6th International Conference on Computer and Knowledge Engineering (ICCKE 2016),
October 20-21 2016, Ferdowsi University of Mashhad

Quality of Claim Metrics in Social Sensing Systems: A case study on IranDeal

Pooria Taghizadeh Naderi[1], Hadi Tabatabaee Malazi[1], Mona Ghassemian[1], Hamed Haddadi[2]
[1] Faculty of Computer Science and Engineering, GC
Shahid Beheshti University, Tehran, Iran
Email: {po.taghizadeh, h_tabatabaee, m_ghassemian}@sbu.ac.ir
[2] School of Electronic Engineering and Computer Science
Queen Mary University of London, United Kingdom
Email: hamed.haddadi@qmul.ac.uk

***Abstract*—There is an ongoing trend in social sensing where people act as sensors and report the events happening in their surroundings. These claims are often reported by smartphones and need to be processed to discover new patterns of events. Since these claims are not generated with consistent quality, the processing and evaluation tasks can become a challenge. In this paper, we address questions on how the quality of each claim can be evaluated, and which factors should be considered to qualify the quality of the claims. To do this, we investigate the sources of low-quality claims an propose a new form of Quality of Claim (*QoC*) metrics. We categorize the Quality of Claim factors into two classes of Content Measure and Feedback Measure. The study is performed on Two datasets. The main dataset is the #IranDeal extracted from Twitter. To compare the quality metrics, a second dataset is crawled from the Fouresqure social network. The metrics follow the power law pattern and are modeled by a Zipfian distribution function. The results show the power degree varies from 1.75 to 5. A number of factors are discussed as an influencer of the variation, such as the query criteria of the extracted dataset, the characteristics of the *QoC* metric, and the type of the social network.**

***Index Terms*—Data Profiling, Social Sensing, Participatory Sensing, Crowdsourcing, Information Quality, Social Media, QoC, Big Data.**

I. Introduction

The combination of human capabilities with the embedded smartphone equipment forms smart sensor networks with greater potentials compared to the traditional ones [1]. Considering the numerous applications on smartphones, this new type of sensor networks [2], which are called social sensors, is one of the emerging sectors that attracts the researchers. Similar to traditional sensor network applications, social sensing is comprised of a number of sensors, but the contrasting point is the diversity of sensors ranging from electronic devices to human being. A group of people who can socially interact online and exchange their sensed information can be utilized for monitoring an environment or even detecting social events. Social networks, such as Twitter and Instagram, with huge amounts of claims and posts about events in the surrounding of the users [3], can play a significant role. The posts can be further utilized to detect events and communities [4]–[6] or be used to integrate contents of different social networks [7].

From an application point of view, a typical social sensing system is composed of a front-end and a back-end. The former consists of a networking platform such as social networks, human capabilities to discover a phenomenon in the surroundings, and smartphones as a mean of communication as well as their embedded sensors. It is responsible for gathering the user data. The latter has the duty of processing and analyzing the collected data and is responsible for prohibiting the invalid claims to be entered. Besides, it should assess the level of trust for the newly entered claims before processing.

Most of the social sensing systems are sensitive to invalid or inaccurate claims and reports. Gossiping, sensor inaccuracy, and user inaccuracy are some of the sources of claim uncertainty. To protect from these deficiencies, the system should analyze the trust level of claims to the invalid ones. In this paper, we address a research gap on the quality of claim in social sensing applications which are based on social networks. We introduce two categories of metrics. The first one is the *Content Measure* that includes metrics such as *content diversity*, *user tagging*, the *number of used keywords*, the *number of used hashtags*, and *geotags*. The second category is the *Feedback Measure* that takes the number of provoked reactions into an account. To analyze the metrics we use two datasets extracted from Twitter and Foursquare. The criteria to extract the first dataset is a hashtag, and the second one is crawled based on users. The main reason to choose these datasets is to analyze the effect of the datasets on the defined *QoC* metrics. The proposed metrics are modeled by curve fitting to the power law function to analyze the results.

The rest of this paper is organized as follows: Section II investigates the source of invalid, uncertain, or inaccurate claims, and reviews several existing validation assessment approaches. In Section III, new evaluation metrics to assess the *QoC*s are introduced. In Section IV a number of datasets are extracted and analyzed from Twitter and Foursquare social networks to further investigate our discussed metrics. Finally, Section V concludes the paper and suggests further works.

II. Uncertainty and invalidity in claims

The main task of the social sensing systems is to discover the interesting phenomenon happening in the monitoring environment based on the reported claims. One of the challenging parts is to find out whether something happens or not, based on

978-1-5090-3586-1/16/$31.00 ©

the reporting users with unknown levels of trust. Considering that a user may report inaccurate or even false claims, and in the absence of the ground truth, the system should be capable of discovering the truth and identify the trust level of individuals to maintain the integrity of claims. In this section, the sources of invalid claims are introduced. Then a number of well-known solutions are reviewed.

A. Sources of claim uncertainty and invalidity

Generally, the sources of inaccurate and uncertain claims are either the observer or the sensing devices.

1) Gossip: Social networks are powerful tools to disseminate news and events, but some of the propagated news are not true and take the form of a gossip. Besides, some of the users may willingly/ unwillingly help spread the gossip.

Zhao et al. in [8] worked on detecting rumors in social media. They use the regular expressions to separate rumors from other contents. The phrases are used to create the regular expressions "Is this true?", "Really?", and "What?". Some suggested regular expressions are "is (that | this | it) true" or "wh[a]*t[?!][?1]*". These regular expressions are used for filtering inquiries and corrections. The work is evaluated by a number of Twitter datasets.

2) Spam: Spamming is one of the prevalent challenging issues in systems in which the Internet is their source of input. In web-based systems, various technologies like CAPTCHA are utilized to differentiate among spamming robots and human. In social networks, spams can be detected by analyzing the inputs such as tags, links, tips and comments. However, this would be more difficult when a real user attempts to spam. Consequently, the large scale information validation process of a social sensing system is challenging.

The research presented in [9] introduces six features that can be used to detect spams and spammers. The features are TagSpam, TagBlur, DomFp, and NumAds. TagSpam parameter shows the probability of the tag is spam. For instance, if 100 users use tag A in their spam posts and 150 users use tag B in their spam posts, then the TagSpam of post A will be smaller than B. TagBlur is a parameter that shows the number of unrelated tags. Besides, spammers usually put many tags under every post and many of them are unrelated. DomFp is used to estimate spams base on the content structure. Some spams are generated by software and have the same structure. NumAds presents the number of ads in pages since most of the spam contents redirect users to external ads web pages.

3) Inaccuracy of users: People are the core element of the social sensing system and one of the main weak points of the system is that it is affected by human errors. As a result, any submitted claim cannot be fully trusted. The error margin of every individual is related to the physical conditions of the person and his/her commitment towards the contributed data. Sometimes, elderly individuals, who are less comfortable with smartphones, are more prone to submitting erroneous data.

B. Claim validation assessment

One of the important challenges of social sensing systems is to identify valid reports and claims from invalid ones. The following works are some of the well-known approaches.

Let C be the set of claims and c_i be a sample claim. The set U defines the users and accordingly, the UU matrix depicts the relationships between users. The value for each element of the matrix shows the type of relationship such as a follower, a following, and a friend. For instance, if the user j in Twitter is a follower of the user k, then the value of UU_{jk} will be *"follower"*. The matrix is used to find source dependencies and their weight. By further processing, a new matrix CU can be built, which shows the relation between claims and users. As a result, $CU_{ij} = m$ states that the user j reports the value m for the claim i. For instance, in imdb.com users may report a rate between 1 to 10 for movies or series (claim).

The issue of finding the truth was introduced on websites before emerging social sensing systems. Several solutions are proposed for the problem such as Sums [10], Average Log [11], Investment [11], Pooled Investment [11], and TruthFinder [12]. These methods are applied to find out the truth on the web. In social sensing systems, the quality of results directly depends on the quality of gathered information. The main challenges are to decrease the false reports and missing values.

To apply the aforementioned solutions for the context of social sensing systems, it must be considered that the conditions of the web and social sensing systems can be different. For instance, in the case of the *Sum algorithm* [10] the link between sites is used which is similar to the relations between users of social networks in social sensing systems.

To assess the validation of claims, various approaches are introduced so far, ranging from machine learning and natural language processing techniques to statistical ones. Some of the notable works are preventing spam input [13], data mining methods [14], clustering methods [15], labeling methods [16], and statistical algorithms [11]. Given the diverse set of social sensing applications and conditions, each case may have different requirements to maintain the quality of claim. The proper method for each application type should be selected based on the application requirements.

Bayesian method is among the well-known statistical methods for claim validating. The approach determines the integrity and the validity of the contents based on the degree of user confidence. The research presented in [17] is one of the notable approaches that uses this method. The approach is devised for independent claims and uses the network information and the topology of graph data to estimate the reliability of information. The approach gets the sets C and S as inputs. The former is the set of claims such as website news or comments and tweets in social media and the latter is their sources (which can be websites users in social media). Moreover, it uses the probability of *true* and *false* for every claim and produces the reliability of each user. The performance of the approach is compared with [11], based on a randomly generated dataset. The selected approach does not support source relationships, such as gossiping or information sharing. It is also not suitable for cases where claims are dependent, for instance, a user views the restaurant rating before vote.

In [18] Expectation-Maximization (EM) algorithm is introduced. It exploits a user vector and a claim vector. Given the existing incomplete vector values, the algorithm attempts to convey the values toward a complete one. Each element of the user vector represents the confidence degree of the user ($u_i \in [0,1]$), where 1 is the highest confidence. Similarly, the claim vector shows the validity of each claim ($c_i \in [0,1]$). The value of the most valid claims is 1. EM algorithm consists of two stages. During the first stage, new vector values are chosen using statistical formulas. In the next stage, elements are chosen to achieve more accurate results iteratively. The algorithm iterates over the two stages until the estimated parameters converge to a specified value. In each iteration, the outputs are applied as the inputs of the next iteration.

The issue of data conflict in EM algorithm is discussed in [19]. The users are able to enter non-binary values for a claim; thereby, various values can be considered for a claim. Since each claim may refer to unique phenomena with a specific value, a conflict happens when different values are submitted by users and as a result the valid value has to be identified. Furthermore, it is possible to reduce the degree of confidence for the users responsible for submitting invalid values. The authors introduce a modified version of the EM algorithm in [19]. One of the other contrasting points of the work is that in [18] binary values are used for claims, whereas in [19] there is no limitation on possible values of claims. Among the challenges in this section is how to pace calculation operations. Since the number of users and claims can be quite large theses algorithms for identifying valid data must be capable of a faster performance.

III. Quality of Claim metrics

The quality of claim is one of the key players in the success of social sensing applications. The quality is measured by a number of factors. According to the type of the front end, the *QoC* factors can be grouped into two categories. The first one deals with social network based systems, and the second one relates to mobile application based ones. The focus of this paper is the social network based ones. We introduce two classes of quality measure.

A. Content Measure

The richness of the claim contents facilitates the back-end applications. The following factors influence this measure.

- **Content diversity**: The diversity of the type of information that users of social networks can share between friends and connections. For instance, in Twitter users can use text to share their opinion and observations, or in Instagram the users can use short videos or photos to share their observations and messages for their opinions.
- **User tagging**: In a number of social network types, users can be mentioned and notified by each other by tagging their names. It provides new information about the importance of the claim, which can be used for further analysis. This mentioning can be analyzed to find debates between users. For instance, In Instagram, Facebook and Twitter everyone can use @ sign before the username to tag them.
- **Quantity of used keywords**: In the absence of natural language processing tools, the keywords are one of the main means to gain insight about the meaning of the reported claims in social networks. It is worth mentioning that the set of keywords is dependent on the subject and needs a prior knowledge. The set can be extracted by pre-processing the claims and extract frequently used terms, which are related to the main topic of interest. The keywords are further exploited in clustering, classifying, and analyzing the frequent patterns of the collected claims. The higher number of used keywords will increase the value of the claims in the analysis phase.
- **Quantity of used hashtags**: Hashtags are one of the main approaches to query the posted claims over a specific period of time. Analyzing hashtags are usually easier than the keywords since the keywords, which are in the form of a phrase, are difficult to extract and process. One of the shortcomings of analyzing hashtags instead of keywords is the relatively low quantity of hashtags.
- **Geo-tagging**: Many social networks, such as Facebook, Instagram, and Twitter, provide a Geo tagging feature for the users. It is used to pin the locations of the users who report a claim. The information is valuable in location base analysis to cluster the reporting user.

B. Feedback (Popularity) Measure

Each claim published on a social network may provoke reactions. Some reactions arise from the users judgment, which is valuable in truth discovery and sentiment analysis applications. The second type of reaction is in the form of redistributing the claim, mostly used in social event detection applications as well as truth discovery.

- **Opinion reaction**: When a user shares his information other users can express their opinion about it by hitting like, dislike, or giving a rate to the post. This parameter can help validate the information by unknown users. For instance, in Facebook, the users may convey their judgment by hitting the ***like*** button, or in Instagram by the red heart. In some of the systems, users may rate by giving stars. The users may also provide comments on the post, which can be further analyzed the asses the feedbacks. On Facebook, a user can reply to other user's comments or comment on users' pages and posts. In Twitter, users are able to reply to other users tweets and in Instagram users can leave their comments on another friend's claim.
- **Redistribution**: In a number of social networks, users are able to share the claims with their own friends. The number of reclaims shows the popularity of the claim and can be used in validating the claims and also in event detection applications.

Table I illustrates the *QoC* features support in a number of the social networks.

Social networks	Content Measure				Feedback Measure	
	Content Diversity	keywords Tagging	Geo Tagging	User Tagging	Opinion Reaction	Redistribution
Facebook.com	Text, Video, Image	√	√	√	Like - Reply	Share
Twitter.com	Text, Image	√	√	√	Favorite - Reply	Re-tweet
Foursquare.com	Text, Image	√	√	√	Like, Rate	-
Instagram.com	Text, Image, Video	√	√	√	Like	-
plus.google.com	Text, Video, Image	√	√	√	Like	Share
Linkedin.com	Text, Video, Image	√	√	√	Like	Share
about.me	Text, Image	-	√	-	Like - Reply	-
pinterest.com	Image	-	√	-	Like-Reply	-
thumbler.com	Text, Video, Image	√	-	-	Like - Reply	Share

TABLE I
QoC FEATURES SUPPORT IN SOCIAL NETWORKS.

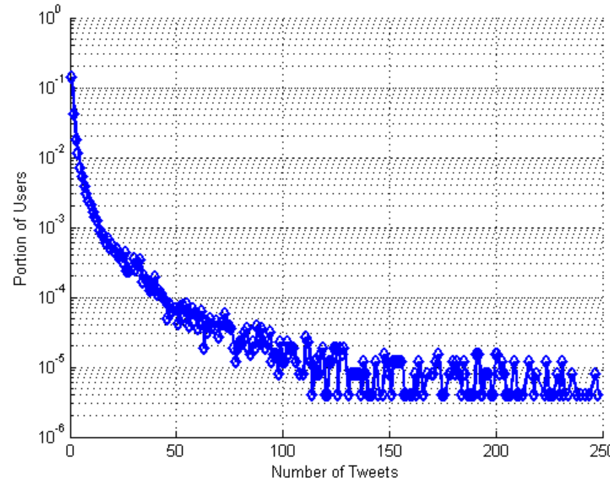

Fig. 1. The distribution of the number of *tweets* per user in the Twitter dataset.

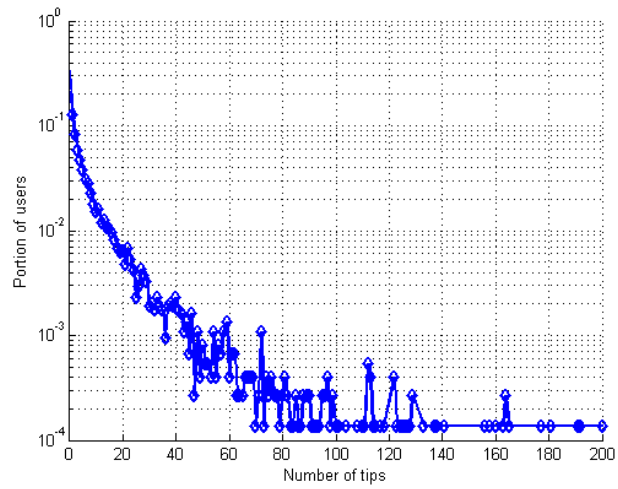

Fig. 2. The distribution of the number of *tips* per user in the Foursquare dataset.

IV. Evaluation and analysis

Two hashtag-centric and user-centric datasets are gathered by the crawler for the evaluation.

The first dataset is extracted from the Twitter based on **IranDeal** hashtag which was in the news headlines in 2015. Accordingly, 260,000 tweets are crawled that belongs to 66,238 users.

The second dataset is extracted from the Foursquare social network. The dataset is comprised of 7,402 users collected based on the recursive friend's lists of a random user and their tips are gathered. Therefore, the dataset is crawled based on users. The refined dataset is populated with 4,903 users, who post at least one tip for a restaurant. The total number of extracted tips is 40,741 for 35,503 restaurants.

A. Quantity of comments per user

User trust level in active users, who may post comments more often, can be calculated easier compared to the passive ones. The users are grouped according to the number of reported claims (tweets or tips). The population of each group indicates the number of users who posted an exactly certain number of tips. The populations of groups depend on the total number of users being analyzed. Therefore, we calculate the portion of users in each group of the total number of users. For instance, if the population of a group of users who posts only 10 tips is 100 and there are 10,000 users in the dataset, the portion of users who posted exactly 10 tips will be 0.01.

In the Twitter dataset, we investigate the distribution of the number of tweets per user. Although the dataset is hashtag based, the metric behaves similarly (Figure 1). About 14% of the users (36663 users) post exactly 1 tweet, but only 4% have two posts. The percentage decreases as the number of tweets increases. The numbers of posts for the groups of 100 and more, oscillate between 0 to 5 users. The maximum number of tweets belongs to a user with 2328 posts. One of the interesting features that can be learned from Figure 1 is, whenever the sparsity of the value decreases (the range of 0 to 50 on the horizontal axis) fewer fluctuations is observed. For instance, in the range of 100 to 250 in X axis, many vibrations are recognized.

The distribution of the quantity of comments per user for the Foursquare dataset is depicted in the Figure 2. The horizontal axis represents the number of tips, and the vertical axis shows the fraction of users who post the specific quantity of tips. The figure shows the quantity of users who posts zero to 200 tips. The values for more than 200 tips are sparse and based on the dataset only 23 users post more than 200 tips. The highest number of tips belongs to a user with 1809 tips. The results show a sharp decline by increasing the number of comments. 934 users post only 1 tip, and in the case of 20 tips, the number of users shrinks to 47. The results also demonstrate that 17 users post only 40 tips and only 1 user posts 200 tips. The

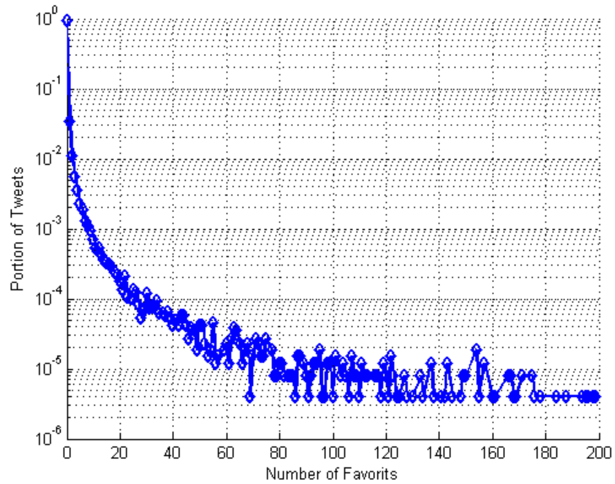

Fig. 3. The distribution of *favorites* per *tweet* in the Twitter dataset.

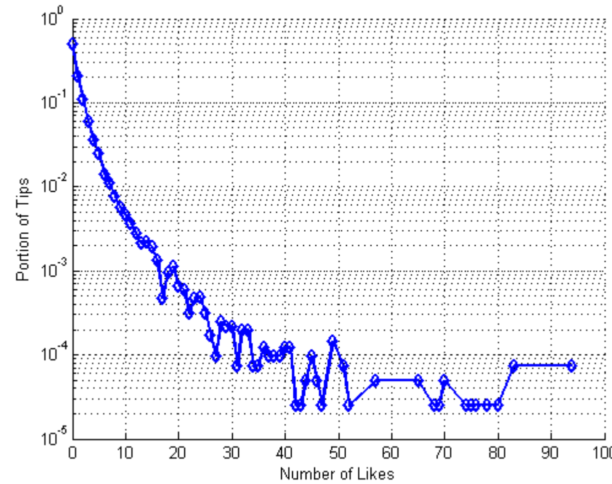

Fig. 4. The distribution of *likes* per *tip* in the Foursquare dataset.

number of users does not decline smoothly, Figure 2 shows a number of fluctuations in 46 tips and 112 tips. The figure shows the power law behavior.

In order to model the metric a power law distribution function is applied. We used the Zipf law of Equation 1 and apply curve fitting to map the results and the outcome is shown in Table II. The sum of squares due to error (SSE) is used as a metric to evaluate the goodness of the fit. One of the important characteristics of Zipf law is the value of S which shows the degree of curve slope. The higher values for s shows a sharper decline in the vertical axis.

$$f(x) = \frac{a}{(x+b)^s} \tag{1}$$

The calculated values of s for Figures 2 and in 1 are 1.751 and 2.715 respectively.

B. Popularity of comments

The number of likes for each comment shows its popularity. To analyze this metric, the comments are categorized based on their number of likes. The population of each category is divided by the total number of comments. The outcome demonstrates the fraction of comments, which may get a certain amount of likes. The approach is applied to both datasets to compare the results. The results for the Twitter dataset is shown in Figures 3, where the X axis represents the number of favorites (likes) and the Y axis shows the portion of the tweets (comments) who is received a certain number of favorites. In this analysis 260,000 tweets are used and the total number of favorites are 145,292. A large fraction of tweets (93%) does not get any favorites. One of the main reasons can be the topic of the dataset (#IranDeal). The portion of tweets that gets 1 and 2 favorites are 3.4% and 1.1% respectively, which shows a sharp decline. The results for more than 200 favorites become sparse and are not shown in this figure. The highest favorites obtained by tweet is 4367. Interestingly, the calculated value of the s (in Equation 1) is 1.775, that shows the users are keen to mark a post as favorite. In Figure 4 the X axis is the number of likes and the Y axis shows the portion of the tips (comments) which receives a certain number of likes. Since the fractions of popular tips are very small, the Y axis is demonstrated logarithmically. The figure depicts the results for the Foursquare dataset. According to the dataset, 40741 tips received 69352 likes. About 49% (20092) of tips do not receive any likes and 8427 tips receive only one like. The number of tips decreases significantly. The number of tips which receive exactly 15, 30, and 45 Likes are 79, 9, and 4 respectively. At the end of the spectrum, 3 tips receive 399 likes. The portions of tips which get more that 100 likes are very sparse; therefore Figure 4 demonstrates only the portions up to 100 likes. The curve fitting calculations for this experiment results in $s = 3.349$. It shows that a few users may post a tip that is interesting for many users.

Comparing the value of s for these datasets implies that the nature of the used social network affects the characteristics of the dataset. The Twitter is a popular social network to disseminate news and follow up events. The Foursquare social network mostly used to introduce perfect places to go with friends and is mostly dependent on the diverse range of preferences.

One of the other popularity metrics is the rate of sharing a comment. There is a general idea that expresses most of the metrics in social networks with the power law. The analysis results of this metric on the Twitter dataset shows different behavior. Figure 5 demonstrates the portion of re-tweets. It expresses the dependency between the *QoC* metrics and the way the dataset is crawled. The exploited Twitter dataset is based on a hashtag. As a result, people who follow the hashtag are eager to share the news headline with their friends. Moreover, the sparsity of the data for the values of higher than 500 also affects the results.

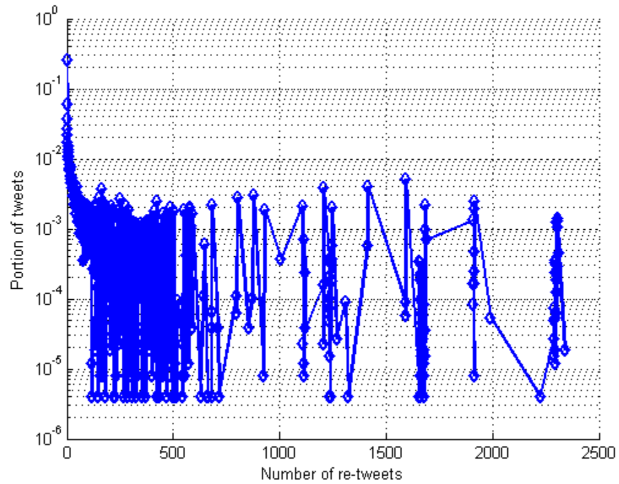

Fig. 5. The distribution of *re-tweets* per *tweet* in the Twitter dataset.

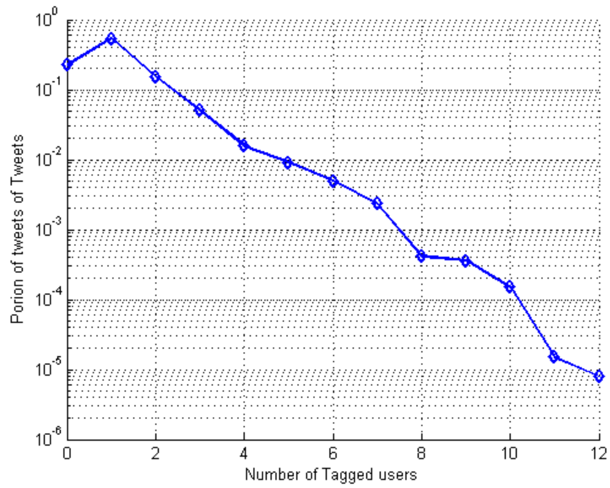

Fig. 6. The distribution of the number of tagged users in Tweets.

C. Number of tagged user per comment

A user can be tagged (mentioned) in a tweet. The tags provide extra information that boosts claims processing applications. Figure 6 illustrates the number of people tagged in tweets based on the second dataset. The tweets containing @ sign are processed to extract the histogram. The highest frequency belongs to the comments with a single tagged user (140191 tweets). The highest population of tagged users in a tweet is mentioned to be 12 people. Similar to the previous results, the number of tagged users in a tweet decreases exponentially as the number of users increases.

The calculated value of s (Equation 1) according to the curve fitting process is 4.666. The value shows that users are interested in tagging a few others in their tweets. About 22.2% of tweets do not contain any mentioned user, while almost 54% mentioned exactly one user. Around 15% of tweets tagged exactly two users and the values decrease in higher numbers.

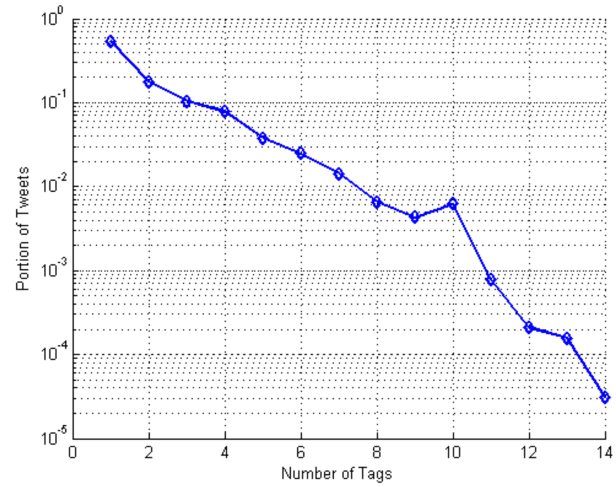

Fig. 7. The distribution of the number of hashtags used in Tweets.

TABLE II
THE ZIPF CURVE FITTING PARAMETERS FOR QOC METRICS

QoC metric distribution	Dataset	a	b	s	SSE
Comments per user	Foursquare	1.422	3.015	1.751	$6.739 * 10^{-5}$
Comments per user	Twitter	1	1.232	2.715	$1 * 10^{-5}$
Likes per comment	Foursquare	33.6	3.566	3.349	$4.636 * 10^{-5}$
Likes per comment	Twitter	1	3.015	1.774	$0.4407 * 10^{-5}$
Tagged users	Twitter	104.1	2.088	4.666	$10.52 * 10^{-5}$
Hashtags used	Twitter	7493	6.563	4.953	$35.81 * 10^{-5}$

D. Number of hashtags per tweet

The frequency of using hashtags is analyzed for the Twitter dataset. The number of hashtags used in each tweet is captured, and the fraction of tweets which use a certain number of hashtags are demonstrated in Figure 7. Considering that the dataset is crawled based on #IranDeal, all the tweets in the dataset have at least 1 hashtag. The number of tweets with exactly one hashtag is 141829 (54.5%). The portion of tweets with exactly 2 hashtags from 17.8% of the dataset, and as the number of hashtags increases, the percentage of tweets declines as well. According to the dataset, the maximum number of hashtags is 14 which used in 8 tweets. The value of s for this metric of the dataset is computed as 4.953. It is the highest value for the s in Equation 1.

V. CONCLUSION

This paper the causes of claims uncertainty are reviewed and a brief overview of the approaches that addresses the issue are presented. It also defines a new set of quality of claims metrics. To analyze the metrics, two different datasets are extracted from two well-known social networks. The analysis results show that most of the metrics follow the power law. But it is not a general rule. The results show that the slope of the curves obtained from the results vary. The type of the metric, the characteristics of the exploited social network, the query criteria in which the data is crawled, and the level of sparsity are some of the main influencers.

References

[1] H. T. Malazi, K. Zamanifar, and S. Dulman, "Fed: Fuzzy event detection model for wireless sensor networks," *Int. Journal of Wireless & Mobile Networks (IJWMN)*, vol. 3, no. 6, pp. 29–45, 2011.

[2] R. Mortier, H. Haddadi, T. Henderson, D. McAuley, and J. Crowcroft, "Human-data interaction: The human face of the data-driven society," *Available at SSRN 2508051*, 2014.

[3] Y. Mejova, H. Haddadi, A. Noulas, and I. Weber, "#foodporn: Obesity patterns in culinary interactions," in *Proceedings of the 5th International Conference on Digital Health 2015*. ACM, 2015, pp. 51–58.

[4] E. Moradi, M. Fazlali, and H. Tabatabaee Malazi, "Fast parallel community detection algorithm," in *Proceedings of the International Symposium on Computer Architecture and Digital Systems (CADS)*, 2015.

[5] Z. Masdarolomoor, R. Azmi, S. Aliakbary, and N. Riahi, "Finding community structure in complex networks using parallel approach," in *Embedded and Ubiquitous Computing (EUC), 2011 IFIP 9th International Conference on*. IEEE, 2011, pp. 474–479.

[6] Z. Masdarolomoor, S. Aliakbary, R. Azmi, and N. Riahi, "Distributed community detection in complex networks," in *Computational Intelligence, Communication Systems and Networks (CICSyN), 2011 Third International Conference on*. IEEE, 2011, pp. 281–286.

[7] M. Neshati, E. Asgari, D. Hiemstra, and H. Beigy, "A joint classification method to integrate scientific and social networks," in *Advances in Information Retrieval*. Springer, 2013, pp. 122–133.

[8] Z. Zhao, P. Resnick, and Q. Mei, "Enquiring minds: Early detection of rumors in social media from enquiry posts," in *Proceedings of the 24th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, 2015, pp. 1395–1405.

[9] B. Markines, C. Cattuto, and F. Menczer, "Social spam detection," in *Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web*. ACM, 2009, pp. 41–48.

[10] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM (JACM)*, vol. 46, no. 5, pp. 604–632, 1999.

[11] J. Pasternack and D. Roth, "Knowing what to believe (when you already know something)," in *Proceedings of the 23rd International Conference on Computational Linguistics*. Association for Computational Linguistics, 2010, pp. 877–885.

[12] X. Yin, J. Han, and P. S. Yu, "Truth discovery with multiple conflicting information providers on the web," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 20, no. 6, pp. 796–808, 2008.

[13] N. Jindal and B. Liu, "Review spam detection," in *Proceedings of the 16th International Conference on World Wide Web*. ACM, 2007, pp. 1189–1190.

[14] M. Jamali and M. Ester, "A matrix factorization technique with trust propagation for recommendation in social networks," in *Proceedings of the fourth ACM conference on Recommender systems*. ACM, 2010, pp. 135–142.

[15] Y. Sun, Y. Yu, and J. Han, "Ranking-based clustering of heterogeneous information networks with star network schema," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge discovery and data mining*. ACM, 2009, pp. 797–806.

[16] V. S. Sheng, F. Provost, and P. G. Ipeirotis, "Get another label? improving data quality and data mining using multiple, noisy labelers," in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge discovery and data mining*. ACM, 2008, pp. 614–622.

[17] D. Wang, T. Abdelzaher, H. Ahmadi, J. Pasternack, D. Roth, M. Gupta, J. Han, O. Fatemieh, H. Le, and C. C. Aggarwal, "On bayesian interpretation of fact-finding in information networks," in *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on*. IEEE, 2011, pp. 1–8.

[18] D. Wang, L. Kaplan, H. Le, and T. Abdelzaher, "On truth discovery in social sensing: A maximum likelihood estimation approach," in *Proceedings of the 11th International Conference on Information Processing in Sensor Networks*. ACM, 2012, pp. 233–244.

[19] D. Wang, L. Kaplan, and T. F. Abdelzaher, "Maximum likelihood analysis of conflicting observations in social sensing," *ACM Transactions on Sensor Networks (ToSN)*, vol. 10, no. 2, p. 30, 2014.