# Believe it or Not? Analyzing Information Credibility in Microblogs

Byungkyu Kang
Department of Computer Science
University of California
Santa Barbara, CA 93106
bkang@cs.ucsb.edu

Tobias Höllerer
Department of Computer Science
University of California
Santa Barbara, CA 93106
holl@cs.ucsb.edu

John O'Donovan
Department of Computer Science
University of California
Santa Barbara, CA 93106
jod@cs.ucsb.edu

***Abstract*—This paper describes a series of experiments to identify and evaluate key factors that influence credibility perception in microblogs. Specifically, we report on a demographic survey (N=81) followed by two user experiments (N=102, N=646) in order to answer the following research questions: (1) What are the important cues that contribute to information being perceived as credible. (2) Can we separate these cues from the content and quantify their influence on credibility perception?, and (3) To what extent is such a quantification portable across different microblogging platforms? To answer the third question, we study two popular microblogs, *Reddit* and *Twitter*. Key results include that significant effects of individual factors can be isolated, are portable, and that links, profile pictures and image content are the strongest influencing factors in credibility assessment.**

Microblogs such as Twitter and Reddit are well established global sources of real-time news and information. As with all platforms that support user-provided content, they suffer from an abundance of noisy and unreliable data. With the emergence of microblog messages as serious sources for news, as indicators and early informers of natural phenomena such as earthquakes and severe weather, and even as medium for financial transactions (tweet-to-pay functionality was recently introduced by several banks), the importance of identifying credible information and information sources on microblog platforms continuously increases.

In recent years, microblogging services have transformed from online journal or peer-communication platforms [1] to powerful online information sources operating at a global scale in every aspect of society, largely due to the advance of mobile technologies. Today's technology enables instant posting and sharing of text and/or multimedia content, allowing people on-location at an event or incident to serve as news reporters. In fact, a recent study of traditional media journalist practice [2] shows that they rely heavily on social media for their information. Another study [3] reported that 53.8% of all U.S. journalists use microblogs to collect information and to report their stories. Several recent studies [4], [5] show how user-provided microblog content is an effective mechanism for understanding crisis situations such as earthquakes, hurricanes or political conflicts.

The mass proliferation of microblog usage also brought about a shift in the interaction mechanisms and information flow within them. In particular, a 2013 PEW research report [6] shows that an increasing number of users search microblogs by keyword or hashtag as opposed to the traditional content stream or message exchange practices. This means that a larger portion of information is coming from complete strangers, accessed via keyword matching than from sources that a user is actively following and are likely to be known by the user. This reduced window of information about the source presents a difficult challenge in assessing credibility of information, and requires a more comprehensive understanding of the components of a microblog message and their potential impact on the information consumer's assessment of information credibility. This is bolstered by the fact that the majority of users in Twitter (52%) and Reddit (60%) treat the system as their primary source of news information [6].

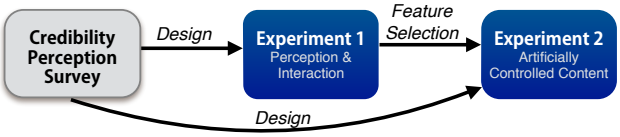



**Fig. 1:** Overview and dependencies between the initial survey (N=81), Experiment 1 (N=102) and Experiment 2 (N=646)

Recently there have been many efforts to study information credibility in microblogs, ranging from automated algorithms to model and predict credibility of users [7] and messages [8], [9] at general [8], [10] and topic-specific [9] levels, to visualization and interaction applications such as [11]. However, with the exception of [12], [13], little research has focused on isolating the impact of individual microblog features such as profile images, links and other available metadata on perceptions of credibility –a problem that is increasingly important as a growing portion of news information gets produced by people the information consumer does not know. It is in view of this problem that we attempt to answer the following research questions:

1) What are the important cues that contribute to information being perceived as credible on microblogs.
2) Can we separate them from the content and quantify their impact on a credibility assessment?
3) To what extent is such a quantification portable across different microblogging platforms?

*a) Contributions:* Figure 1 shows a high level overview of the three key phases in this study, and the dependencies between them. First, a crowd sourced survey of 81 users was performed to assess patterns in users assessments of information credibility across different topics and for demographic groupings. From this survey, a large set of microblog features was analyzed and a subset was identified for further evaluation.

Second, a user experiment (N=102) was designed to place

users in context across two microblog platforms and elicit a more refined set of salient factors that influenced their judgement of information credibility. To do this, heatmaps were computed from mouse click behavior in the microblog interfaces. To address the question of portability, we perform experiments across two of the most popular microblogs for news-consumption: *Reddit* and *Twitter*.

Third and finally, we conduct an experiment (N=646) in which variables from phase 2 are experimentally controlled. This allows us to assess the impact of the each individual variable on credibility assessments in a range of contexts.

Results of the three studies provide insight on the ways users evaluate and perceive information credibility in different contexts on microblog platforms.

## I. Related Work

Having described our experiment at a high level, we will now discuss it within the context of relevant related research, before detailing the experimental setup and results. We begin by defining credibility, and continue by comparing and contrasting different approaches for modeling information credibility in microblogs, at the computational and user-perception levels.

*b) Credibility Definition:* Information credibility is a concept that has received research attention from a variety of disciplines over many decades, and there are many conflicting definitions of the topic. A recent study by [14] provided a definition for credibility which has been frequently relied on by subsequent research. They describe credibility as "(a) a *perceived quality* (b) made up of multiple dimensions, mostly *trustworthiness* and *expertise*."

[14] claim that credibility does not reside in an object, a person or a piece of information. Similarly in this work, we consider credibility as a function of *perception* including the person and the object being perceived. Many studies of credibility (e.g.: [15]) find that it is comprised of two primary dimensions–*trustworthiness* and *expertise*.

*c) Credibility models:* [16] studied how people evaluate information credibility online and proposed the idea of *Prominence-interpretation theory* which, as the name suggests, is comprised of the two values: "prominence" (important or distinct entities) and "interpretation" (user assessment of the entities). Following from this, other researchers proposed models or algorithms that either measure or predict credibility of users in different contexts. For example, [7] build a model of location and topic affinity to identify credible, relevant users in crisis situations. [8] and [9] both propose models for identifying credible sources of news information based on computational models. Going beyond theory, some researchers place a focus on empirical evaluation of credibility and the problem of ground truth. For example, [10] proposed a two pronged approach to gathering ground truth information on the credibility of microblog data by combining manually annotated scores with observed network statistics (e.g: retweets) from the data to achieve a "more stable" estimate of credibility.

*d) Perceived Credibility:* While there is much discord over a standard definition of credibility across different disciplines, most researchers agree that "credibility" that is inherent to an entity and perceived credibility of that entity are not necessarily equivalent. The latter could be viewed as a subjective function of the former. Perceived credibility can fluctuate from (inherent) credibility based on the way in which the entity is represented, and based on the characteristics of the person making the credibility assessment. Accordingly, numerous researchers from different disciplines have attempted to identify a set of salient factors that contribute to our perception process. Visual and textual components have been studied in the same vein [17] to reveal complex relationships between data, metadata and context that inform our perception of information credibility. Other researchers have studies influence at the network level, identifying trends and patterns that lead to tipping points of credibility and popularity [18] and surges in influence within the network [19].

In recent years, features or cues that affect perceived credibility of information in microblogs have been studied [13], [20]. A common theme in these recent studies is to first select numerous candidate features that are likely to contribute to the assessment processes, and then analyze them in both qualitative and quantitative ways through online surveys or user studies. For example, [20] found disparity between the features considered for evaluating information credibility between search engines and Twitter. Morris also reported a controlled experiment that revealed the features through which users assess information credibility on Twitter. Their study also provided insight and suggestions for interface design to improve credibility perception from the end-user perspective. Perceived credibility has also been compared across different cultural settings by [13] Their study reports on experimental and survey data that compares and contrasts the impact of several features of microblog updates (authors gender, name style, profile image, location, and degree of network overlap with the reader) on credibility perceptions among U.S.(Twitter) and Chinese(Weibo) audiences. Their goal was to design new user experiences which can maximize both *credibility* (as a property of entity) and *perception of credibility* (an end-user subjective function, to which the entity is an argument) of the contents on social media. Perceived credibility can be impacted by personality characteristics such as those modeled by [21]

## II. Credibility Perception Survey

To gather a fair assessment of candidate features to evaluate in our main experiments, and to gather insight about credibility decisions in microblogs, we conducted a crowd sourced study targeting Twitter users in late 2013 using Amazon's Mechanical Turk (MTurk) platform. A total of 81 respondents were asked a series of questions to explore what information from microblogs they mostly consider when they need to search for credible information about particular events. The 59 male and 22 female participants were from different parts of the world, with a majority from the United States and India. Participant age ranged between 18 and 60 with an average of 28. 60% of the subjects used microblog on a weekly (22%) or daily (38%) basis. Participants reported basic information such as Twitter usage, educational and cultural backgrounds and yearly income.

The overarching goal of the survey was to explore the following general hypothesis through self-reported metrics and to identify the set of Twitter features (E.g: links, profile images

**TABLE I:** Primary use of information on Twitter

| | | | |
|---|---|---|---|
| Business | 22(27.2%) | Information Sharing | 21(25.9%) |
| Social Purpose | 16(20.0%) | Information Search | 15(18.5%) |
| Serendipitous Search | 4(4.9%) | Other | 3(3.7%) |

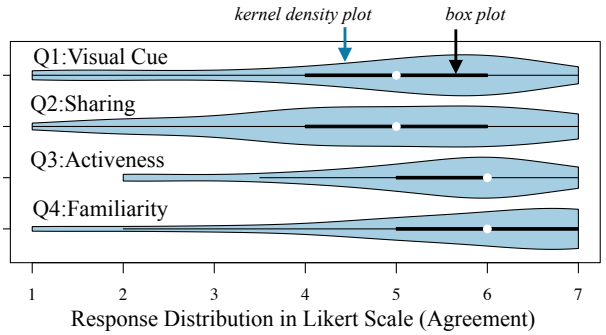



**Fig. 2:** Survey responses for key questions (marked with $*$ in Table II. Each line and enclosed curve shows the response distribution for questions in (1) a box-whisker-chart indicating medians and quartiles and (2) a kernel density plot, respectively. Responses are provided on a Likert scale (Strongly disagree (1) – Strongly agree (7)).

etc.) reported as most influential in credibility assessment. The two additional experiments in this paper expand the general hypothesis into 6 additional hypotheses, and were both designed based on analysis of the results from this survey. In the survey, 20% (16/81) of the participants reported that they consider visual cues as a major factor that affects their credibility assessments. Details of the resulting design decisions are discussed in Section III.

**Research Question (RQ) 1:** Does the display of metadata in microblogs influence perceived credibility of the associated content.

HYPOTHESIS. *Metadata display (textual or visual) influences perceived credibility of microblog content. The direction of influence is dependent on the specific content displayed.*

To gauge usage patterns and credibility perception, participants were asked 17 questions in a web survey, covering aspects such as activity rate, perceived impact of visual cues, and sharing frequency, among others. A selection of these survey questions are shown in Table II, and response distributions are shown in Figure 2. The results indicate that the majority of the participants consider themselves active information consumers on microblogs (79% for 'Activeness'). These users also share their own content frequently with their followers (57%, 'Sharing'). The most common usage reasons (Table I) were reported as business (27%) including online marketing, information sharing (26%) and social use (20%). Interestingly, 68% of participants reported that visual elements have significant impact on their credibility assessment in the microblog, as indicated by Q1 in Figure 2. Our population exhibited reasonably heavy use of Twitter (38% daily use) and a good standard of education across participants, with 67/81 possessing at least a bachelor's degree.

*e) Credibility Factors:* The main section of the survey investigated what kinds of attributes participants consider as a primary factor when they search for credible information on

| Label | Questions |
|---|---|
| Activeness* | Do you consider yourself as an active online information consumer? |
| Sharing Freq* | Do you frequently share your information with the people in your network (followers)? |
| Primary Usage | What is your primary usage of information on microblogs? |
| Familiarity* | Are you familiar with microblog services? |
| 1st Cred Factor | Which do you consider as a primary factor for measuring information credibility? |
| 2nd Cred Factor | Which do you consider as a secondary factor for measuring information credibility? |
| Visual Cues * | Do you think that visual cues are important for judging credibility? |
| Url Relevance | Do you think that the presence of URLs in a tweet, which point to an external information source, can enhance information credibility? |
| In-Person Friends | About how many of your "friends" on Twitter have you met in person? |
| Non-Human Friends | About how many companies or organizations do you currently follow on Twitter? |
| Celebrity Friends | About how many celebrities do you currently follow on Twitter? |
| Time-On-Others | On Twitter, about how much time do you spend looking at what other users have posted? |
| Time-On-Me | On Twitter, about how much time do you spend posting tweets about yourself? |

**TABLE II:** Survey Questions. * denotes further detail in Figure 2

microblogs. We can intuitively expect that both content and information source would be highly ranked and the results indeed support this. However, it is interesting to note that *visual cues* such as design and layout were also reported as influential in the process of credibility assessment. 10% of participants responded that design/layout was the primary factor (20% elected it a major factor) in their credibility assessments.

*f) Correlation Analysis:* People consider many different factors during credibility assessment with microblog information. Numerous researchers concluded that, ultimately, credibility can be perceived or measured in different ways based on the given context, cultural background, language, etc. [16]. We also find that many microblog users agree with this statement from pre-study offline interviews. Thus, we designed our questionnaire to find underlying correlation, if any, between demographic background and the question responses. Results are shown in Figure 3. Table II provides a full description of each element in the correlation plot of Figure 3. Some notable correlations include Twitter use and general information use. There was a positive correlation between employment type and content use –this may have been a result of the number of users who said they used the microblogs for marketing purposes. There was a strong correlation between locality (size of city lived in) and amount of information shared on microblogs. People in larger cities shared more information than those in small cities and towns. Predictably, employment type was positively correlated with primary usage of the microblog. We also find a correlation between gross income / ethnic origin and microblog usage, complementary to [22], who found strong correlation between these factors and browsing behavior. Demographic factors (both age and cultural background) correlated with usage rate, and with the impression of visual cues as an information credibility factor –younger people had higher usage rates and were more influenced by visual cues.

In summary, the initial analysis from the survey highlights visual cues as a useful factor for further study, incorporating aspects of content, and metadata about the source/provenance of microblog messages.

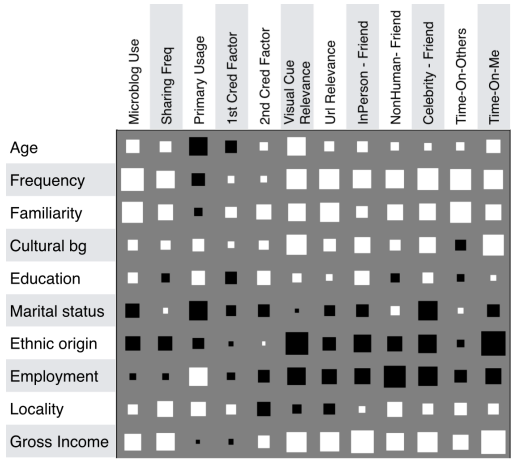



**Fig. 3:** Correlation matrix between the demographic information of the participants and the responses to the survey. The matrix is visualized in a Hinton Map (white and black squares represent positive and negative correlation, respectively. Square size is proportional to the absolute value of the score [0–1].)

## III. Experimental Setup

To further explore what factors influence credibility perception in microblogs, we designed and conducted two different user experiments guided by insights from the credibility perception survey. Both experiments were conducted on MTurk. In this section, we detail the design of both experiments (`Exp1` and `Exp2`) and discuss a refined set of experimental hypotheses.

### A. `Exp1`: *Perception and Interaction*

The initial survey highlighted that content (meaning) and sources (origin) of microblog posts are the most influential factors in credibility assessment. However, the representation of information such as metadata also plays a significant role in user perception of information credibility. According to these findings, we refine the initial research questions to include the following three questions/hypotheses.

**RQ2:** Do different features influence credibility by different amounts?

Hypothesis. *Features have varying degree of influence over credibility perception*

**RQ3:** What are the effects of different classes of microblog features on perceived credibility?

Hypothesis. *Visual factors will have the most influence, followed by network and content-based (text) factors.*

**RQ4:** Can our models of feature influence be ported successfully to different microblogs?

Hypothesis. *Influence of features is consistent across platforms*

To test these hypotheses, a study was designed to place users in a familiar/typical microblog context and provide them with a simple mechanism to highlight the specific features that they felt had an influence on their perception of content. To address the last hypothesis, the study was designed to be cross-platform, comparing features from Twitter and Reddit. Figure 4 shows two example interfaces from the study (N=102). Users were requested to click on or close to items that they felt had *any* impact on their perception of information, regardless of positive or negative direction, which is evaluated separately in our third experiment. They were given no a-priori information on specific feature lists. This mechanism for identifying influential features was used in an effort to avoid bias from manual or expert selection of a feature set (intended for detailed analysis in Experiment 2). Domains (Twitter and Reddit) were a between-subjects variable, and only participants with significant prior experience with a domain were allowed perform the task.

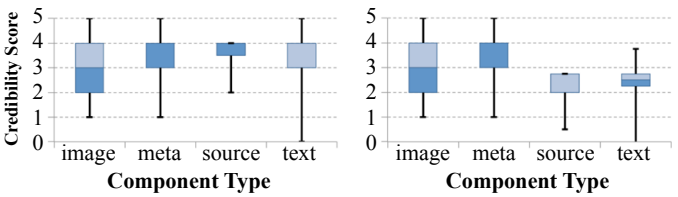


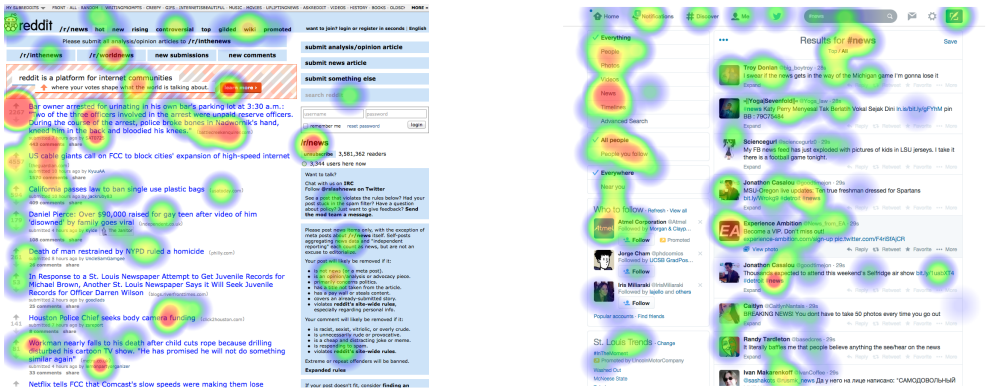

**Fig. 5:** Click frequency/perceived credibility by UI component type for (a) Twitter and (b) Reddit (`Exp1`)

First, participants were asked 6 general questions about their microblogging practice. Then they were shown 3-4 screenshot images of the microblog (3 for Reddit and 4 for Twitter). To capture the aforementioned features avoiding possible bias, we let participants select three visual elements instead of having them rank an arbitrary selection of features we provide. On each click, a slider selector was shown below the image to record the amount of impact the element that the user clicked on has on her credibility assessment. We collected coordinates of each click and its corresponding score in likert scale (0 for no effect to 5 for major effect).

*g) Feature extraction:* In order to extract meaningful features from this experiment, we analyzed the results from a heatmap visualization (Figure 4) and statistical analysis using five-number summaries (Figure 5). As can be seen in Figure 5, there is overall similarity in credibility ratings on different elements for both Reddit and Twitter users. However, users of Reddit express higher priority on both information sources and textual elements for their credibility assessments. This observation may be due to small differences between two social platforms: For example, most of the posts in Reddit are directly connected to external webpages and this makes the source (URLs) more important during credibility assessments. Additionally, posts in Reddit are longer than in Twitter, which could account for the higher text credibility score for Reddit. Although the metadata scores are similar in Figure 5 for both platforms. Twitter does provide a richer set of metadata (e.g. classifications, hashtags, retweet counts etc.) on their page layouts, and this is evident from the increased number of clicks on metadata components in the heatmap (Figure 4(b)).

### B. `Exp2`: *Artificially Controlled Content*

From the previous experiment, we selected a set of features on which to base our experimental evaluation for `Exp2`. Once again, our research question and hypothesis was refined based on information from the previous studies. By artificially controlling values for each target feature, we can assess the



(a) News subreddit page in Reddit (b) A search result page with 'news' keyword in Twitter

**Fig. 4:** Heatmap visualization of the user annotations for `Exp1`, where clicks on an entity indicate perceived credibility.

| Type | Feature | Treatment 1 | Treatment 2 |
|---|---|---|---|
| Visual | Embedded image | Present | Not present |
| Visual | Profile image(person) | Present | Not present |
| Visual | Profile image(logo) | Professional | Unprofessional |
| Network | # of friends | 95th percentile(7,524) | 5th percentile(8) |
| Network | Classification | News | Non-news |
| Network | # of comments | 95th percentile(565) | 5th percentile(5) |
| Network | # of shares | 95th percentile(933) | 5th percentile(0) |
| Network | Posting time | 95th percentile | 5th percentile |
| Content | Sentiment degree | No sentiment | High sentiment(95%) |
| Content | Sentiment polarity | Negative value(-0.95) | Positive value(+0.95) |
| Content | Tags | Tags present | No tags |
| Content | Links | Links present | No links |

**TABLE III:** List of features and their independent variables in the second user experiment. (For each treatment, posts from both outlets, NYTimes and The Onion, are presented to the participants.)

directional effect of metadata content on credibility perception. Furthermore, to avoid topic-specific biases in assessing the stability of feature influence on credibility perception across topics, we incorporated a variety of common topics (e.g.: World, Health, Politics, Entertainment) into the evaluation.

**RQ5:** How do different treatments of metadata variables influence credibility perception?

HYPOTHESIS. By applying artificially controlled values for metadata from the 5th and 95th percentiles of sampled real world data, we will observe differences in perceived credibility of the associated information.

**RQ6:** Is the influence of displayed metadata on credibility perception consistent across different topics?

HYPOTHESIS. *Feature influence varies across topics.*

*h) Study design:* In this experiment, we aim to test hypotheses 1 through 6 by artificially controlling metadata values for each of 12 salient factors identified in the previous experiment, and eliciting credibility assessments from participants. To achieve this, we construct 12 different formulated lists of microblog postings, controlling one independent variable on each list to capture how much impact that individual factor (e.g: profile image, link, number of friends) has on perceived credibility of information. Figure 6 shows a screen shot of the interface used in the study. The treatment in this case is a default profile image, which is the only controlled variable in this example. A five point Likert scale for feedback on perceived credibility is shown beneath the blog post.

A larger user experiment (N=646) was deployed on MTurk to evaluate the effects of artificially controlled treatments of each feature. In order to determine in which direction each factor impacts on the perceived credibility, we designed the study with two treatments and one baseline for each feature. *Treatment 1* uses 'feature present' in case of binary features and 95th percentiles (high values) in case of numeric features. Correspondingly, *Treatment 2* exhibits 'feature absent' in case of binary features and 5th percentiles (low values) for numeric features. The exception to this choice of percentiles are our two "sentiment" related features. Since we assume that low sentiment indicates more credibility, in view of the fact that objectivity is linearly correlated with credibility, we mapped the 5th percentiles to Treatment 1 and 95th percentile to Treatment 2 here. The complete list of features tested in this experiment (`Exp 2`) are shown in Table III. The leftmost column shows a categorization of each feature into one of three classes:

- *Visual* This is the set of highly visual factors in the microblog, including profile pictures, attached images, and photos.
- *Network* This is the set of network-based factors, including static features such as number of friends or followers, and dynamic/conversational features such as retweets, votes or mentions.
- *Content* This is the class of features solely based on text, including sentiment terms, hashtags, and links.

Table III describes our selection of treatments for each target feature. While we cannot exhaustively evaluate all values for a particular feature, our aim was to construct a reasonably diverse set of values for each, based in some cases on analysis of real world distributions (e.g., for numeric attributes such as number of friends), and on manual selection for others (e.g: profile image content). The two rightmost columns give examples of controlled metadata treatments for each variable in our study. For instance, in the case of profile images, one treatment was selected from a popular satirical website. A second treatment (showing an image of President Obama giving a speech), was selected from a major US news outlet. The main goal of Table III is to show that differences can be produced in perceived credibility compared to a baseline, by manipulating a single visual feature, and our results show that this is indeed the case. For data with a range, such as number of friends, scores were taken based on large-scale distribution

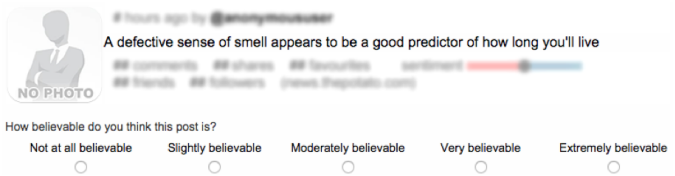



**Fig. 6:** Screenshot showing part of the microblog interface from Expt 2.

analysis for each feature, as described in the Random Sampling section below. The majority of examples were straightforward to construct, for example, network features such as number of friends or retweets were represented as low or high values. Recency was evaluated by grouping messages by posting time and displaying 5th or 95th percentile values of the resulting distribution. Sentiment features were more difficult however: for polarity and sentiment degree, a negative value (negative sentiment used) and low value (little sentiment used) were assigned respectively. These were indicated using "score bars", shown in the example column of Table III. For content/text features such as tags and links, a binary value (present or not present) was used. A baseline condition was also used. The baseline consisted of raw text with no associated metadata. Participants were asked to answer the same type of general questions about their demographic information and experience with microblogs as in `Exp1`. To avoid topic-specific biases and to explore our earlier hypothesis, 5 topics (World, Health, Politics, Entertainment, Business) and 10 posts from both *New York Times* and *The Onion* accounts on Twitter were manually collected. We purposely sampled two outlets that represent two different aspects of online journalism (objectivity and satire) in order to reflect real-world contexts.

*i) Random sampling:* To estimate reasonable values for the treatments listed in Table III, 233,037 messages and user profiles were crawled using the Twitter Streaming API and Reddit API. A distribution analysis of feature values in this data allowed us to find reasonable thresholds (and extremes) to select values for our experimental treatments. From this dataset, we extracted 5th and 95th percentiles from the feature distributions.

## IV. Results and Discussion

In this section, we provide the findings from our two main user experiments. This section provides an overview of participant statistics for both studies, and following that, is organized around the six research questions and hypotheses posed earlier.

### A. Study Participants

`Exp1` had 102 participants. The average interaction time for both Twitter and Reddit users was 5 minutes. Users annotated three items each for a total of 306 annotations. 646 users participated in `Exp2`. Most reported that they were active daily on Twitter and had been active for more than a year. Most used the the official application, on a combination of mobile and desktop platforms. Most users did NOT guess that content was sourced from either the New York Times or the Onion. Facebook and Twitter was the most common response for the two source platforms. Participants spent an average of 9 minutes completing the survey and were paid $0.40. 55% were male and 45% female. They ranged in age between 18 and 60, with the majority between 18 and 29.

### B. Influence of Metadata

**RQ1:** Does the display of metadata in microblogs influence perceived credibility of the associated content?

Hypothesis. *Metadata display (textual or visual) does influence perceived credibility of microblog content. The direction of influence is dependent on the specific content displayed.*

All three experiments produced results that reveal an impact of metadata on perceived credibility in microblogs. The subjective results in Survey 1 show a strong (but self-reported) indication that Content of a message and Origin (author) of a message are the strongest influencing factors. This is followed by visual features including design of the UI and visual components such as profile pictures and other metadata. The discussions that follow here further illuminate and reinforce this basic result.

### C. Cross-Feature Analysis

**RQ2:** Do different features influence credibility by different amounts?

Hypothesis. *Features have varying degree of influence over credibility perception*

Figure 7 shows a list of the 12 evaluated features ranked according to the observed difference in influence between the treatments from Table III. The distribution clearly supports our initial hypothesis on RQ2. From this list, the profile picture/logo is the most influential factor on credibility perception, showing a significant increase over other features in terms of the difference in credibility rating between the treatments (ANOVA $p = 7.79e-9$ and $p < 0.05$ in Tukey post-hoc tests). Factors reinforced by the underlying network (Number of Friends –static, and Number of Shares –dynamic) are next-most influential. On the opposite end of the scale, sentiment polarity and posting time were the only two features where the lower value treatment achieved a better score than the higher value treatment. It seems that users in this study were not too concerned with recency. The more sentiment a message contained, the more likely it was to be deemed not credible.

Figure 7 shows the individual differences between the treatment 1 (white, top bar), baseline (dark gray, middle bar) and treatment 2 (light gray, bottom bar) for each of the 12 features. Overall, it is clear that the treatment 1 had a much stronger influence compared to the baseline (no controlled metadata) for almost all features. Effects were significant between the treatment 1 and the baseline in most cases but not significant between the treatment 2 and the baseline.

### D. Influence of Feature Classes

**RQ3:** What are the effects of different classes of microblog features on perceived credibility?

Hypothesis. *Visual factors will have the most influence, followed by network and content-based (text) factors.*

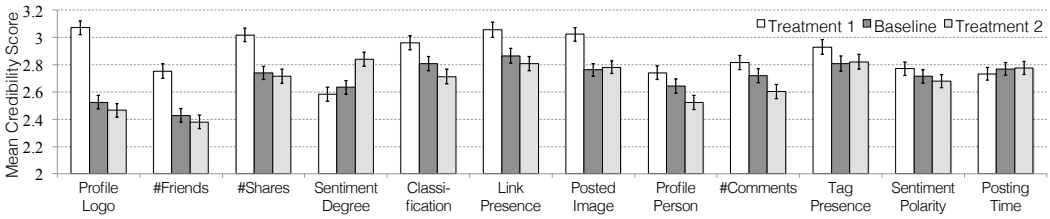



**Fig. 7:** Mean credibility scores across different features for the three treatments including baseline. Groups are sorted from left to right by the difference between the treatment 1 and treatment 2 scores.

Many existing credibility models perform feature classification to arrive at a credibility prediction for a given message. To evaluate our simple classification of (Network, Visual and Content)-based factors, we examined the overall group-wide credibility ranking and the results are shown in Figure 8 (a). Our initial hypothesis was that visual factors would be most influential, and it appears from the left column in Figure 8 (a) that this is in fact the case, at least for the high value treatments of the features in the group. No significant effect was shown for the lower value feature differences. The visual features produced a 10% increase over the baseline, compared with 7% for the Network and 3% for the Content group. This result further supports the notion that manipulating visual components can have a large effect in terms of how strangers perceive a microblog profile.

### *E. Cross-Platform Analysis*

**RQ4:** Can our models of feature influence be ported successfully to different microblogs?

HYPOTHESIS. *Influence of features is consistent across platforms*

To recap, `Exp1` evaluated self-reported importance of microblog features across two platforms: Reddit and Twitter, allowing users to place simple clicks in context of the actual interface. Figure 5(a) and (b) show the results for Reddit and Twitter respectively. This result disproves our initial hypothesis that feature ratings are invariant, since the text features achieved a higher credibility rating in Reddit. This is likely to be related to the fact that there is significantly more text per post allowed in Reddit. Image features appeared to garner similar ratings across the two platforms, which is a significant finding especially coupled with the fact that the Visual/Image-based features are the strongest influences on credibility perception.

### *F. Impact of Different Treatments*

**RQ5:** How do high and low values of metadata contents influence perceived credibility?

HYPOTHESIS. *L*ow values for metadata have a stronger effect than high values

Figure 8(d) shows a small (7%) improvement across all features and topics for treatment 1 over treatment 2. There is also a significant improvement shown for treatment 1 over the baseline (6%), Specifically, a single factor ANOVA test over the treatments shows the result is statistically significant ($F\ value$:87.54, $Pr(> F) < 2e-16$).

### *G. Cross-Topic Analysis*

**RQ6:** Is the influence of displayed metadata on credibility perception consistent across different topics within a domain?

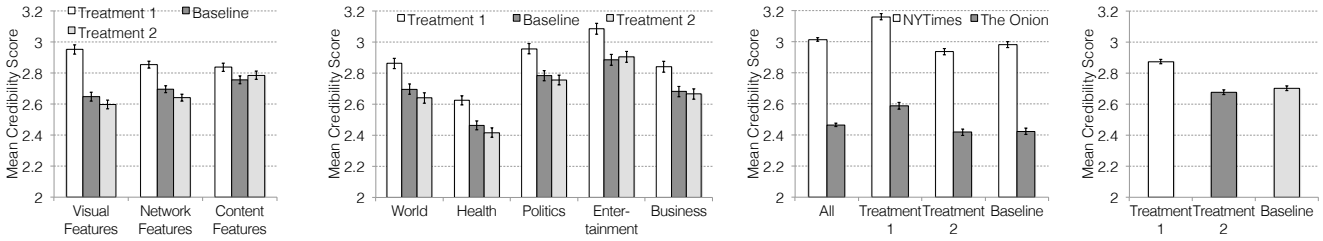
HYPOTHESIS. *Influence of the treatments varies across topics.*

To determine whether or not credibility ratings for our feature sets vary across different topics, and also to avoid introducing topic-specific biases in our other evaluations, we examined a collection of 5 diverse topics (World, Health, Politics, Entertainment and Business). This approach could be relevant for a system such as "CatStream" [23] which classifies Twitter feeds automatically based on topic interests. Messages from two well-known websites were sampled for each topic on a specific date in September 2014. The two websites were the New York Times (News) and The Onion (Satire/Comedy). Figure 8 (b) shows the results of an analysis of our different treatments across each of the five topics individually. From the results, it is immediately clear that the recurring trend of treatment 1 having a stronger effect than treatment 2 is invariant across all topics. ANOVA tests for differences showed $p < 2e-16$. Another interesting result is that the overall ratings for the health-related topic was significantly lower than all other topics, across all treatments. This might be a reflection of participants' cautious behavior when approaching a serious topic such as health. Conversely, the Entertainment topic produced the highest credibility ratings. Perhaps it is easier to appear credible in a domain that has far fewer constraints.

In addition to the cross-topic analysis, we also examined the overall ratings of articles from "New York Times (NYT)" and "The Onion" across every rating context, broken down by the three treatment conditions. Figure 8 (c) shows the results of this analysis. There is a significant improvement of 25% (ANOVA, $p < 2e-16$) in credibility rating between the two sources, with NYT at the upper end. Over all features, treatment 1 of metadata produces an equivalent increase in both NYT and Onion ratings, keeping the difference at approximately 25%. When comparing treatment 2 to the baseline treatment, we did not observe any significant change.

## V. CONCLUSIONS AND FUTURE WORK

To conclude, in this set of three interlinked studies we evaluated ways in which individual components of microblogs can influence end-user perceptions of information credibility. In particular, an initial survey provided general insight into a candidate set of factors that influence credibility most; a second study (N=102) examined these factors in the context of two microblog domains, Reddit and Twitter, to allow real




(a) Groups (Feature types) (b) Topics (c) Information Sources (d) All

**Fig. 8:** Mean credibility scores across different treatments

users to communally identify the most influential factors. A set of 12 factors were evaluated in detail in a third study that artificially controlled values for each feature and assessed credibility opinions from 646 participants in a crowdsourced evaluation.

Six hypotheses related to the impact of different microblog elements and treatments on human-provided assessments of credibility were tested and the results were discussed in detail. Key findings from the study show that 1) metadata from the high end of observed real-world distributions (e.g. high number of friends, high number of hashtags) have a far stronger effect on credibility ratings than treatments from the lower end of the distribution. 2) Visual factors, in particular, display of a Profile Picture Logo had the most positive impact on reported credibility. 3) Participants in the study did not view recency of posts as an important factor in credibility assessment. 4) Factors that influence perceived credibility did not remain constant across platforms. Text-based features scored higher on Reddit while Visual features did remain constant.

In follow-up studies, the authors plan to further examine some of the findings from this work. For example, what are the exact reasons for 'negative' traits, such as low numbers of friends, comments, and shares, not exhibiting as much of an effect on the mean credibility score as the positive ones across most features and topics? Future work will also apply findings from this experiment to improve prediction models from real world data such as [8], [9], predicting newsworthiness, credibility, or actions such as retweets, upvotes or shares.

## VI. Acknowledgment

This work was partially supported by the U.S. Army Research Laboratory under Cooperative Agreement No. W911NF-09-2-0053; The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of ARL, NSF, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

## References

[1] A. Java, X. Song, T. Finin, and B. Tseng, "Why we twitter: Understanding microblogging usage and communities," in *Proceedings of the 9th WebKDD / 1st SNA-KDD Workshop on Web Mining and Social Network Analysis*. ACM, 2007.

[2] D. L. Lasorsa, S. C. Lewis, and A. E. Holton, "Normalizing twitter: Journalism practice in an emerging communication space," *Journalism Studies*, vol. 13, no. 1, pp. 19–36, 2012.

[3] L. Willnat and D. H. Weaver, "The american journalist in the digital age," School of Journalism, Indiana University, Tech. Rep., 2014.

[4] M. Mendoza, B. Poblete, and C. Castillo, "Twitter under crisis: Can we trust what we rt?" in *Proceedings of the first workshop on social media analytics*. ACM, 2010, pp. 71–79.

[5] A. Burns and B. Eltham, "Twitter free iran: An evaluation of twitter's role in public diplomacy and information operations in iran's 2009 election crisis," 2009.

[6] J. Holcomb, J. Gottfried, and A. Mitchell. (2013) News use across social media platforms.

[7] S. Kumar, F. Morstatter, R. Zafarani, and H. Liu, "Whom should i follow?: identifying relevant users during crises," in *Proceedings of the 24th ACM conference on Hypertext and social media*. ACM, 2013, pp. 139–147.

[8] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter," in *Proceedings of the 20th international conference on World wide web*. ACM, 2011, pp. 675–684.

[9] B. Kang, J. O'Donovan, and T. Höllerer, "Modeling topic specific credibility on twitter," in *Proceedings of the 2012 ACM international conference on Intelligent User Interfaces*. ACM, 2012, pp. 179–188.

[10] S. Sikdar, B. Kang, J. O', T. Höllerer, and S. Adalı, "Understanding information credibility on twitter." in *SocialCom*, 2013, pp. 19–24.

[11] M. Rios and J. Lin, "Visualizing the "pulse" of world cities on twitter," in *Proceedings of the Seventh International Conference on Weblogs and Social Media, Cambridge, MA, USA, July 8-11, 2013.*, 2013.

[12] J. O'Donovan, B. Kang, G. Meyer, T. Höllerer, and S. Adalı, "Credibility in context: An analysis of feature distributions in twitter," in *Privacy, Security, Risk and Trust (PASSAT), International Conference on Social Computing (SocialCom)*. IEEE, 2012, pp. 293–301.

[13] J. Yang, S. Counts, M. R. Morris, and A. Hoff, "Microblog credibility perceptions: comparing the usa and china," in *Proceedings of the conference on Computer supported cooperative work*. ACM, 2013.

[14] B. Fogg and H. Tseng, "The elements of computer credibility," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 1999, pp. 80–87.

[15] J. O'Donovan, B. Gretarsson, S. Bostandjiev, T. Höllerer, and B. Smyth, "A visual interface for social information filtering," in *Computational Science and Engineering, 2009. CSE'09. International Conference on*, vol. 4. IEEE, 2009, pp. 74–81.

[16] B. Fogg, "Prominence-interpretation theory: Explaining how people assess credibility online," in *CHI'03 extended abstracts on human factors in computing systems*. ACM, 2003, pp. 722–723.

[17] B. Fogg, C. Soohoo, D. R. Danielson, L. Marable, J. Stanford, and E. R. Tauber, "How do users evaluate the credibility of web sites?: a study with over 2,500 participants," in *Proceedings of the 2003 conference on Designing for user experiences*. ACM, 2003, pp. 1–15.

[18] M. Gladwell, *The Tipping Point: How Little Things Can Make a Big Difference*. Back Bay Books, Jan. 2002.

[19] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi, "Measuring user influence in twitter: The million follower fallacy," in *Proceedings of international AAAI Conference on Weblogs and Social*, 2010.

[20] M. R. Morris, S. Counts, A. Roseway, A. Hoff, and J. Schwarz, "Tweeting is believing?: understanding microblog credibility perceptions," in *Proceedings of the conference on Computer Supported Cooperative Work*. ACM, 2012, pp. 441–450.

[21] J. Mahmud, M. X. Zhou, N. Megiddo, J. Nichols, and C. Drews, "Recommending targeted strangers from whom to solicit information on social media," in *Proceedings of the 2013 International Conference*

*on Intelligent User Interfaces*, ser. IUI '13. New York, NY, USA: ACM, 2013, pp. 37–48.

[22] S. Goel, J. Hofman, and M. Sirer, "Who does what on the web: A large-scale study of browsing behavior," 2012.

[23] S. Garcia Esparza, M. P. O'Mahony, and B. Smyth, "Catstream: categorising tweets for user profiling and stream filtering," in *Proceedings of the 2013 international conference on Intelligent user interfaces*. ACM, 2013, pp. 25–36.