User Trust Graph: A Model to Measure Trustworthiness

Ali Arya[1]
arya@carleton.ca

Michael Hine[2]
mike.hine@carleton.ca

Amirsam Khataei[1]
amirsam.khataei@carleton.ca

[1]School of Information Technology, Carleton University, Ottawa, Canada
[2]Sprott School of Business, Carleton University, Ottawa, Canada

Abstract. In this paper, we address the problem of evaluating the trustworthiness of a user in different types of Twitter graphs. We discuss this within the context of a persuasive recommender system that aims at creating personalized content using social media and other personal information. Twitter has been established as an alternative type of information resource due to its simplicity and its enormous number of users who transmit diverse information in real time. There is evidence that people are becoming increasingly reliant on others' opinion through their social network accounts. Trustable opinion is influential in persuading individuals to purchase products or support policy makers. The domain varies from consumer market to news and politics. Evaluating trust between two people is a delicate subject. In Twitter, neither the social graph nor structured data such as total number of likes or common retweets are sufficient to measure the trustworthiness of a user within a given social group. It is important to consider the actual sentiment associated with the tweets shared between the two parties. Existing approaches only consider relationships among users based on structured data. In this paper, we introduce a new approach to calculate a trust score as a function of time and tweet sentiment.

Keywords: Trust, Persuasive, Twitter, Network, Personalization.

1 Introduction

Twitter offers a venue where people can express their feelings, share their opinion, and get reviews from other users. These activities result in the generation of a rich source of information on different aspects of life. Such user-generated information ranges from health and politics to product and service reviews. Still, "opinionative information" can be generated by different people in different format (image or text) with various relations to us. The opinion can be a picture that is shared by an iconic public figure, an organization who we strongly support or a credible colleague who we personally know. Alternatively, the opinion might be a tweet that is published by someone who is part of our social network but not trusted, or with different taste and standards.

In: R. Orji, M. Kaptein, J. Ham, K. Oyibo, J. Nwokeji (eds.): Proceedings of the Personalization in Persuasive Technology Workshop, Persuasive Technology 2018, Waterloo, Canada, 17-04-2018, published at http://ceur-ws.org

There is evidence that people are becoming increasingly reliant on others' opinion through their social network accounts like Twitter [1][2]. Their online peers are influencing their knowledge, opinions, and behaviors through the information stream and social dynamics within these sites [3][4][5][6]. Such influence can be essential in persuading people to perform an action, buy a product, or attend an activity. For example, a restaurant's promotional material can be more persuasive to an individual if they are customized to each viewer and incorporate photos from past festive events attended by the receiver or their friends, or positive quotes from trusted friends about that restaurant. We are more likely to support a cause or perform any action if our friends show support for it, but it depends on how much we trust those friends. With the rapid growth of user-generated content published on Twitter, a tool for mining said content to assess sentiments and opinions while evaluating the trustworthiness of the content on a large scale becomes essential. There have been many studies within the field of trust-based social recommender systems. Some preliminary literature demonstrates the advantages of applying factor of trust in recommendation marking [9] [10].

Previous work by the authors [7] has made initial inroads into this topic. Our initial goal was to demonstrate the value of aggregating trustable opinionative information with product reviews to persuade individuals by creating personalized content. Their objective was to introduce the essential components involved in collecting supporting opinions from reader's trustworthy sources to support the personalized content. Pang and Lee [8] and Ravi [1] published an inclusive survey of methods and approaches in the field of opinion mining. Though, without identifying the trustable source for the opinion, the personalized content may not become persuasive.

Due to the importance of product and service reviews to product vendors and policy-makers, there is extensive interest in this line of research [9][10][1]. The focus of the work referenced above is to design a system to enable organizations to analyze and aggregate their customers' attitudes towards a product or service. They consider different dimensions such as time, geographical location, and personal/professional experience. With years of research, we are now able to collect and aggregate opinionative information but measuring the trustworthiness of the opinion source has yet to be incorporated.

However, the above approaches were limited in their reliance on social graphs and it did not consider user trust. On Twitter, most users follow back their followers in accordance with mere formal courtesy. Only a few percent of users in follow relationships communicate with each other [6]. Therefore, an algorithm that is using only a Twitter social graph consisting of only following relationships is not sufficient.

Based on the above need, we propose the new User Trust Graph which consists of nodes (corresponding to user accounts and tweets), and edges (corresponding to follow and retweet relationships). Unlike the Twitter social graph, which is relatively static, the user trust graph is dynamic and reconstructed when a user mentioned the other user in a tweet, both users share a common retweet/hashtag, and a user likes the other user's tweet.

After discussing the technical issues of creating the user trust graph by using opinion-oriented information, we focus on the central problems of designing the actual model. In the next section, the related work and the existing gaps in research have been reviewed. Following that, we present some background information about our proposed model, a general overview of the trust calculation procedure, and some implementation notes. This is followed by some conceptual results and examples. Finally, we describe the primary application that will benefit from the model, how this model can be improved and what work remains to be done in the future phases of the research.

2 Related work

As discussed in the previous section, the most natural and intuitive graph to represent follow-followee relation is the social graph. This graph is extensively studied by Kawak [10]. He demonstrated social graph flaws as a mean to investigate the role of Twitter as a social or information network [11]. As previously discussed, this form of graph does not capture any reliable information on trust.

The idea of ranking tweets is introduced and explored in TwitterRank [12] and TURank [13]. These approaches measure the users' influence considering the link structure of follow relationships, the similarity between users, and the number of posts. The models are based on a false assumption that a user only retweets a tweet if it appears to contain useful information because he/she wants to share it with his/her followers. After all, people retweet sometimes to show their opposition and disagreement with a news or a content. In addition, both models do not reflect tweets' sentiment in their algorithms. Although this research is close to our research, measuring the influence is inadequate to be interpreted as measuring trust strength.

Srijith and colleagues [14] propose to assess the trustworthiness of tweets based on the analysis of the entire tweet ecosystem spanning across tweets, users, and the web. Although they are using tweet popularity as a factor to demonstrate the credibility of the tweet space, in the user space their model is limited to implicit links between the users based on the follower-followee relationships. On the other hand, as noted previously, ranking tweets based on popularity should not count as trustworthiness.

The Mention Graph was first designed to improve the identification of authoritative accounts, to discover active and dynamic communities, or to assign weights to the follow relationships [14]. It has also become an alternative to Social Graph. In the mention graph, each node represents a Twitter account. A directed edge between two nodes *a* and *b* exists if the account *a* mentions the account *b* in at least one tweet. The Mention Graph is also suitable to identify the authoritative users who are main original content publishers within a social group. However, being the content publisher should not be the sole indicator of trustworthiness.

Current research has not yet taken into proper consideration that people's relations are not just the matter of total number of retweets, likes, or followers. The role of feelings (sentiments) need to be considered more. The User Trust Graph is a step toward inclusion of sentiments in the subject of social graphs.

3 Proposed Model

We model interactions between two Twitter users by defining the User Trust Graph. In this graph, each node represents a user. There are two types of edges that exist in this graph: directed and undirected. A directed edge between two nodes *a* and *b* exists if user *a* mentions user *b* in at least one tweet or likes a tweet from user *b* or retweets at least one of user *b*'s tweets. When two users share a common hashtag or retweet, there is an undirected edge between two nodes. This is illustrated in Fig. 1.

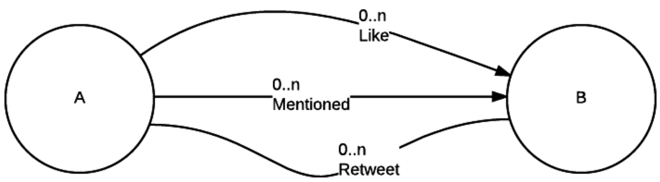

Fig. 1. Trust User graph overview

We perform sentiment analysis on the interaction between users and assign a positive, neutral or negative score. Despite the sentiment of a tweet's content, when a user retweets another person's tweet, the interaction is considered as a positive score. But a negative quote by the user on the same tweet would take precedence over the positive score. Intuitively, a greater positive score is interpreted as stronger trust strength between the two users. Besides, in contrast with the social graph which is static, this graph is reconstructed whenever a new activity such as a retweet, mentioned a user in a tweet or adding a like occurs.

Human relations may change over time and correspondingly the strength of trust varies between two people over time as well. As discussed by Dai and Davison [15], information should be ranked based on recency. Further, recent interaction with a positive sentiment is interpreted as a shared current interest between the users. To apply recency during weight calculation, edges (user's interactions) are labeled with the event timestamps. This allows us to filter edges based on temporal parameters to measure the trend within a given time frame.

Having obtained a user trust graph, we then try to calculate the weight of each interaction. Equation 1 is based on the idea of defining trust as a function of time. Here if the edge (interaction) from node *a* to node *b* exists with sentiment value *e*, the trust weight for the interaction is $w = e / \mu_t$ where μ_t is the weight coefficient for the given timestamp. The weight coefficient depends on two factors:

1. ΔTE_{ab}: The elapsed time from the interaction time (e_t) between user *a* and b, to the current time.

2. ΔTI_{ab}: The elapsed time from the first interaction time between user *a* and b, to the current time.

$$\mu_t = \Delta TE_{ab} / \Delta TI_{ab} \quad (1)$$

This leads us to Equation 2, where DT is the direct trust experience from *a* to *b* within a given domain and W is the sentiment (weight) associated with direct interaction between the users.

$$DT_{ab} = (W_{DL}, W_{DM}, W_{DR}) \quad (2)$$

$$DTW_{ab} = \sum w_{ab} / |W_{ab}| \qquad 0 < W_{ab} < 1 \quad (3)$$

In Equation 3, $W_{ab} = \{w_{ab1}, w_{ab2}, \ldots, w_{abn}\}$: a set of all form of sentiment score in the form of likes, retweets, and mentioned that user *a* gives user *b*'s content.

While *DT* is the direct trust strength, there are also common preferences and content in the form of retweets and hashtag exist between two users. This data would also be used as a second source of input to improve the trust strength prediction, though there is no direct interaction or no trust path from the user *a* to user *b*. This value is demonstrated as *CT.* As demonstrated in Equation 4, *DT* and *CT* become the foundation for Twitter Trust Weight (T_{ab}).

$$T_{ab} = \alpha DT_{ab} + (1 - \alpha) CT_{ab} \quad (4)$$

$$where\ CT_{ab} = (W_{CR}, W_{CH}) \quad (5)$$

$$CTW_{ab} = \sum w_{ab} / |W_{ab}| \qquad 0 < W_{ab} < 1 \quad (6)$$

The rationale behind introducing parameter α is to establish minimum criteria to have full confidence on scores calculated based on direct communication. To have full confidence in direct trust experience value, a certain number of direct interaction between two users is needed. As the number of direct interaction increases, the amount of confidence (reliability) increases until it reaches a point which signifies a close relationship between two users. On the other hand, if user *a* has a small number of direct interactions with user *b*, the value for DT_{ab} becomes unreliable to quantify the trust value. Thus, the combination mechanism (Equation 4) is proposed to reduce the weight a for *DT* and increases the importance of the common interaction trust value. CT_{ab} helps complement the uncertainty from the lack of direct feedback. To calculate the reliability of DT_{ab} we rely on Equation 7, which is proposed by Xu et al [16] and Sabater [17].

$$\alpha_{ab} = \begin{cases} \sin(\pi/_2 . \, ^{|W_{ab}|}/_{N_{min}} & |W_{ab}| \in [0, N_{min}] \\ 1 & Otherwise \end{cases} \tag{7}$$

Where N_{min}: the minimum number of sentiment score in the form of likes, retweets, and mentioned that user *a* gives on user *b*'s content.

4 Evaluation

We performed a pilot study of user trust graph using a Twitter collection that was built by monitoring the activity of 10 users who may also follow and interact with each other on Twitter as well. To simplify the sentiment analysis, we decided to only use tweets in the English language. Due to Twitter API limitation, we were restricted to collect up to a maximum of 500 tweets per user. For each user, the model returned a list of top ten trustworthy users. Users were asked to evaluate the generated result by making any adjustment to the list. We repeated the same experiment using TwitterRank [12] which correlates trust with tweets popularity. For both models, we applied Bobadilla formula [18], to calculate the error measurement of the predicted trust strength for each user. Fig. 2, shows the differences between the two approaches.

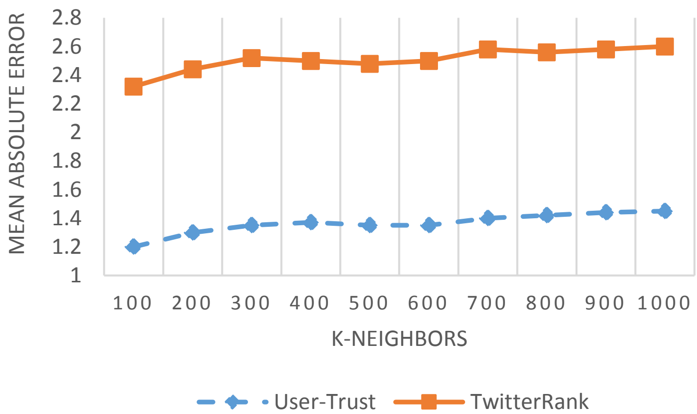

Fig. 2. Error measurement of the predicted trust strength

The proposed model has a lower error rate than TwitterRank. In some places, we noticed that our model's predicted value is 40% to 45% more accurate than TwitterRank. Although the results are very promising, the sample that we used to evaluate is relatively small and only adequate for a pilot study. Further analysis is required to examine the findings.

5 Application in Persuasive Content Generation

This research is the crucial part of a bigger project aimed at automatically creating persuasive content based on personalization [19]. While the detail description of this persuasive content generator is beyond the scope of this paper, here we briefly review its main concepts and the use of the proposed trust network.

Customization has been widely considered as a means of increasing persuasion and effectiveness through the process of preparing content for a specific person based on stated or implied preference [20][21][22], and it has resulted in breaking groups of people into micro-segments [23]. While a good starting point, such “segmentation” treats all members of a group similarly, ignoring personal differences. Fig. 3 shows a high-level overview of the proposed persuasive system which tries to achieve “personalization’ through three layers of incorporating personal characteristics.

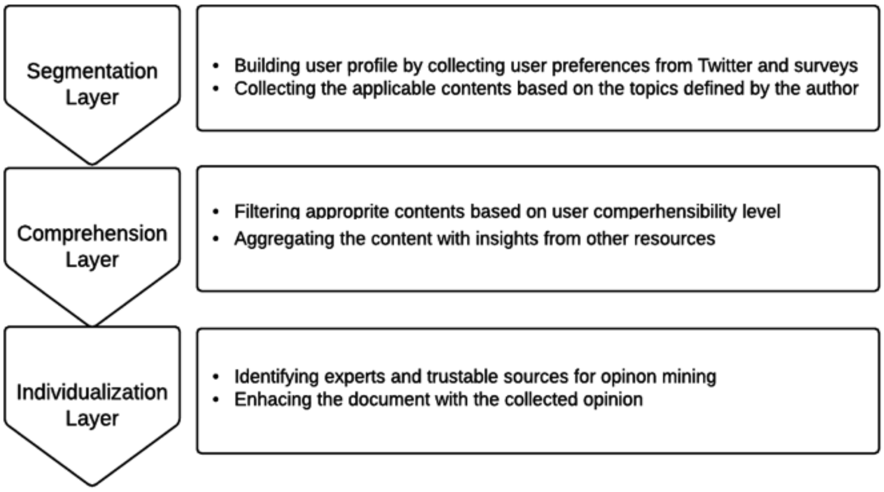

Fig. 3. Proposed persuasive system and its three layers of personalization

In the design of this system, we rely on the four persuasive factors defined by the Yale Attitude Change (YAC) model. According to YAC model [24], in order to maximize the chance to persuade the audience to take action or change their opinion, we should first gain their attention, adjust the comprehension level so the user can consume the message, make sure the argument is accepted, and, finally, ensure that the message is remembered.

YAC is our main inspiration for the design of the persuasive content model. As Fig. 3 shows, the proposed persuasive system consists of three layers: Segmentation, Comprehension and Individualization. First, the content is selected and adjusted in the segmentation layer based on each user’s interests. The process corresponds to YAC’s ‘gaining user attention’ factor. To avoid introducing an inappropriate level of complexity for the user, we tailor the selected information to match the reader’s readability score. This is an attempt to address YAC's 'comprehension criteria' as a part of the Comprehension layer. Finally, the third component is 'acceptance through trust.' It has been frequently demonstrated that highly trustworthy and credible communicators inspire a more positive attitude toward the position they advocate than do those with a lower level of credibility [25]. Thus, as part of the Individualization layer, readers' trustable sources (such as close friends or respected celebrities) are identified. These sources are later used to collect and encompass the main content with applicable and trustable opinions.

Our propose trust network is used in the Individualization layer, designed based on the idea of enriching the personalized content with the flavor of trustable personal data such as opinionative information that support the content.

6 Conclusion

In this paper, we measure trust strength between two Twitter users by defining the user trust graph through applying classification and sentiment analysis. It also elucidates that the data which would be extracted from Twitter Trust Weight is essential for persuaders like, recommender systems. For a recommendation to be effective, a consumer must find the opinion (tweets) trustworthy. Persuasive content generator [19] is the primary application that would benefit from user trust graph. These applications rely on opinionative information as trustworthy support to maximize the persuasion effectiveness. While further research is required to fine-tune all major parts of the model, the current design and findings are promising and show the potential use in many educational and otherwise informative applications, such as customer briefing, e-learning, etc. In addition to further research on technical aspect of the proposed system, more theoretical work is required to investigate the value of social network-based trust and who we can trust, the ethical issues associated with such persuasions (abuse of trust, privacy, etc.). We observed 40% improvement in accuracy of computation in our pilot study. Though, it is required to do a similar evaluation against other types of models with a bigger sample to validate our findings. As ideas for future research, it is also interesting to classify tweets to measure the strength of trust in a given domain. This will result in identifying experts for a given domain in a user trust graph

References

[1] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications," *Knowledge-Based Syst.*, vol. 89, pp. 14–46, 2015.

[2] M. Tsytsarau and T. Palpanas, "Survey on mining subjective data on the web," *Data Min. Knowl. Discov.*, vol. 24, no. 3, pp. 478–514, May 2012.

[3] L. Bode, "Political News in the News Feed: Learning Politics from Social Media," *Mass Commun. Soc.*, vol. 19, no. 1, pp. 24–48, Jan. 2016.

[4] R. M. Bond *et al.*, "A 61-million-person experiment in social influence and political mobilization," *Nature*, vol. 489, no. 7415, pp. 295–298, Sep. 2012.

[5] S. Messing and S. J. Westwood, "Selective Exposure in the Age of Social Media," *Communic. Res.*, vol. 41, no. 8, pp. 1042–1063, Dec. 2014.

[6] J. Turcotte, C. York, J. Irving, R. M. Scholl, and R. J. Pingree, “News Recommendations from Social Media Opinion Leaders: Effects on Media Trust and Information Seeking,” *J. Comput. Commun.*, vol. 20, no. 5, pp. 520–535, Sep. 2015.

[7] A. Khataei and A. Arya, “Trust-based Individualization for Persuasive Presentation Builder,” in *Human-Computer Interaction*, 2015, pp. 418–428.

[8] B. Pang and L. Lee, “Opinion Mining and Sentiment Analysis,” *Found. Trends Inf. Retr.*, vol. 2, no. 1–2, pp. 1–135, 2008.

[9] J. Liu, Y. Cao, C. Lin, Y. Huang, and M. Zhou, “Low-Quality Product Review Detection in Opinion Summarization,” in *Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 2007.

[10] A. Ghose and P. G. Ipeirotis, “Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics,” *IEEE Trans. Knowl. Data Eng.*, vol. 23, no. 10, pp. 1498–1512, 2011.

[11] H. Kwak, C. Lee, H. Park, and S. Moon, “What is Twitter, a social network or a news media?,” in *Proceedings of the 19th international conference on World wide web - WWW ’10*, 2010, p. 591.

[12] J. Weng, E.-P. Lim, J. Jiang, and Q. He, “TwitterRank: Finding Topic-sensitive Influential Twitterers,” in *WSDM*, 2010.

[13] Y. Yamaguchi, T. Takahashi, T. Amagasa, and H. Kitagawa, “TURank: twitter user ranking based on user-tweet graph analysis,” in *International conference on Web information systems engineering*, 2010, pp. 240–253.

[14] G. Amati, S. Angelini, and M. Bianchi, “Moving beyond the Twitter follow graph,” in *Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 2015.

[15] N. Dai and B. D. Davison, “Freshness matters: in flowers, food, and web authority,” in *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval - SIGIR ’10*, 2010, p. 114.

[16] M.-L. Germain and D. McGuire, “The Role of Swift Trust in Virtual Teams and Implications for Human Resource Development,” *Adv. Dev. Hum. Resour.*, vol. 16, no. 3, pp. 356–370, Aug. 2014.

[17] J. Sabater and C. Sierra, “REGRET: reputation in gregarious societies,” in *Proceedings of the fifth international conference on Autonomous agents - AGENTS ’01*, 2001, pp. 194–195.

[18] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez, “Recommender systems survey,” *Knowledge-Based Syst.*, vol. 46, pp. 109–132, Jul. 2013.

[19] A. Khataei and A. Arya, “Personalized Presentation Builder for Persuasive Communication,” *Commun. Des. Q. Rev.*, vol. 3, no. 3, pp. 25–32, 2015.

[20] A. Cöner, “Personalization and Customization in Financial Portals,” *J. Am. Acad. Bus.*, vol. 2, no. 2, pp. 498–504, 2003.

[21] M. Lou Roberts and D. Zahay, *Internet Marketing: Integrating Online and Offline Strategies: Mary Lou Roberts, Debra Zahay: 9781133625902: Amazon.com: Books*, 3rd ed. Cengage Learning, 2003.

[22] J. Vesanen, “What is personalization? A conceptual framework,” *Eur. J. Mark.*, vol. 41, no. 5/6, pp. 409–418, Jun. 2007.

[23] Tianyi Jiang and A. Tuzhilin, “Improving Personalization Solutions through Optimal Segmentation of Customer Bases,” *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 3, pp. 305–320, Mar. 2009.

[24] C. I. Hovland, I. L. Janis, and H. H. Kelley, “Communication and persuasion; psychological studies of opinion change.” Yale University Press, 1953.

[25] B. Sternthal, D. Ruby, and C. Leavit, “The Persuasive Effect of Source Credibility: Tests of Cognitive Response,” *J. Consum. Res.*, vol. 4, no. 4, pp. 252–260, 1978.