Correlation Analysis of Reader's Demographics and Tweet Credibility Perception

Shafiza Mohd Shariff[1,2], Mark Sanderson[1], and Xiuzhen Zhang[1]

[1] School of Computer Science and IT, RMIT University, Australia
[2] Malaysian Institute of IT, Universiti Kuala Lumpur, Malaysia

Abstract. When searching on Twitter, readers have to determine the credibility level of tweets on their own. Previous work has mostly studied how the text content of tweets influences credibility perception. In this paper, we study reader demographics and information credibility perception on Twitter. We find reader's educational background and geo-location have significant correlation with credibility perception. Further investigation reveals that combinations of demographic attributes correlating with credibility perception are insignificant. Despite differences in demographics, readers find features regarding topic keyword and the writing style of a tweet to be independently helpful in perceiving tweets' credibility. While previous studies reported the use of features independently, our result shows that readers use combination of features to help in making credibility perception of tweets.

1 Introduction

Tweets from reliable news sources and trusted authors via known social links are generally trustworthy. However, when Twitter readers *search* for tweets regarding a particular topic, the returned messages require readers to determine the credibility of tweet content. How do readers perceive credibility, and what features (available on Twitter) do they use to help them determine credibility? Since Twitter readers come from all over the world, do demographic attributes influence their credibility perception?

There are several pieces of research regarding the automated detection of tweet credibility using various features, especially for news tweets and rumours [3], [12], [7], [17]. However, these studies focus on building machine learned classifiers and not on the question of how readers perceive credibility. Other research that studies reader's credibility judgments were conducted on web blogs, Internet news media, and websites [24], [6], [5], [23]. Quantitative studies were conducted on limited groups of participants to identify particular factors that influenced readers' credibility judgments. Since these user studies focused on certain factors, the subjects for readers' credibility assessment were controlled and limited.

We have found that there is a gap in understanding Twitter readers and their credibility judgments of news tweets. We aim to understand the features readers use when judging, especially when tweets are from authors unfamiliar to them. Therefore in this study, we address the following research questions:

1. Do Twitter readers' demographic profiles correlate with their credibility perception of news tweets?
2. Do the tweet features readers use for their credibility perception correlate with reader's demographic profiles?

To answer the research questions, we design a user study of 1,510 tweets returned by 15 search topics, which are judged by 754 participants. The study explores the correlation between readers' demographic attributes, credibility judgments, and features used to judge tweet credibility. We will focus only on tweet content features as presented by the Twitter platform and available directly to readers.

2 Related work

A class of existing studies focus on tweet credibility prediction by supervised learning using tweet content and textual features, the tweet author's social network, and the source of retweets. The credibility of newsworthy tweets is determined by human annotators that are then used to predict the credibility of previously unseen tweets [3]. The tweet credibility model presented in [7] were used to rank the tweets by credibility. Both works used a current trending topics dataset. Other studies focused on the utility of individual features for automatically predicting credibility [17] and on the credibility verification of tweets for journalists based on the tweet authors' influence [19].

Another class of research has examined the features influencing readers' credibility perception of tweets. Examining only certain tweet features, Morris et al. [16] studied just under 300 readers from the US. The authors identified that a tweet written by authors with a topically related display name influenced reader credibility perception. Similar research was conducted [23], comparing readers from China and the US. People from different cultural background perceived the credibility of tweets differently in terms of what and how features were used. The differences in tweet credibility perception for different topics was also reported in [20]. The study found eight tweet-content features readers use when judging the credibility level of tweets.

Some research has considered credibility perception in media other than Twitter. In the work by [5], they discovered that different website credibility elements such as interface, expertise and security are influenced by users' demographic attributes. Another study found that the manipulation level of news photos influenced credibility perception of news media [6]. The study showed that people's demographics influenced the perception of media credibility.

A Taiwanese-based study of reader's credibility perception regarding news-related blogs found belief factors can predict user's perceived credibility [24]. They also found that reader's motivation in using news-related blogs as a news source influenced credibility perception. Demographic variables were also shown to affect credibility. In another study [11], demographic attributes are also found to correlate with visual features as information credibility factors for microblogs, especially by younger people.

3 Methodology

We describe the collection of credibility judgments and the techniques that we use to analyze the data.

3.1 Data collection

Since we are aiming for broad participation in our study, a crowdsourcing platform was used to recruit participants. The use of crowdsourcing for annotating tweet credibility can be found in prior works [20], [7], [3]. We designed a questionnaire on the Crowdflower[3] platform. We divided the questionnaire into two parts. We first part of the questionnaire regards the basic demographic questions: gender, age, and education level. The country information is supplied to us by CrowdFlower platform as it is part of the workers' information upon their registration on the platform. The workers are regarded as tweet readers in this paper.

The second part of the questionnaire regards perceptions of the credibility of news-related tweets. We compiled tweets from three news categories: breaking news, political news, and natural disaster news, the same categories used in past studies [16], [23]. Each news category consists five world news topics reported by news agencies including BBC, Reuters and CNN from 2011 until May 2014. We made sure the news topics were evenly divided between trending and not trending topics. Trends were determined from the trending list on Twitter and What the Trend[4]. The tweets were examined to ensure they were topic relevant tweets and unique (i.e. each tweet contained a different message about the particular topic).

Readers were shown tweets as they would be shown in a Twitter search result page, retrieved in response to a search topic. Workers were also shown the topic and a topic description. Without expanding the tweet to see any additional comments, the readers were asked to give their perception on the credibility level of the tweet. Four levels are listed: very credible, seem credible, not credible, and cannot decide [3], [7]. Upon judging, readers were asked to describe what feature/s of the tweet they use to make the judgment. We prompted readers with a list of features reported in previous research [3], [20] as well as encouraging them to describe other features in the free-text interface.

In the news tweet collection, two writing styles of tweets are included – a style expressing authors opinion or emotion towards the topic and another just reporting factual information. The writing styles were used after results from a pilot user study, which indicated that readers also find tweets expressing an author's feelings regarding a topic as credible.

To ensure the quality of answers by readers, the readers were required to answer a set of gold questions at a minimum 80% qualifying level before they were allowed to progress. The gold questions were standard awareness questions,

[3] http://www.crowdflower.com/

[4] http://whatthetrend.com/ - a HootSuite Media company that lists Twitter's trending topic and explain why it is trending.

e.g. determining whether a topic and a tweet message were about the same news topic. The gold questions were not counted as part of the user study. A number of pilot studies were run to determine the optimal number of tweet judgments readers were willing to make. Twelve judgments per reader was the figure chosen. The dataset ground truth is available at http://www.xiuzhenzhang.org/downloads/.

3.2 Statistical analysis method

The chi-square test of independence is used to establish if two categorical variables have significant correlation. The test calculates the difference between observed data counts and expected data counts. The cutoff acceptance for the relationship is based on the accepted probability value (p-value) of 0.05. The chi-square statistic test can be calculated as follows, where O_i and E_i are the observed value and expected value for cell i of the contingency table:

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} \tag{1}$$

In this study, in addition to correlation analysis regarding a single demographic attribute and credibility judgments, we also aim to analyze how combinations of demographic attributes correlate with credibility judgments. Therefore, multi-way chi-square tests are also performed. Let V_1, ..., and V_k be k binary variables, the contingency table to calculate the χ^2 for these k binary variables is $(V_1, \bar{V_1}) \times (V_2, \bar{V_2}) \times ... \times (V_k, \bar{V_k})$. For example, when there are three binary variables A, B and C, to find out if variable A and B are correlated with variable C, the χ^2-statistic would be $\chi^2(\text{ABC}) + \chi^2(\text{AB}\bar{\text{C}})$ [1]. Note that the chi-square statistic is upward-closed, this means that the χ^2 value of ABC would always be greater than the χ^2 value of AB. Therefore, if AB is correlated, adding in variable C, ABC must also be correlated. Refer to [1] for proof of the theorem.

In our problem setting, we apply the theorem to prevent false discoveries for multi-way chi-square analysis. Assuming that A and B are independent variables for demographic attributes and C is the dependent variable for credibility levels. If A and B are correlated, even if A, B, and C are correlated, we would not be able to tell if the association between credibility level and the demographic attributes is due to an actual effect or to the non-independence of observations.

We first apply chi-square analysis between individual demographic attributes and the credibility judgments. If the result is insignificant, multi-way correlation analysis for combination of demographic attributes will be applied. To this end, the correlation for pairwise demographic attributes is first analyzed. If the attributes are significantly correlated, we will not continue the χ^2 test between the pair and credibility judgments. We similarly analyze the correlation between demographic attributes and features readers use for credibility judgments.

We also measure which cell in the contingency table influences the χ^2 value. The interest or dependence of a cell (c) is defined as I(c)= Oc/Ec. The further away the value is from 1, the higher influence it has on the χ^2 value. Positive dependence is when the interest value is greater than 1, and a negative dependence is those lower than 1 [1].

3.3 Slicing reader demographics

In this study the demographic data collected from the readers are used for chi-square analysis, refer to Table 1. The readers' demographic data, except for gender, are also categorized in binary and categorical setting based on other research [5], [6] to examine any correlation of demographic attributes or combinations of demographic attributes with tweet credibility perception. The different ways of partitioning demographic data are as follows:

- Age: Binary {Young adult ($\leqslant$ 39 years old), Older adult ($\geqslant$ 40 years old)} and Categorical {Boomers (51-69 years old), Gen X (36-50 years old), Gen Y(21-35 years old), Gen Z (6-20 years old)} [14]
- Education: Binary {Below university level, University level} and Categorical {School level, Some college, Undergraduate, Postgraduate}
- Location: Binary {Eastern hemisphere, Western hemisphere} and Categorical {Asia-Pacific, Americas, Europe, Africa}

We conduct the correlation analysis for each single demographic attribute for all the different slicing with credibility judgments or features.

4 Results

A total of 10,571 credibility judgments for 1,510 news tweets were collected from the user study. Only 9,828 judgments from 819 crowdsource workers were accepted for this study because only those workers answered the demographic questions and completed all 12 judgments. For any credibility judgments that were found to not describe the features used to make the credibility judgment or gave nonsensical comments, all judgments of the reader were discarded. We also discarded judgments of two readers from Oceania continent and three readers that did not have any education background, due to their low values undermine the required minimal expected frequency to apply χ^2 analysis. We were left with a final dataset for analysis from 754 readers with 9,048 judgments.

4.1 Overall demographics

Our final collection of data includes readers from 76 countries with the highest number of participants coming from India (15%). We then group the countries into continents due to the countries' sparsity. Out of the 754 readers, the majority (69.0%, n=521) of readers were male, similar to prior work that uses crowdsource workers for user study [11]. Most of the readers were in the age group of 20-29 years old (43.4%, n=327). In regards to the readers' education background, the majority had a University degree (38.1%, n=287). Table 1 shows the readers demographic profiles.

Table 1: Demographic profiles distribution

Demographic	**Value**	**Frequency**	**%**
Gender	Male	521	69.2
	Female	233	30.8
Age	16-19 years old	58	7.7
	20-29 years old	327	43.4
	30-39 years old	243	32.2
	40-49 years old	89	11.8
	50 years and older	37	4.9
Education	High school	127	16.8
	Technical training	58	7.7
	Diploma	81	10.7
	Professional certification	50	6.6
	Bachelor's degree	287	38.1
	Master's degree	137	18.2
	Doctorate degree	14	1.9
Location	Asia	275	36.5
	Europe	247	32.8
	South America	130	17.2
	North America	65	8.6
	Africa	37	4.9

4.2 Features

The features reported by readers are features of the tweet message itself, content-based and source-based. For features reported in free text, we applied a summative content analysis based on the list of features identified beforehand [9]. Table 2 (column 2) lists the features reported by readers when making their credibility judgments. Since the features are sparse, it is difficult to analyze their influence in the readers' credibility judgment. Therefore, we categorize the features into five categories and will be using the feature categories in all of our analysis related to the features:

- **Author**: features regarding the person who posted a tweet, including the Twitter ID, display name, and the avatar image;
- **Transmission**: features in a tweet message for broadcasting the messages on Twitter;
- **Auxiliary**: auxiliary information external to the textual message, including URL links, pictures, or videos;
- **Topic**: words and phrases indicating the search topic or news type, including search keywords and alert phrases such as "breaking news";
- **Style**: writing style of a tweet, including language style as well as message style as expressing opinion or stating facts.

4.3 Findings

We report our findings based on the research questions.

Table 2: Features reported by readers to judge credibility for news tweets

Category	Feature	Description
Author	Tweet author	Twitter ID or display name e.g. `Sydneynewsnow`
Transmission	User mention	Other Twitter user's Twitter ID mentioned in the tweet starting with the @ symbol e.g. `@thestormreports`
	Hashtag	The # symbol used to categorise keywords in a tweet e.g. `#Pray4Boston`
	Retweet	Contain the letters RT (retweet) in the tweet and the retweet count
Auxiliary	Link	Link to outside source - URLs, URL shortener
	Media	Picture or video from other sources embedded within the tweet
Topic	Alert phrase	Phrase that indicate new or information update regarding a news topic - e.g. `Update`
	Topic keyword	The search keyword regarding a news topic e.g. `Hurricane Sandy`
Style	Language	The language construction of the tweet (formal or informal English)
	Author's opinion	Tweet that conveys the author's emotion or feeling towards the news topic
	Fact	Factual information on the tweet regarding the news topic

RQ1: Do Twitter reader's demographic profiles contribute to the credibility perception of news tweets?

The correlation analysis for individual demographic attributes for each data setting (as described in subsection 3.3): Original (O), Binary (B), Categorical (C), and the credibility perceptions is shown in Table 3. At the original data setting, Education and Location are significantly correlated with credibility judgment, $\chi^2 = 49.43$,p<0.05 and $\chi^2 = 80.79$,p<0.05. Only Location is significantly correlated at all levels of partitioning. A post hoc analysis on the interest value of cells in the contingency table $Education \times Credibility$ for the original data found the cell that contributes most to the χ^2 value is readers with a 'Professional certification', who commonly gave 'not credible' judgments. In regards to the contingency table $Location \times Credibility$, we found there was a correlation between the readers from the African continent and the 'cannot decide' credibility perception in the original and the categorical data setting with a positive dependence. Both cells interest values are far from 1, indicating strong dependence. In the contingency table for $Location \times Credibility$ in the binary data setting, the interest value in each cell is close to 1, therefore there are no strong dependence.

We then conduct multi-way correlation analysis between combinations of demographic attributes and credibility judgments. Since Location is significantly correlated at all data levels, due to the upward closeness of χ^2 statistics (Section 3.2), we will not analyze combinations including Location. The correlation result for the rest demographic attribute pairs is shown in Table 4. In analyzing the

Table 3: Demographic profiles and credibility perception chi-square results

Demographic	**Data setting**	**Credibility**
Gender	Original	1.51
	Binary	1.51
	Categorical	1.51
Age	Original	4.87
	Binary	4.68
	Categorical	9.84
Education	Original	**49.43**
	Binary	4.78
	Categorical	12.29
Location	Original	**80.79**
	Binary	**39.62**
	Categorical	**80.33**

combination of demographic attributes, Bonferroni corrections of the *p-values* ($p<0.003$) are applied. Table 4(b) shows that only for the binary setting the (Age, Education) pair is not significantly correlated. Therefore, we further analyze the correlation of the (Age, Education) pair with credibility judgments. The correlation analysis outcome for $Age \times Education \times Credibility$ is $\chi^2 = 3.70, p>0.003$, accepting the null hypothesis. The result indicates that the joint independent demographic attributes of Age and Education in the binary setting do not correlate with the credibility judgments.

Table 4: Chi-square result for demographic attribute pairwise correlation

(a) (Age, Gender) & (Education, Gender)

		Gender
Age	O	**107.71**
	B	**77.40**
	C	**82.18**
Education	O	**105.89**
	B	**48.67**
	C	**61.80**

(b) Age, Education

	Education		
Age	O	B	C
O	**1791.23**	**763.96**	**1579.96**
B	**105.89**	2.18	**47.96**
C	**1732.96**	**749.53**	**1549.49**

RQ2: Do the tweet features readers use for their credibility perception of tweets correlate with reader's demographic profiles?

To answer this research question, we analyze the correlation between reader demographic attributes and the features readers reported for credibility judgments. From Table 5, all demographic attributes are significantly correlated with credibility perception features reported by readers. In the last column of Table 5, for the analysis of demographic attributes and the Transmission feature, as over 20% of expected values of the contingency table have expected value of less than 5, Fisher's Exact Test is used [15]. Table 5 is based on demographic data at the original setting, and similar results are obtained for data at binary and categorical settings. As all demographic attributes are correlated with credibility perception features, due to the upward closeness of chi-square statistics, any combination of demographic attributes is also correlated with credibility perception features.

Table 5: The chi square correlation between demographics and features used in credibility perception

Demographic	Feature Categories				
	Author	Topic	Style	Auxiliary	Transmission
Gender	0.01	**18.15**	**23.27**	1.59	0.59*
Age	**16.63**	**26.65**	**41.99**	8.65	1.00*
Education	11.12	**31.87**	**50.12**	**16.53**	**0.03***
Location	**46.87**	**83.81**	**67.35**	**13.60**	1.00*

*Calculated using Fisher's Exact Test

Topic and Style features have the most significant correlation with the demographic profiles while the Transmission feature has the least significant correlation with demographic attributes. *Age* and *Location* are significantly correlated with Author, and *Education* and *Location* are correlated with Auxiliary features. Meanwhile, only *Education* has significant correlation with Transmission. We are curious to know if there are combination of features readers reported to use when perceiving the credibility level of tweets. Using association mining to find the frequent combination of features [8], we found that Transmission, Author, and Auxiliary are frequently used with other features. Table 6 shows the frequent features that meet the support threshold of 1% or 90 times. The support threshold refers to the feature/s frequency of occurrence in the dataset. A low support threshold would help to eliminate uninteresting patterns [22].

5 Discussion

In regards to our first research question, readers' education background and their geo-location have significant correlation with credibility judgments. This finding is different from [24], [11], [6], as these studies do not find any correlation between tweet credibility perception and the education background. From our analysis, readers with a 'Professional certificate' and who judge tweets as 'not credible' are the ones that contribute to the significant χ^2 result. It is likely that education background may be connected with experience and thus such readers are more careful in making credibility judgments. Another possible reason may

Table 6: Frequent pattern mining of feature category

Frequent patterns	**Support (%)**
Topic	14.1
Style	12.7
Topic, Style	6.1
Auxiliary, Style	5.2
Auxiliary, Topic	4.7
Auxiliary, Topic, Style	4.6
Auxiliary,Topic,Style,Transmission	3.7
Auxiliary,Topic,Transmission	2.7
Author	2.7
Author,Auxiliary,Topic,Style,Transmission	2.6
Topic,Style,Transmission	2.5
Style,Transmission	2.0
Auxiliary,Style,Transmission	1.9
Author,Topic,Style	1.8
Author, Style	1.8
Topic,Transmission	1.6

be the absence or a low number of higher education level participants in past studies.

Although other researchers also found location correlated with credibility judgments, our dataset of international readership shows that readers from Africa have positive dependence with the 'cannot decide' credibility judgment. The political conflicts in countries on the Africa continent may have influenced the skeptical attitude towards media by the readers [4]. Therefore, tweets that readers find ambiguous resulted in their indecisive judgments on the tweet credibility judgements [18]. Other demographic attributes Age and Gender are not correlated with tweet credibility perception, which is a result similar to the work by [2]. Moreover the combination of Age and Gender does not have any significant correlation with tweet credibility perception either.

For the second research question, we find that all demographic attributes are significantly correlated with credibility perception features reported by readers. Especially the Topic features, including topic keyword and news alert phrase, and the tweet writing Style are important features used by readers for credibility perception. More than 26% of credibility judgments rely on Topic and Style features.

Features that are used in broadcasting tweets, the Auxiliary feature and Author feature, seems to be not considered by readers when judging the tweets' credibility level. Our results show a perspective different from that in [21], [10], [3], [13]. We also find that Auxiliary and Author features are mostly combined with other feature categories when readers make credibility judgements of news tweets, a result that was missing in other works since they are previously studied separately.

6 Conclusion

Although research on Twitter information credibility has been reported, most work focuses on automatic predicting or detecting tweet credibility. Our focus is on understanding Twitter readers and what influences their credibility judgments. In this study, we provided new insights in the correlation of reader demographic attributes with credibility judgments of tweets and the features readers used to make those judgments. Furthermore, the richness of data collected for this study – derived from a wide range of demographic profiles and readers across countries – is the first to offer insights on Twitter reader's direct perception of credibility and the features readers use for credibility judgements. For future work, we plan to examine if the type of news tweets has any influence on a reader's credibility perception. We would also like to investigate deeper on the features readers use, and the type of credibility level relates to those features and news type.

7 Acknowledgment

This research is partially supported by Universiti Kuala Lumpur (UniKL), Majlis Amanah Rakyat (MARA), and by the ARC Discovery Project DP140102655.

References

1. Brin, S., Motwani, R., Silverstein, C.: Beyond market baskets: Generalizing association rules to correlations. In: ACM SIGMOD Record. vol. 26:2, pp. 265–276. ACM (1997)
2. Cassidy, W.P.: Online news credibility: An examination of the perceptions of newspaper journalists. Journal of Computer-Mediated Communication 12(2), 478–498 (2007)
3. Castillo, C., Mendoza, M., Poblete, B.: Information credibility on twitter. In: WWW2011. pp. 675–684. ACM (2011)
4. Cozzens, M.D., Contractor, N.S.: The effect of conflicting information on media skepticism. Communication Research 14(4), 437–451 (1987)
5. Fogg, B., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P., et al.: What makes web sites credible?: a report on a large quantitative study. In: SIGCHI2001. pp. 61–68. ACM (2001)
6. Greer, J.D., Gosen, J.D.: How much is too much? assessing levels of digital alteration of factors in public perception of news media credibility. Visual Communication Quarterly 9(3), 4–13 (2002)
7. Gupta, A., Kumaraguru, P.: Credibility ranking of tweets during high impact events. In: PSOSM2012. pp. 2–8. ACM (2012)
8. Hahsler, M., Grün, B., Hornik, K.: Introduction to arules – mining association rules and frequent item sets. SIGKDD Explorations (2007)
9. Hsieh, H.F., Shannon, S.E.: Three approaches to qualitative content analysis. Qualitative health research 15(9), 1277–1288 (2005)
10. Hu, M., Liu, S., Wei, F., Wu, Y., Stasko, J., Ma, K.L.: Breaking news on twitter. In: SIGCHI2012. pp. 2751–2754. ACM (2012)

11. Kang, B., Höllerer, T., O'Donovan, J.: Believe it or not? analyzing information credibility in microblogs. In: ASONAM2015. pp. 611–616. ACM (2015)
12. Kang, B., O'Donovan, J., Höllerer, T.: Modeling topic specific credibility on twitter. In: IUI'12. pp. 179–188. ACM (2012)
13. Liu, Z.: Perceptions of credibility of scholarly information on the web. Information Processing & Management 40(6), 1027–1038 (2004)
14. McCrindle, M., Wolfinger, E.: The ABC of XYZ: Understanding the global generations. Sydney: University of New South Wales Press (2009)
15. McDonald, J.H.: Handbook of biological statistics, vol. 3. Sparky House Publishing Baltimore, MD (2014)
16. Morris, M.R., Counts, S., Roseway, A., Hoff, A., Schwarz, J.: Tweeting is believing?: understanding microblog credibility perceptions. In: Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work. pp. 441–450. ACM (2012)
17. O'Donovan, J., Kang, B., Meyer, G., Höllerer, T., Adalii, S.: Credibility in context: An analysis of feature distributions in twitter. In: Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom). pp. 293–301. IEEE (2012)
18. Rassin, E., Muris, P.: Indecisiveness and the interpretation of ambiguous situations. Personality and Individual Differences 39(7), 1285–1291 (2005)
19. Schifferes, S., Newman, N., Thurman, N., Corney, D., Göker, A., Martin, C.: Identifying and verifying news through social media: Developing a user-centred tool for professional journalists. Digital Journalism 2(3), 406–418 (2014)
20. Shariff, S.M., Zhang, X., Sanderson, M.: User perception of information credibility of news on twitter. In: Advances in Information Retrieval, pp. 513–518. Springer (2014)
21. Sundar, S.S.: Effect of source attribution on perception of online news stories. Journalism & Mass Communication Quarterly 75(1), 55–68 (1998)
22. Tan, P.N., Kumar, V.: Chapter 6. association analysis: Basic concepts and algorithms. Introduction to Data Mining. Addison-Wesley. (2005)
23. Yang, J., Counts, S., Morris, M.R., Hoff, A.: Microblog credibility perceptions: comparing the usa and china. In: Proceedings of the 2013 conference on Computer supported cooperative work. pp. 575–586. ACM (2013)
24. Yang, K.C.C.: Factors influencing internet users perceived credibility of news-related blogs in taiwan. Telematics and Informatics 24(2), 69–85 (2007)