2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

Followee Management: Helping users follow the right users on Online Social Media

Anjali Verma*, Ashima Wadhwa*, Navya Singh*, Shivangi Beniwal*,
Rishabh Kaushal† and Ponnurangam Kumaraguru†
* *Department of Information Technology*
Indira Gandhi Delhi Technical University for Women, Delhi, India.
{anjaliverma471, ashima.wadhwa.1996, navya183, shivi17ben}@gmail.com
† *Precog Research Group*
Indraprastha Institute of Information Technology, Delhi, India.
{rishabhk, pk}@iiitd.ac.in

***Abstract*—User timelines in Online Social Media (OSM) remains filled with a significant amount of information received from followees. Given that content posted by followee is not under user's control, this information may not always be relevant. If there is large presence of not so relevant content, then a user may end up overlooking relevant content, which is undesirable. To address this issue, in the first part of our work, we propose suitable metrics to characterize the user-followee relationship. We find that most of the users choose their followees primarily due to the content that they post (*content-conscious* behavior, measured by *content similarity scores*). For a small number of followees, a high degree of social engagement (likes and shares) irrespective of the content posted by them is observed (*user-conscious* behavior, measured by *user affinity scores*). We evaluate our proposed approach on 26,516 followees across 100 random users on Twitter who have cumulatively posted 234,403 tweets. We find that on average for 60% of their followees, users exhibit very low degree of content similarity and social engagement. These findings motivate the second part of our work, where we develop a *Followee Management Nudge* (FMN) through a browser extension (plugin) that helps users remain more informed about their relationship with each of their followees. In particular, the FMN nudges a user with a list of followees with whom they have least (or never) engaged in the past and also exhibit very low similarity in terms of content, thereby helping a user to make an informed decision (say by unfollowing some of these followees). Results from a preliminary controlled lab study show that 62.5% of participants find the nudge to be quite useful.**

***Index Terms*—Online Social Media, Profile Management, Recommendation System, Nudge Design.**

I. Introduction

The usage of Online Social Media (OSM) in general and Twitter, in particular, has diversified over time with a number of people using their social media profiles for a variety of reasons [1] like staying connected with their friends, getting news update, brand advertisement and so on. To consume relevant content, users on Twitter follow other users (*followees*). There are numerous guidelines available on *whom to follow* on the web [2], [3] in the hope that content received from these followees on user's timeline is relevant. Twitter also takes care by ordering the content on user's timeline[1] based on the likelihood of relevance of the content for the user. Since the content posted by followees is not under user's control, despite all these guidelines and Twitter's ordering, the content displayed in timeline may not always be relevant. As the number of followees grows, the timeline gets filled with more such content which is not relevant, consequently, likelihood of relevant content getting overlooked increases. From user experience perspective, this would be undesirable and it would make them lose interest in the OSM platform itself. This would adversely affect monthly active users and thereby impact on the reputation of OSM platform. Users would like to follow only those people whose content would matter to them and they would want to unfollow the people whose content does not interest them.

In this work, we address this issue in two steps. In *first* step, we characterize users behavior with each of his/her followees. We hypothesize two behaviors in the context of user-followee relationship namely *user conscious* and *content-conscious* behaviors. User-followeee pair exhibit user conscious behavior when the user engages with the followee primarily due to social bonding irrespective of content. This engagement could be in terms of likes, shares and retweets of the content received by the followee. However, the content similarity between the posts of user and followee would be very low in user conscious behavior. User-followeee pair exhibit content conscious behavior when the user engages with the followee primarily due to the kind of content posted by the followee. There will be high degree of similarity in the tweets posted by user and followee in content conscious behavior. To quantify user conscious and content conscious behaviors, we propose two metrics namely *user affinity score (UAS)* and *content similarity score (CSS)*, respectively. User affinity score measures the closeness between the user and followee, exhibited through social engagements irrespective of content. It is summation of the number of times a user has retweeted, liked and mentioned each followee. Content similarity score finds similarity between content as well as interests of user and followee through word-level and topic-level similarities.

We implement our proposed metrics on 26,516 followees

[1]Twitter Timeline: https://help.twitter.com/en/using-twitter/twitter-timeline

IEEE/ACM ASONAM 2018, August 28-31, 2018, Barcelona, Spain

followed by 100 random users on Twitter and analyze over 234,403 tweets posted by these followees. We find that majority of users choose their followees primarily due to the content that they posted (content-conscious). For a small number of followees, a high degree of engagement (likes and shares) irrespective of the content posted by them is observed (user-conscious). Most interestingly, we find that for 60% of user-followee pair the user affinity scores and content similarity scores are almost close to zero, thereby suggesting very minimal social engagement and content similarity. This forms the motivation for the *second* step of our work in which we develop *Followee Management Nudge* (FMN) through browser extension (plugin) that helps users remain more informed about their relationship with each of the followees. In particular, the FMN nudges (as depicted in Fig 1) a given user with a list of followees with whom they have least (or never) engaged in the past and also exhibited very low similarity in terms of content, thereby helping user to make an informed decision (say by unfollowing the followees). We conduct a preliminary

Fig. 1. Followee Management Nudge (FMN) implemented through a browser plugin which displays the list of followees with whom user has least (or never) engaged in the past and the content similarity of tweets posted by user with those posted by these followees is also very less.

controlled lab study to evaluate our proposed nudge. 62.5% of participants find the nudge to be quite useful, it enabled them to unfollow followees posting irrelevant content, thereby, making their timelines more relevant.

II. Related Work

Our work comprises of two parts, *first*, in which we perform user-followee behavioral characterization and *second*, where we leverage this characterization study to develop a nudge implemented through browser extension.

For the first part of our work, we draw inspirations from the work related to user characterization [4] and tie strength [5] which quantifies the quality of relationship between a user pair on social media platform. Banks et. al. [6] were among the earliest to introduce interaction count based approach to determine relationship strength. Benevenuto et. al. [4] performed user characterization on social network based on click-stream data. Sousa et. al. [7] characterized Twitter reply network based on whether the tie is due to social reasons or due to the content (topical). Wagner et. al. [8] leveraged social media features to classify Twitter users on the basis of profession and personality.

For the second part of our work, we draw inspirations from the works of nudges ([9]–[11]) which help users take informed decisions. In particular, we found work of Wang et. al. [12] and Kaushal et. al. [13] quite interesting. Privacy nudges [12] for Facebook users were designed to prevent them from inadvertently posting privacy sensitive information. In the other work, nudge was developed [13] to help users control linkability between their identities on multiple social media platforms.

In the context of our problem of followee management, we found plenty of work ([14]–[16]) addressing the issue of *whom to follow*. However, once a user starts following a particular followee, whether the user has made the right decision or not, is an area which we felt is under explored. Our work, therefore, fills this gap by performing user-followee behavioral characterization and subsequently proposing followee management nudge implemented through browser plugin which helps user unfollow irrelevant followees.

III. Proposed Methodology

We explain the first step of our work on user-followee behavioral characterization in this section, Fig 2 describes our proposed methodology. We begin by *data collection* step in which random users are selected over Twitter along with their followees and the tweets posted by them. To characterize user-followee behavior, in the next step, we compute *user affinity score* and *content similarity score* for a user-followee pair, Table I explains the nomenclature used in defining these scores. Finally, we leverage these scores and develop browser extension (plugin) to nudge a given user with the names of followees whom the user can potentially *unfollow*.

TABLE I
Different Notations Used

Description	Notation
Twitter User whose followees are to be studied	U
k^{th} Followee of Twitter User	F_k
Number of Original Tweets posted by U	m
Number of Tweets from all followees which are engaged (liked or retweeted or mentioned) by U	n
Original Tweet posted by Twitter User U	OT
Tweet engaged (liked/retweeted/mentioned) by U	ET

A. User Conscious Behavior

By *user conscious* behavior, we refer to the association of a given user towards one (or more) of his/her followee irrespective of the content posted by followee. Such a behavior would exhibit high degree of engagement between user and

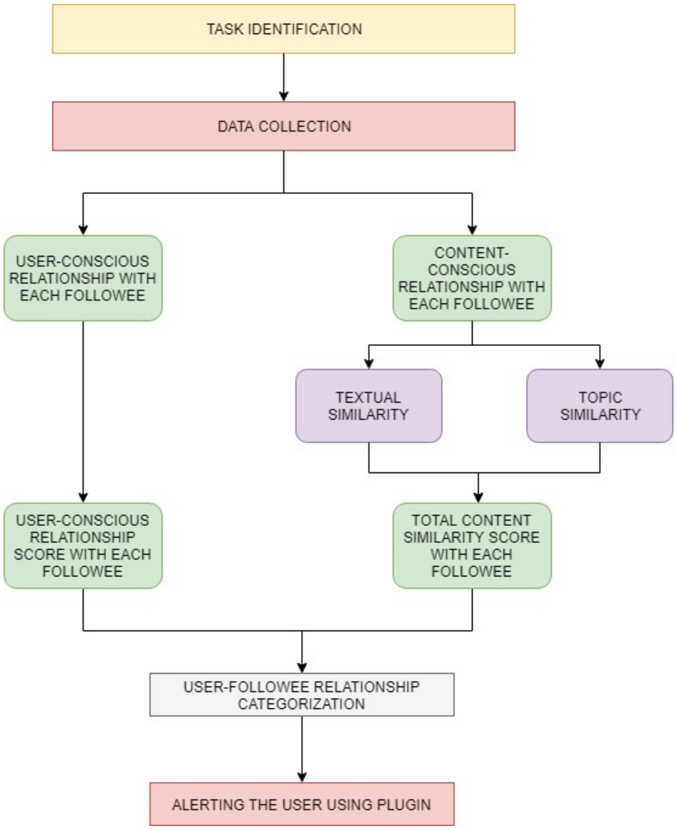

Fig. 2. Proposed Methodology. Starting with data collection, user affinity score and content similarity score are computed which are eventually used to characterize user-followee behavior.

followee. At the same time, content similarity between the tweets posted by user and followee would be extremely low. Engagement between user-followee on Twitter can be found out by computing the number of times user has liked, shared or mentioned tweet posted by his/her followee. To quantify user conscious behavior, we define *User Affinity Score (UAS)* between a user U and his/her k_{th} followee F_k as below.

$$UAS(U, F_k) = count(\sum_{i=1}^{n} ET_i^j(j = F_k)) \qquad (1)$$

where ET_i^j refers to i^{th} engaged tweet posted by j^{th} followee. Engagement can be either retweet (R) or like (L) or mention (M). $count$ means that we are interested in number of times user U has engaged (through retweet or like or mention) with his/her k_{th} followee F_k among the total n engaged tweets.

B. Content Conscious Behavior

By *content conscious* behavior, we refer to the association of user towards a followee owing to the content being posted. For content conscious user, the post made by the followee is relevant to the user only if it's content is similar to the content posted the user or user interest. To quantify content conscious behavior, we compute *Content Similarity Score (CSS)* between a user U and his/her k^{th} followee F_k. We do this by comparing each of the m original tweets posted by U with each of the n engaged tweets (liked or retweeted or mentioned) by U (which would have been posted by one of the followees). This comparison is done in two ways, one by using *word-to-word (or textual) similarity* and second by using *topic similarity*, therefore, *Content Similarity Score (CSS)* is given as below.

$$CCS = (\alpha \times T_xS(U, F_k)) + (\beta \times T_pS(U, F_k)) \qquad (2)$$

1) Textual Similarity: Number of similarity measures have been used in prior works to determine how textually similar two different strings are, such as, Cosine Similarity, Jaccard Coefficient, Pearson Correlation and so on. However, we consider Google's Word2Vec[2] which converts input text into word vectors as output. In our work, we used vectors pre-trained Google News dataset (comprising of about 100 billion words) in the Word2Vec model and evaluated the similarity between the tweets.

$$T_xS(U, F_k) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m}\sum_{j=1}^{m} w2v\_cmp(OT_j^U, ET_i^F(F = F_k)) \qquad (3)$$

where OT_j is the j^{th} original tweet content posted by user U and ET_i^F is the i^{th} engaged tweet posted by followee F engaged by user U.

2) Topic Similarity: We have used Latent Dirichlet Allocation[3] (LDA), one of the popular topic modeling technique to determine topics from the tweets of the user-followee pair to compute topic similarity between them.

$$T_pS(U, F_k) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m}\sum_{j=1}^{m} tp\_sim(TOT_j^U, TET_i^F(F = F_k)) \qquad (4)$$

where TOT_j is the j^{th} topic-list from original tweet posted by user U and TET_i^F is the i^{th} topic-list from engaged tweet posted by followee F engaged by user U.

C. User-Followee Relationship

On the basis of the user affinity scores and content similarity scores as explained in earlier sections, we categorize user-followee relationship into one of the four categories.

- User-Conscious Relationship: When the behavior of users towards their followees is due to personal relationship (social) measured quantitatively by user affinity score.
- Content-Conscious Relationship: When the behavior of the users towards their followees is driven only by the content similarity and topic similarity measured quantitatively by content similarity score.
- Mixed Behavior Relationship: When the behavior of the users towards their followees is driven both by personal relationship with the followees and content similarity as well.
- Inconclusive Relationship: When the behavior of the users towards their followees is driven neither by personal relationship with the followees nor by similarity of content.

[2]https://code.google.com/archive/p/word2vec/
[3]https://github.com/lda-project/lda

IV. Evaluation and Results

We begin by explain the data collection methodology followed by giving details of evaluation set-up and results.

A. Data Collection Methodology

Given that we were to perform user-followee characterization on Twitter, we started data collection by randomly picking 100 Twitter users. Table II explains the details of the data collected.

TABLE II
Data Collection for User-Followee Characterization

Description	Count
Number of Twitter Users for Characterization Study	100
Number of original Tweets from these Twitter Users	234,403
Total Number of Followees of these Twitter Users	26,516

Apart from above data, we needed more attributes to analyze the user-followee behavior. Using the Twitter API[4], for each of the 100 Twitter users, we collected the following:-

- Tweets from each of followee which were retweeted by the user.
- Tweets from each of followee which were liked by the user.
- Tweets from each of followee which were mentioned by the user.

B. Distribution of Scores

Using eq (1) and (2), we computed user affinity score and content affinity score for 26,516 user-followee pairs, respectively. Fig 3 gives distribution of these scores. It turns out that for over 90% of users, the user affinity scores are less than 0.1 whereas the same amount of users have less than 0.65 as their content similarity scores. Interestingly, it is evident from Fig 3 that almost 60% of user-followee relationships have almost zero user affinity and content similarity scores. This clearly suggests that for a large number of followees, users' relationship with them is *incomprehensible* which is reason enough to investigate further as to what kind of relationship do users want to establish with these followees. To this end, we develop a followee management nudge through browser plugin which nudges users about these followees and prompts them to take necessary action if required, one of which being, the act of unfollowing the followeee. More on this shall be discussed later in this paper.

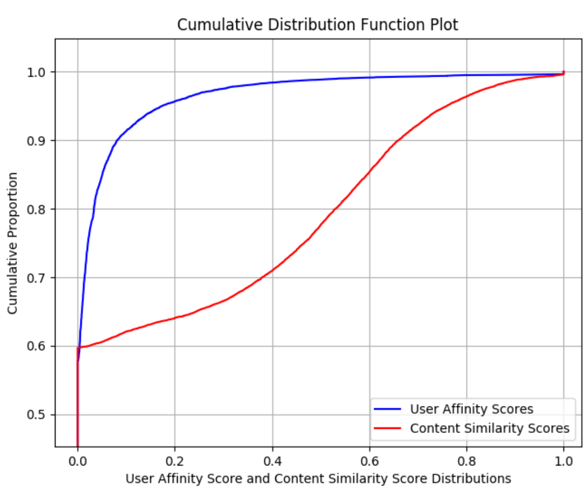

Fig. 3. Cumulative Distribution Function plot showing User Affinity and Content Similarity scores for 100 Twitter Users

[4]https://developer.twitter.com/

C. Categorization of Followees of User

After having studied the user affinity and content similarity scores in Fig 3, we see that 75% of user-followee pairs have user affinity score ((UAS) and content similarity score (CSS) less than 0.04 (th_{UAS}) and 0.5 (th_{CSS}), respectively. Henceforth, we take these values as thresholds to characterize user-followee behavior into either *user-conscious* or *content-conscious*. Empirically, we define a user to behave user-consciously with its followeee if UAS is greater than th_{UAS}. Similarly, we define a user to behave content-consciously with its followee if CSS is greater than th_{CSS}. In Fig 4, we plot number of followee with user-conscious behavior and number of followee with content-conscious behavior for each Twitter user under study. Circles in 'blue' represent those users who have more user-conscious followees than content-conscious followees whereas circles in 'red' means vice-versa.

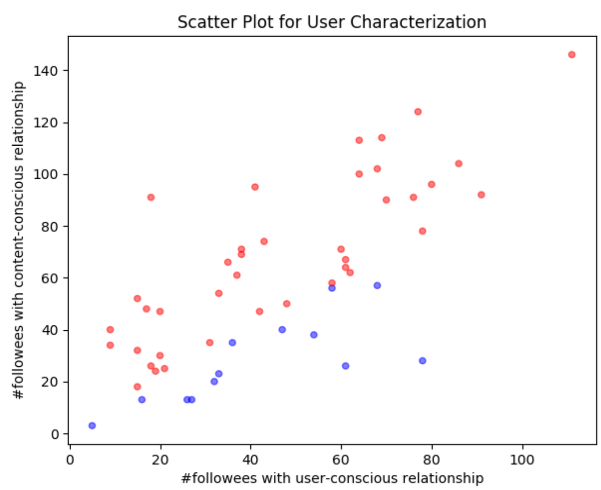

Fig. 4. Scatter Plot depicting Twitter Users. Each point (shown in blue or red) represents a user. X-axis shows number of followees with whom a user shows user-conscious behavior. Y-axis shows number of followees with whom a user shows content-conscious behavior

V. Followee Management Nudge

Our work is aimed at improving the user experience for Twitter users by helping them manage their followee so that they can view relevant content in their timeline. User-Followee behavioral characterization helps in listing those followees with whom user has user affinity score less that th_{UAS} and content similarity score less than th_{CSS}. These followees are shown to the user in our *Followee Management Nudge* (Fig 1) so that user is informed that these are the followees with whom user has minimally engaged nor does user has posted content similar to the content posted by these followees. Along with each of these followees, an option to *unfollow* is also provided to the user.

A. System Design

Followee Management Nudge is implemented as Google Chrome plugin. Chrome Extensions are software programs that help in customizing the browsing experience. They enable developers to modify Chrome Browser functionality and behavior as per user needs or preferences.

B. Evaluation of Nudge

Participants: In order to gauge users perceptions and opinions with respect to followee management nudge, we engaged with around 15 participants from our friend network. A pre-study questionnaire was prepared to gauge demographics and Twitter usage of the participants. Subsequently, we filtered out and recruited only 8 participants for controlled lab study who had more than 150 followees on Twitter. These participants were within the age group of 18-26 years, with 62.5% female and 37.5% male comprising of all undergraduate engineering students studying computer science.

Study Design: We conducted controlled lab study in following three steps: 1. Pre-Study questionnaire was filled by all initial participants. 2. Shortlisted participants were asked to use our plugin on their respective browsers. They were expected to sign-into their Twitter accounts following which they were displayed a list of their followees. 3. Lastly, these participants were asked to fill the Post-Study questionnaire.

Results: Here we present our observations and outcomes of user interactions with our nudge. 5 out of 8 (62.5%) participants found our nudge to be useful. After unfollowing few of suggested followees, 3 participants said that it helped them in making their timelines more relevant. However, the participants who didn't find it helpful had almost similar comment conveying that *even though they don't 'interact' with their tweets, but, they still would like to receive them.*

VI. Conclusion and Future Work

In this paper we have presented an approach that would be helpful to some users in reducing noise in their timelines. We proposed two metrics namely user affinity score and content similarity score using which we characterize user-followee relationship as either user-conscious or content-conscious, respectively. We leverage this characterization to develop Followee Management Nudge that would help user unfollow some of their followees with whom they rarely engage nor do they find their content relevant.

With that said, this is a preliminary work in the direction of followee management. Ways to figure out user's relevance with their followee can be improved further, in particular, *how to quantify relationship between a user-followee pair who has minimally or never interacted.* User base needs to be scaled up. Nudge may be put into field trial to find more interesting user observations and feedbacks.

References

[1] C. Sturk, "A look at the many different uses of twitter," *Life Wire*, May 2018, [Online; posted 23-May-2018]. [Online]. Available: https://www.lifewire.com/what-is-twitter-for-3288888

[2] J. Glum, "Who to follow on twitter in 2018," *News Week*, December 2017, [Online; posted 14-December-2017]. [Online]. Available: http://www.newsweek.com/best-twitter-accounts-2018-follow-747168

[3] F. Bridges, "Top 25 best twitter accounts to follow to improve your life," *Forbes*, February 2017, [Online; posted 28-February-2017]. [Online]. Available: https://www.forbes.com/sites/francesbridges/2017/02/28/the-25-best-twitter-accounts-to-follow-to-improve-your-life

[4] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida, "Characterizing user behavior in online social networks," in *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement*. ACM, 2009, pp. 49–62.

[5] E. Gilbert and K. Karahalios, "Predicting tie strength with social media," in *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2009, pp. 211–220.

[6] L. Banks and S. F. Wu, "All friends are not created equal: An interaction intensity based approach to privacy in online social networks," in *Computational Science and Engineering, 2009. CSE'09. International Conference on*, vol. 4. IEEE, 2009, pp. 970–974.

[7] D. Sousa, L. Sarmento, and E. Mendes Rodrigues, "Characterization of the twitter@ replies network: are user ties social or topical?" in *Proceedings of the 2nd international workshop on Search and mining user-generated contents*. ACM, 2010, pp. 63–70.

[8] C. Wagner, S. Asur, and J. Hailpern, "Religious politicians and creative photographers: Automatic user categorization in twitter," in *Social Computing (SocialCom), 2013 International Conference on*. IEEE, 2013, pp. 303–310.

[9] T. C. Leonard, "Richard h. thaler, cass r. sunstein, nudge: Improving decisions about health, wealth, and happiness," 2008.

[10] D. G. Goldstein, E. J. Johnson, A. Herrmann, and M. Heitmann, "Nudge your customers toward better choices," *Harvard Business Review*, vol. 86, no. 12, pp. 99–105, 2008.

[11] D. M. Hausman and B. Welch, "Debate: To nudge or not to nudge," *Journal of Political Philosophy*, vol. 18, no. 1, pp. 123–136, 2010.

[12] Y. Wang, P. G. Leon, K. Scott, X. Chen, A. Acquisti, and L. F. Cranor, "Privacy nudges for social media: an exploratory facebook study," in *Proceedings of the 22nd International Conference on World Wide Web*. ACM, 2013, pp. 763–770.

[13] R. Kaushal, S. Chandok, P. Jain, P. Dewan, N. Gupta, and P. Kumaraguru, "Nudging nemo: Helping users control linkability across social networks," in *International Conference on Social Informatics*. Springer, 2017, pp. 477–490.

[14] M. G. Armentano, D. Godoy, and A. A. Amandi, "Followee recommendation based on text analysis of micro-blogging activity," *Information systems*, vol. 38, no. 8, pp. 1116–1127, 2013.

[15] T. Chen, L. Tang, Q. Liu, D. Yang, S. Xie, X. Cao, C. Wu, E. Yao, Z. Liu, Z. Jiang *et al.*, "Combining factorization model and additive forest for collaborative followee recommendation," *KDD CUP*, 2012.

[16] H. Chen, X. Cui, and H. Jin, "Top-k followee recommendation over microblogging systems by exploiting diverse information sources," *Future Generation Computer Systems*, vol. 55, pp. 534–543, 2016.