

When it comes to authoritative information, journalists and the media are also well established, although their status may have eroded recently. When examining the role of source credibility in the vaccination debate, Stueckemann (2019) found that well-established news outlets had higher credibility evaluations. Based on the previous literature, this study uses health officials, health professionals, academic, and established media sources as authority cues, as described in the Appendix. Original tweeters in the sample were coded by two independent coders using the coding instructions in the Appendix. Interrater reliability, estimated using Cohen's kappa, was 0.7 for the medical category, 0.7 for the academic category, 0.4 for the government/public health/medical official category, and 0.7 for the news media category. Table 1 shows the frequency of authority cues by category.

The kappa results were interpreted as: Values ≤0 no agreement, 0.01-0.20 none to slight, 0.21-0.40 fair, 0.41-0.60 moderate, 0.61-0.80 substantial, and 0.81-1.00 almost perfect agreement (Landis & Koch, 1977). Cohen's suggested interpretation implies that a score as low as 0.41 might be acceptable (McHugh, 2012). The kappa values for all categories, except for government/public health/medical official, are above 0.41. Since the government/public health/medical official category made up only 5%, it is unlikely that it will affect overall results. Therefore, all the categories are included in the analysis.

3.3.2. Bandwagon Cues

Adopting the measures by Lee and Sundar (2013), the number of followers is used in this study as a bandwagon cue.

3.3.3. Social Presence Cues

Adopting the measures by Son et al. (2020), status count (number of posts) and affiliation length/age of account (number of days since the account was created up to February 11, 2021) are used in this study as social presence cues. We chose February 11, 2021, as the cut-off date since it was the most recent date available in the dataset. Hence, all accounts in the dataset have an affiliation length or age of 0 or greater than 0.

3.4. Analysis

As supported by prior studies, retweets were considered an indicator of social influence in this study. Thus, RQ1 was addressed by finding the proportion of original tweets/tweeters and retweets/retweeters, and for RQ2 by plotting the distribution of retweet count to original tweets/tweeters in the dataset using the poweRlaw R package (Gillespie, 2015). Since a minimum retweet value of 1 is required to plot a power-law distribution, 1 was added to all retweet counts when plotting the retweet distribution. To investigate the effect of credibility heuristics cues on retweet counts, the Mann-Whitney U test and multiple linear regression were selected. Because the dependent variable retweet count was highly skewed, the Mann-Whitney U test was used to test the difference in median retweet counts between the two groups: The one with authority cues and the one without.

When choosing a suitable regression model in citation and altmetrics studies, as the data is skewed and the variance is greater than the mean, some altmetrics studies have used negative binomial multiple regression (Pandian et al., 2019). However, Thelwall and Wilson (2014) showed that ordinary least squares regression, after a log transformation, is the most suitable regression strategy for citation and altmetric data as it takes into account very high values, which are typical for skewed distributions. Thus in this study, multiple linear regression was conducted with log-transformed retweet counts ln(1+retweet count) as the dependent variable. Continuous independent variables (follower count, number of tweets posted, and number of days since the account was created) were mean-centred to make the results more interpretable, and to avoid multicollinearity when testing for interaction effects.

Table 1. Frequencies of authority cues by category (n=265)

Category	n (%)
Medical	77 (29)
Academic	115 (43)
Public health	13 (5)
Media	41 (16)

4. RESULTS AND DISCUSSION

4.1. Wisdom of Crowds or Stupidity of Herds

Out of 238,339 tweets for 53 articles, 61,328 (26%) were original tweets tweeted by 38,172 unique original tweeters, and the remaining 177,011 (74%) were retweets tweeted by 125,264 unique retweeters. All 53 articles had received at least 22 tweets, perhaps from publicity by the journal's Twitter account @TheLancet, which had around 724,000 Twitter followers as of August 2022. There were varying proportions of retweets between articles, from 40% to 93% (Fig. 1), in line with the general pattern of tweets and retweets ratios discussed above. Addressing RQ1, the results confirmed that tweeting research papers, at least for medical research papers relating to COVID-19, is not an independent process. In fact, the extent of influence, as revealed by the proportion of original tweets/tweeters and retweets/retweeters in total tweets/tweeters, is substantial. Fortunately, it is easy to address this issue in tweet/tweeter counts as altmetric indicators. In addition to reporting the total number of tweets/tweeters, as done by altmetric data aggregators like Altmetric.com, the number of original tweets/tweeters could also be reported.

4.2. Influence of Original Tweeters

Distributions of retweet counts for both original tweets and original tweeters were generated. Fig. 2A shows the distribution of retweet counts for original tweets, and the line follows a power law reasonably well (exponent=1.89) with p=0.43 for a goodness-of-fit test, failing to reject the null hypothesis that data is generated from a power-law distribution.

Since some original tweeters tweeted the same paper more than once, original tweeters were consolidated for each paper by taking their average retweet count for the same paper. After consolidation, there were 11,892 original tweeters associated with 177,011 retweets retrieved from Altmetric.com. Fig. 2B shows the distribution of retweet counts for 11,892 original tweeters, and the line has a power-law fit (exponent=1.89) with p=0.37 for a goodness-of-fit test, failing to reject the null hypothesis that data is generated from a power-law distribution. The distributions of retweets for both original tweets and original tweeters are very similar since most data points are the same in both cases. Overall, it is clear that a large majority of retweets were from a few highly retweeted influential original tweeters, addressing RQ2.

4.3. Credibility Heuristic Cues of Original Tweeters and Retweeting

Out of the 265 manually classified original tweeters, 72% had authority cues (belonging to at least one category in the Appendix) and 28% did not. Mann-Whitney U testing was applied to determine if there were differences in retweet counts for original tweeters with and without authority cues. The median retweet counts for the two groups were not significantly different (Table 2), rejecting H1.

Next, hierarchical multiple linear regression was conducted to test the main effects of all independent variables as well as interaction effects between authority cues (present vs. absent) and bandwagon cues (follower count) on logged retweet counts. Descriptive statistics for the variables are shown in Table 3. In the first step, all inde-

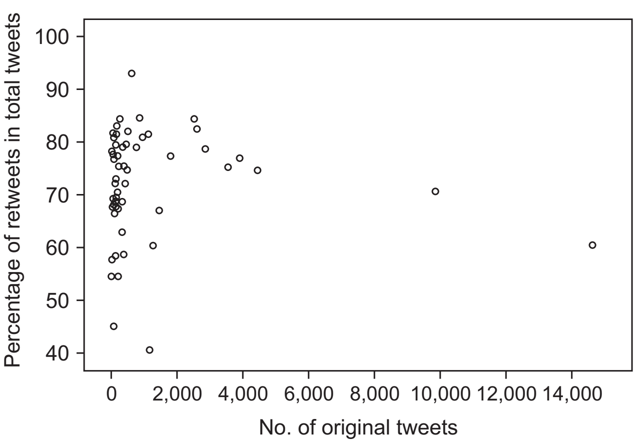

Fig. 1. Proportion of retweets vs. the number of original tweets received by the 53 papers investigated.

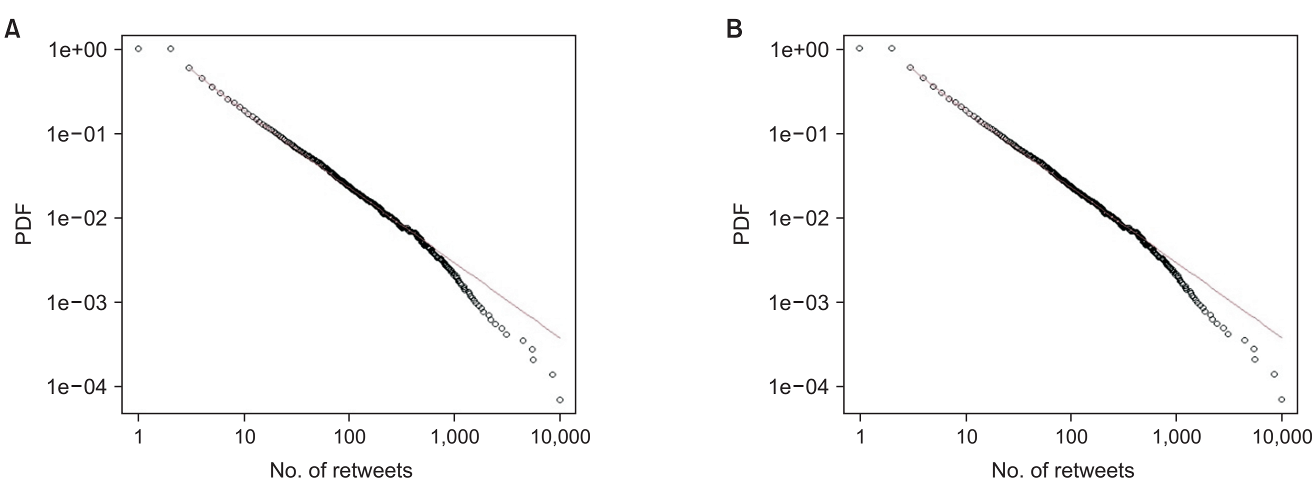

Fig. 2. Distribution of retweets. (A) Distribution of retweet counts per original tweet (probability distribution function [PDF] with a power-law fit [solid line] [exponent=1.89]). (B) Distribution of retweet counts per original tweeter (PDF with a power-law fit [solid line] [exponent=1.89]).

 https://doi.org/10.1633/JISTaP.2025.13.1.1

Table 2. Mann-Whitney U test results for retweet counts of original tweeters with and without authority cues

Group	n (%)	Median retweet count	Mean rank	Z	Mann-Whitney U	*p*
Users with authority cues	192 (72)	82.83	134.75	-0.60	6,672.50	0.55
Others	73 (28)	92	128.40			

Table 3. Descriptive statistics of variables

	Minimum	Maximum	Skewness	Mean	Medium	Standard deviation
Dependent variable						
Retweet count (original variable not included in the analysis)	0	2,826	4.23	198.10	85	330.70
Logged retweet count (transformed variable used in the analysis)	0	7.95	-0.11	4.46	4.45	1.33
Independent variables						
Follower count	5	7,964,976	9.57	179,124.99	36,784	601,546.43
Status count	178	474,770	3.24	44,836.74	19,923	70,127.90
Age of account	0	5,102	-0.75	3,123.76	3,410.76	1,279.04

Table 4. Results of hierarchical regression analysis

Step	Predictors	Standardized coefficients		R^2/R^2 change	*F*	*p*	Tolerance/variance inflation factor
		β	*p*				
1				0.094/0.094	6.70	<0.001	
	Authority	0.03	0.612				0.954/1.04
	Follower count	0.248	<0.001				0.983/1.02
	Status count	-0.152	0.018				0.862/1.16
	Age of account	0.131	0.037				0.889/1.13
2				0.11/0.01	6.181	<0.001	
	Authority	0.024	0.689				0.951/1.051
	Follower count	0.581	0.001				0.105/0.105
	Status count	-0.144	0.024				0.859/1.165
	Age of account	0.121	0.055				0.882/1.134
	(Follower count* Authority)	-0.351	0.053				0.106/9.446

*The interaction between variables.

pendent variables were entered as independent predictors of logged retweet counts (Table 4). In the second step, the interaction effect was tested by entering the product of the authority cues (present vs. absent) and mean-centred follower count as an additional predictor (Table 4).

The regression results showed that both bandwagon cues (follower count) and social presence cues (status count and age of account) are significant predictors of retweet counts at a significant level p=0.05 in step 1 (Table 4), supporting H2 and H3. The strongest predictor of higher retweet rates was the follower count. Although the main effect of authority cues was not significant, the interaction effect between authority cues and bandwagon cues (follower counts) was close to significant with a p-value

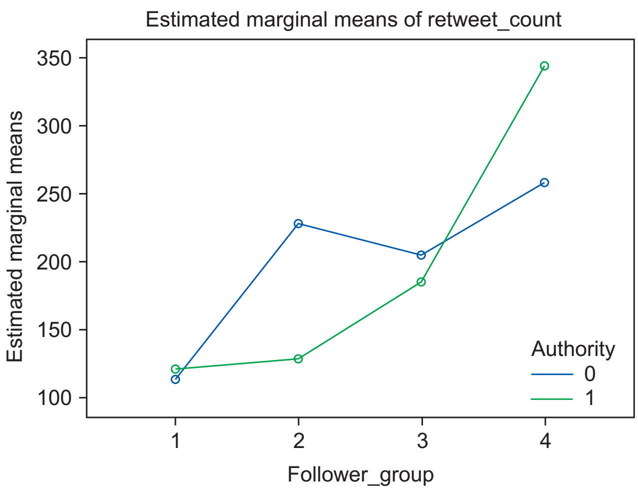

Fig. 3. Interaction between authority cues and bandwagon cues (follower count) on retweet counts.

of 0.053 in step 2 (Table 4), providing marginal support for H4. To plot the interaction effect, the follower count variable was recoded as a categorical variable with 4 levels. The original tweeters in the first quartile with the lowest number of followers were coded as 1, those in the second quartile as 2, the third quartile as 3, and the fourth quartile with the highest number of followers was coded as 4.

As shown in Fig. 3, in the presence of authority cues, the higher the follower count of original tweeters, the more retweets they receive. The effect of authority cues is strongest for original tweeters with the highest follower count and weakest for those with the lowest follower count. In contrast, for original tweeters with no authority cues, the retweet rate does not seem to follow the pattern of follower count consistently. But among them, in general, the top 75% with the highest follower count received similar retweet rates. Although H1 is not supported, based on the findings supporting H2, H3, and H4, it can be concluded that, overall, the results offered evidence for credibility cues of original tweeters as significant factors associated with retweet rates, addressing RQ3. These results support previous research that highlights the role of credibility heuristic cues in shaping the magnitude of retweets received by tweeters.

5. CONCLUSIONS

Although based on a single topic, year, and journal, this study found evidence of social influence, as manifested by retweets, in the overall tweet/tweeter counts of research papers. The high proportion of retweets and the power-law-like distribution of retweets to original tweets and tweeters in the data is consistent with herd-like behaviour and against the wisdom of crowd effect. However, as for citation counts, not every condition giving rise to a power law challenges the validity of the indicators. Further analysis on the characteristics of original tweeters associated with retweets revealed that the retweet rate was associated with the credibility cues of original tweeters, which are associated with the credibility perception of the retweeters. If retweets are mostly positively associated with credible original tweeters with the expertise to identify important works in the field, retweets can be seen as reinforcements of the importance of the papers to retweeters. In this case, even if Twitter mentions are heavily influenced by retweets, this may not undermine their validity as indicators of research significance.

Examining the findings from another perspective, associations between credibility heuristic cues of original tweeters and retweet rates suggest that retweets are more likely influenced by the perceived credibility of the original tweeters rather than the quality of the research papers they share. This raises an important question: Given that a large majority of tweets are retweets, if retweets are influenced by the credibility of the original tweeters rather than the quality/significance of research papers, can tweets be considered valid indicators of the quality/significance of research papers? They are, evidently, more directly indicative of the extent to which credible tweeters have tweeted the papers. The findings of this research support separating original tweets from retweets, as previously suggested by Fang et al. (2020) and Haustein (2019). The current research goes further by revealing the need to re-evaluate the use of retweets as indicators for research papers. The implicit assumption for including retweets within altmetric indicators is that retweets reflect the significance of research papers to some extent. The current paper's suggestion that they may more directly reflect the credibility of the original tweeters undermines this claim to some extent.

CONFLICTS OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCES

Albalawi, Y., Nikolov, N. S., & Buckley, J. (2019). Trustworthy health-related tweets on social media in Saudi Arabia:

Tweet metadata analysis. *Journal of Medical Internet Research,* 21(10), e14731. https://doi.org/10.2196/14731

Azar, P. D., & Lo, A. W. (2016). The wisdom of Twitter crowds: Predicting stock market reactions to FOMC meetings via Twitter feeds. *Journal of Portfolio Management,* 42(5), 123-134. https://doi.org/10.3905/jpm.2016.42.5.123

Banerjee, A. V. (1992). A simple model of herd behavior. *Quarterly Journal of Economics,* 107(3), 797-817. https://doi.org/10.2307/2118364

Banshal, S. K., Gupta, S., Lathabai, H. H., & Singh, V. K. (2022). Power laws in altmetrics: An empirical analysis. *Journal of Informetrics,* 16(3), 101309. https://doi.org/10.1016/j.joi.2022.101309

Blashfield, R. K. (1982). Feighner et al., invisible colleges, and the Matthew effect. *Schizophrenia Bulletin,* 8(1), 1-6. https://doi.org/10.1093/schbul/8.1.1

Boehmer, J., & Tandoc, E. C., Jr. (2015). Why we retweet: Factors influencing intentions to share sport news on Twitter. *International Journal of Sport Communication,* 8(2), 212-232. https://doi.org/10.1123/ijsc.2015-0011

Cano, A. E., Mazumdar, S., & Ciravegna, F. (2014). Social influence analysis in microblogging platforms – A topic-sensitive based approach. *Semantic Web,* 5(5), 357-372. http://doi.org/10.3233/SW-130108

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology,* 39(5), 752-766. https://doi.org/10.1037/0022-3514.39.5.752

Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In S. Chaiken, & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 73-96). Guilford Press.

Fang, Z., Dudek, J., & Costas, R. (2020). The stability of Twitter metrics: A study on unavailable Twitter mentions of scientific publications. *Journal of the Association for Information Science and Technology*, 71(12), 1455-1469. https://doi.org/10.1002/asi.24344

Firdaus, S. N., Ding, C., & Sadeghian, A. (2018). Retweet: A popular information diffusion mechanism – A survey paper. *Online Social Networks and Media,* 6, 26-40. https://doi.org/10.1016/j.osnem.2018.04.001

Galton, F. (1907). Vox populi. *Nature,* 75(1949), 450-451. https://doi.org/10.1038/075450a0

Gillespie, C. S. (2015). Fitting heavy tailed distributions: The poweRlaw package. *Journal of Statistical Software,* 64(2), 1-16. https://doi.org/10.18637/jss.v064.i02

Golub, B., & Jackson, M. O. (2010). Naïve learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics,* 2(1), 112-149. https://doi.org/10.1257/mic.2.1.112

Gunn, A., & Mintrom, M. (2016). Higher education policy change in Europe: Academic research funding and the impact agenda. *European Education,* 48(4), 241-257. https://doi.org/10.1080/10564934.2016.1237703

Haustein, S. (2019). Scholarly Twitter metrics. In W. Glänzel, H.F. Moed, U.Schmoch, & M. Thelwall (Eds.), *Springer handbook of science and technology indicators* (pp. 729-760). Springer. https://doi.org/10.1007/978-3-030-02511-3_28

Hertwig, R. (2012). Tapping into the wisdom of the crowd—With confidence. *Science (New York, N.Y.),* 336(6079), 303-304. https://doi.org/10.1126/science.1221403

Htoo, T.H.H. & Na, J. C. (2017). Disciplinary differences in altmetrics for social sciences. *Online Information Review*, 41(2), 235-251. https://doi.org/10.1108/OIR-12-2015-0386

Jahng, M. R., & Littau, J. (2016). Interacting is believing: Interactivity, social cue, and perceptions of journalistic credibility on Twitter. *Journalism & Mass Communication Quarterly,* 93(1), 38-58. https://doi.org/10.1177/1077699015606680

Kamiński, M., Szymańska, C., & Nowak, J. K. (2021). Whose Tweets on COVID-19 gain the most attention: Celebrities, political, or scientific authorities? *Cyberpsychology, Behavior and Social Networking,* 24(2), 123-128. https://doi.org/10.1089/cyber.2020.0336

Katz, E. (1957). The two-step flow of communication: An up-to-date report on an hypothesis. *Public Opinion Quarterly,* 21(1), 61-78. https://doi.org/10.1086/266687

Kim, H. S., & Sundar, S. S. (2016). Motivating contributions to online forums: Can locus of control moderate the effects of interface cues? *Health Communication,* 31(5), 583-595. https://doi.org/10.1080/10410236.2014.981663

Lamont, M. (2009). *How professors think: Inside the curious world of academic judgment*. Harvard University Press.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics,* 33(1), 159-174. http://doi.org/10.2307/2529310

Larivière, V., & Gingras, Y. (2010). The impact factor's Matthew effect: A natural experiment in bibliometrics. *Journal of the American Society for Information Science and Technology,* 61(2), 424-427. https://doi.org/10.1002/asi.21232

Lee, J. Y., & Sundar, S. S. (2013). To tweet or to retweet? That is the question for health professionals on Twitter. *Health Communication,* 28(5), 509-524. https://doi.org/10.1080/10410236.2012.700391

Lin, X., Spence, P. R., & Lachlan, K. A. (2016). Social media and credibility indicators: The effect of influence cues. *Computers in Human Behavior,* 63, 264-271. https://doi.org/10.1016/j.chb.2016.05.002

Lorenz, J., Rauhut, H., Schweitzer, F., & Helbing, D. (2011). How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences of the United States of America,* 108(22), 9020-9025. https://doi.org/10.1073/pnas.1008636108

Lu, L., Liu, J., Yuan, Y. C., Burns, K. S., Lu, E., & Li, D. (2021). Source trust and COVID-19 information sharing: The mediating roles of emotions and beliefs about sharing. *Health Education & Behavior,* 48(2), 132-139. https://doi.org/10.1177/1090198120984760

Lu, Y., Zhang, P., Cao, Y., Hu, Y., & Guo, L. (2014). On the frequency distribution of retweets. *Procedia Computer Science,* 31, 747-753. https://doi.org/10.1016/j.procs.2014.05.323

Mariano, E. R., Kou, A., Stiegler, M. A., & Matava, C. (2021). The rise and fall of the COVID-19 aerosol box through the lens of Twitter. *Journal of Clinical Anesthesia,* 69, 110145. https://doi.org/10.1016/j.jclinane.2020.110145

Mathews, P., Mitchell, L., Nguyen, G. T. K., & Bean, N. G. (2017, April 3-7). The nature and origin of heavy tails in retweet activity. *Proceedings of the 26th International World Wide Web Conference* (pp. 1493-1498). International World Wide Web Conferences Steering Committee.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica,* 22(3), 276-282.

Medoff, M. H. (2006). Evidence of a Harvard and Chicago Matthew effect. *Journal of Economic Methodology,* 13(4), 485-506. https://doi.org/10.1080/13501780601049079

Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. University of Chicago Press.

Mukkamala, A., & Beck, R. (2017). *Presence of social presence during disasters*. Paper presented at the 21th Pacific Asia Conference on Information Systems, Langkawi, Malaysia.

Pandian, N. D. S., Na, J. C., Veeramachaneni, B., & Boothaladinni, R. V. (2019). Altmetrics: Factor analysis for assessing the popularity of research articles on Twitter. *Journal of Information Science Theory and Practice,* 7(4), 33-44. https://doi.org/10.1633/JISTaP.2019.7.4.3

Perc, M. (2014). The Matthew effect in empirical data. *Journal of the Royal Society Interface,* 11(98), 20140378. http://doi.org/10.1098/rsif.2014.0378

Prager, R., Pratte, M. T., Unni, R. R., Bala, S., Ng Fat Hing, N., Wu, K., McGrath, T. A., Thomas, A., Thoma, B., & Kyeremanteng, K. (2021). Content analysis and characterization of medical Tweets during the early COVID-19 pandemic. *Cureus,* 13(2), e13594. https://doi.org/10.7759/cureus.13594

Priem, J., Taraborelli, D., Groth, P., & Neylon, C. (2011). *Altmetrics: A manifesto*. https://digitalcommons.unl.edu/scholcom/185/

Remy, C., Pervin, N., Toriumi, F., & Takeda, H. (2013, December 2-5). Information diffusion on Twitter: Everyone has its chance, but all chances are not equal. In K. Yetongnon, A. Dipanda, & R. Chbeir (Eds.), *Proceedings of the 2013 International Conference on Signal-Image Technology & Internet-Based Systems* (pp. 483-490). IEEE.

Shioda, S., & Minamikawa, M. (2019, October 22-25). Features found in Twitter data and examination of retweeting behavior. In M. Alsmirat, & Y. Jararweh (Eds.), *Proceedings of the 6th International Conference on Social Networks Analysis, Management and Security (SNAMS)* (pp. 529-534). IEEE.

Son, J., Lee, J., Oh, O., Lee, H. K., & Woo, J. (2020). Using a heuristic-systematic model to assess the Twitter user profile's impact on disaster tweet credibility. *International Journal of Information Management,* 54, 102176. https://doi.org/10.1016/j.ijinfomgt.2020.102176

Stueckemann, E. (2019). *Examining the role of source credibility in the vaccination debate: An experimental study of the influence of heuristic cues on source credibility assessments and attitude change* (master's thesis). Jönköping University, Jönköping, Sweden.

Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger, & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 73-100). MIT Press.

Surowiecki, J. (2005). *The wisdom of crowds.* Anchor Books.

Tanner-Smith, E. E., & Polanin, J. R. (2016). Brief alcohol intervention trials conducted by higher prestige authors and published in higher impact factor journals are cited more frequently. *Journal of Clinical Epidemiology,* 75, 119-125. https://doi.org/10.1016/j.jclinepi.2016.01.028

Teah Hopper Consulting. (2018). *Why posting consistently matters on social media*. https://www.teahhopper.com/blog/2018/11/15/why-posting-consistently-matters-on-social-media

Thelwall, M. (2021). Measuring societal impacts of research with altmetrics? Common problems and mistakes. *Journal of Economic Surveys,* 35(5), 1302-1314. https://doi.org/10.1111/joes.12381

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. *PLoS One*, 8(5),64841. https://doi.org/10.1371/journal.pone.0064841

Thelwall, M., & Wilson, P. (2014). Regression for citation data: An evaluation of different methods. *Journal of Informetrics,* 8(4), 963-971. https://doi.org/10.1016/j.joi.2014.09.011

Venkatesan, S., Valecha, R., Yaraghi, N., Oh, O. O., & Rao, H. R. (2021). Influence in social media: An investigation of Tweets spanning the 2011 Egyptian revolution. *Manage-*

ment Information Systems Quarterly, 45(4), 1679-1714. https://aisel.aisnet.org/misq/vol45/iss4/6/

Wang, J. (2014). Unpacking the Matthew effect in citations. *Journal of Informetrics,* 8(2), 329-339. https://doi.org/10.1016/j.joi.2014.01.006

Yin, Z., & Zhang, Y. (2012, September 3-5). Measuring pairwise social influence in microblog. In R. Bilof (Ed.), *Proceedings of the 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing* (pp. 502-507). IEEE.

Yoon, J., Hagen, L., Andrews, J., Scharf, R., Keller, T., & Chung, E. K. (2019). On the use of multimedia in twitter health communication: Analysis of tweets regarding the zika virus. *Information Research,* 24(2), Article 823. http://informationr.net/ir/24-2/paper823.html

Yu, T., & Chen, S. H. (2021). Big data, scarce attention and decision-making quality. *Computational Economics,* 57(3), 827-856. https://doi.org/10.1007/s10614-018-9798-5

Yu, T., Chen, S. H., & Wang, C. H. (2018). Information aggregation in big data: Wisdom of crowds or stupidity of herds. In E. Bucciarelli, S. H. Chen, & J. Corchado (Eds.), *Decision economics: In the tradition of Herbert A. Simon's heritage* (pp. 16-27). Springer.

Zhang, J., Tang, J., Li, J., Liu, Y., & Xing, C. (2015). Who influenced you? Predicting retweet via social influence locality. *ACM Transactions on Knowledge Discovery from Data,* 9(3), 1-26. https://doi.org/10.1145/2700398

Zhang, N., Yuan, Q., Xiang, X., & Chen, K. (2021). What can you perceive? Understanding user's information quality judgment on academic social networking sites. *Information Research,* 26(2), paper 896. https://doi.org/10.47989/irpaper896

Appendix: Coding Instrument

No.	Categories	Code
1	Medical - Medical professionals & organizations	1 if the account contains cues (words) indicating that the account represents a medical practitioner (such as MD, doctor, Dr.); a medical specialist; or an infectious disease specialist (OR) a medical department/group in a university, a research organization, or a hospital 0 otherwise.
2	Academic -Research/academic professionals & organizations (including academic publishers/ journals)	1 if the account contains cues indicating that the account represents a scientist, researcher, or academic professional or organization (e.g. PhD, scientist, researcher, or lecturer, but not teacher or name of a university). 0 otherwise.
3	Public health - Government officials/Public health professionals & organizations	1 if the account contains cues indicating that the account represents a monarch, president, prime minister, chancellor, minister of health, or head of health departments or organizations (OR) a public health official, professional, or a person with a public health doctorate degree (e.g., Doctor in public health) (OR) A public health-related organization 0 otherwise.
4	Media - Journalists/News organization	1 if the account contains cues indicating that the account represents a professional journalist or another similar professional science communicator (not blogger) (OR) A professional news media organization 0 otherwise.

 https://doi.org/10.1633/JISTaP.2025.13.1.1