2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

Information Adoption via Repeated or Diversified Social Influence on Twitter

Jaqueline Faria de Oliveira[*†], Humberto Torres Marques-Neto[*] and Márton Karsai[‡†]
[*] Department of Computer Science, PUC Minas, Belo Horizonte, Brazil
[†] Univ Lyon, ENS de Lyon, Inria, CNRS, UCB Lyon 1, LIP UMR 5668, IXXI, F-69342, Lyon, France
[‡] Department of Network and Data Science, Central European University, 1010 Vienna, Austria
jaqueline.oliveira@sga.pucminas.br, humberto@pucminas.br, karsaim@ceu.edu

***Abstract*—Influence arriving via social ties may be relevant for a person to decide to buy a new product, share information, or to adopt a new behaviour. However, quantifying social influence is a difficult task, even in online social systems where the interactions and communication content can be closely followed. Here we study the *information susceptibility* and *adoption thresholds* of users on Twitter. We consider hashtag and retweet adoptions on different aggregation levels: items, users, and topic groups, and study these adoption mechanisms characterized by diversified or repeated influence stimuli. We find both metrics to be heterogeneously distributed, correlated, and dependent on the topics and aggregation level of social influence. We show that users adopt retweets easier than hashtags, and find that new influencing neighbors can effectively trigger adoptions. Our results may inform better models of adoption processes leading to a deeper empirical understanding of simple and complex contagion.**

***Index Terms*—Social Influence, Susceptibility, Adoption Threshold**

I. Introduction

Think about how you choose which news you share next online or the type of phone you are about to buy. Would it be easier to decide after hearing many times the same information from the same people, or if many different friends would tell you the same? This is a central question in studies of social influence where different decision mechanisms of adoptions have been identified behind these two scenarios. In reality, decisions cannot be associated clearly to one of these mechanisms but probably have combined effects on one's conclusion. In this paper, we aim to observe differences between these scenarios by measuring information susceptibility and adoption thresholds of hashtags and retweets on different levels, using a large Twitter data set collected for this purpose.

The adoption of information, products, or any behavioral pattern is partially driven by social influence and may be interpreted as a spreading process, driven by certain mechanisms, between people connected in a social structure. To characterize the effects of these different influence mechanisms one could devise different measures. *Susceptibility* [1] is defined as the fraction of the number of adoption and exposure to a set of items propagated multiple times from the peers of an ego. This metric captures well the effects of *simple contagion* [2]–[4] as it measures how the probability of adoption changes via multiple stimuli coming even from a single neighbor. *Adoption threshold* [5], on the other hand, measures the fraction of already adopted and the total number of friends at the time of a single adoption of an ego. This metric captures directly individual thresholds assumed to drive adoption spreading as suggested by *complex contagion* [5], [6]. Even the differences between the two discussed adoption mechanisms are easy to understand, it is surprisingly difficult to distinguish the effects of simple and complex contagion in local and global observations. To come over these obstacles, first we need to look at real data to understand the emergent heterogeneities and the correlations between adoption susceptibility and thresholds.

Our goal in this paper is to contribute to this challenge by directly measuring the threshold and susceptibility of information adoption by *users*, of *items* and *topics*. We propose to study these measures associated with social contagion mechanisms characterized by diversified or repeated influence stimuli. As a novel approach, we observe adoptions not as a part of a larger diffusion process, but on the individual level by following the adoption dynamics of single users and their egocentric network. We collected a specific data corpus from Twitter for this study, which contains the timeline of ~ 1.2 million users. $8,527$ of them are seed individuals who appeared active during the European Election in 2019 in France, while others being their influencing followee peers.

II. Related work

The understanding of adoption behavior in different contexts like marketing, political campaigns, or information and misinformation diffusion have been in the focus of research since a long time [3]. These earlier studies commonly focused on to identify influential users [7]–[9], or to model information propagation [5], [10], and especially to predict large adoption cascades using machine learning [11], [12] or mechanistic modeling [5], [13].

One central measure to quantify adoption behaviour is called *user susceptibility*. Following some early studies [14], user susceptibility of adoption was mostly studied in online social systems, where such behavior and the underlying social network were easier to follow [15]. These studies reported on adoption of products [14], [16], behavior [17], [18], social relations [15], adherence to information and misinformation [19], or on the relationship between susceptibility and virality in information diffusion [1], [20] and sentiment diffusion [21].

IEEE/ACM ASONAM 2020, December 7-10, 2020
978-1-7281-1056-1/20/$31.00 © 2020 IEEE

They typically employed binary measures [8], [17], [18] to quantify individuals' sensitivity to adopt some behavior or information. Interestingly, some works found that susceptibility appears strongly heterogeneous between people [1], [20]–[22]. Modelling studies [22] incorporated these observations and showed that variable susceptibility and multiple exposures are key to predict diffusion outcomes. Closest to our study, Hoang et al. [1] analyzed information diffusion from a virality and susceptibility point of views and proposed a measure for varying user susceptibility as the fraction of items a user adopts and the number of times any of these items were diffused to him/her. We extend this definition by introducing susceptibility at different levels of aggregation to characterise better adoption behaviour.

Adoption threshold of behavior proposes another broadly accepted measure to quantify adoption behaviour. It is an early concept first formulated by Schelling and Granovetter [4], [6], who recognized that decisions about moving to a new neighborhood or joining a riot can be conditional to a cognitive threshold of social influence. They quantified this threshold via the fraction of people in ones social circle already engaged in such behaviors. Later the underlying network structure and the distribution of individual thresholds were shown to be the condition of globally emerging adoption cascades in these systems [5]. Nevertheless, early threshold models usually assumed for simplicity a random underlying network and a constant threshold for every node. However, social networks are known to be degree heterogeneous [22], [23], and more importantly, it has been shown that even adoption thresholds may be distributed broadly very similar to a log-normal distribution [13].

III. Data description

To measure susceptibility and threshold of adoption we collected a very specific dataset over the open API [24] of the online social system of Twitter.

1) Data collection and filtering: We followed the adoption of some hashtags and retweets from the collected timelines of specific seed users and their influencing followee friends. We focused on the period of the European Union Election in 2019 which took place between the 23 and 26 May. We first collected tweets through the streaming API between May 25 and June 3 2019, and selected users posting hashtags identified in Worldwide trends topics on Twitter to be related to the European Union Election 2019. This way we obtained $187,767$ distinct users posted $432,861$ tweets including hashtags like `#epelection2019`, `#euelection2019`, or `#europeanelection2019`. From them we selected our seed users, all who declared French as their official language in their user profile, and followed more than 100 other users. Finally we collected their ego-networks built from others they followed (called as followees from now on). This led us to $2,082,090$ users (combined) including $8,554$ seed users.

To track the incoming influence and adoption behavior, in the second part of our data gathering, we collected timelines containing all tweets (maximum $3,200$ API limitation) posted by any seed users and their followees between May 1 2018 and May 31 2019. This way $\sim$ 66% of users had their timeline collected, including 99% of the seed users. Users without timelines had whether no tweets during the recorded period or their profiles were private or deleted. In the end we obtained $1,844,978$ users among which $8,527$ were seed users. We collected $\sim$ 1.2 million of timelines and $\sim$ 1.1 billion of tweets of which 42% were retweets. Approximately 23% of the tweets contained at least one of the observed $19,233,668$ distinct hashtags but about $\sim 50\%$ of them were mentioned only once. Note that when tracking the adoption of hashtags, we considered only hashtags which were posted in original tweets by seed users. Although this condition reduced considerably our observation set, it provided a way more accurate identification of hashtag adoptions.

2) Hashtag grouping: Large adoption cascades of single hashtags are very rare events, thus they largely limit the observations of susceptibility and threshold variance of a single spreading item. One solution to this obstacle is grouping hashtags and assuming that hashtags of the same topic are adopted in similar ways. To do so we chose the top $1,000$ most adopted hashtags by seed users, and used the Word2Vec [25] world embedding method to represent all words and hashtags in a lower dimensional vector space. We used k-means clustering to group vectors into 10 clusters (optimal cluster number obtained using Davies-Bouldin Index [26] and the Silhouette Score [27]). Finally, by looking at the grouped hashtags manually, we found most of them grouped in clusters on topics like the *European election*, *Climate change*, *Notre-Dame* and *Protests*, followed by *International politics*, *Gilets jaunes*, *French politics*, *Entertainment*, *Sports* and *Tourism*.

IV. Terminology

Our methodological goal is to extend the conventional definition of adoption susceptibility and threshold for their comparison at different aggregation levels of social influence.

More formally, we observe the adoption of a U set of seed users influenced by their F set of neighbors (followees) to adopt some items from a set X. In this context, the egocentric network of a seed user $u \in U$ is defined by his/her followee set $F(u) \in F$ and taken as static. The set of adopted items of a seed user u is indicated as $X(u)$.

To detect potential influence for adoption, we consider the posting time $t_u(x)$ of an item x by a user u. Further, similar to Hoang and Lim [1], we take a time window τ and say that a followee $v \in F(u)$ influenced a user u to adopt item x at time $t_u(x)$ if (s)he posted x at any time $t_v(x)$, such that $t_u(x) - \tau \geq t_v(x)$. Conversely, we say that a user u was influenced to adopt an item x, if (s)he posted it at $t_u(x)$ for the first time within τ after any of its followee. Note that the action of influence is based on the arguable assumption that posted items by followees appear on the wall of followers. Unfortunately the data allows no better proxy for social influence in this setting. To best estimate τ, we measured its value as $\tau = 2 \times \langle \max(\{t_u(x) - t_v(x)\}_v) \rangle_{u,x}$ for each followee $v \in F(u)$. Averages run over every $u \in U$

and $x \in X_F(u)$ (set of all items introduced to user u by $F(u)$) with the condition that $t_u(x) - t_v(x) < 60$ days. We found $\tau = 2$ days for retweets and $\tau = 20$ days for hashtags, reflecting typically faster adoption of retweets.

To better quantify influence, we say that an item x has been *introduced* to a user u, if x has been posted by any of the followees $v \in F(u)$ of u maximum τ time before the adoption by u. We denote the set of introduced items as $X_I(u,x)$ and by $F_I(u,x)$ the set of influencing followees. We indicate the total (non-unique) set of introduced items by $X_I(u)$, which may contain the same item multiple times if introduced more than once even by the same followee. We also define $X_A(u)$, the set of *adopted* items (called set of diffused items in [1]) including all items what user u adopted within τ time after its last introduction to him/her. Note that $X_A(u)$ is unique as a user can adopt an item only once. Further we note by $F_A(u,x)$ the followee set of user u, who adopted item x earlier than u. For the adoption of items belonging to an item category set X_i we denote by $X_I(u,X_i)$ the set of introduced, and by $X_A(u,X_i)$ the set of adopted items from an item set X_i.

V. Results

Our main analysis is built on the simultaneous observation of the timelines and egocentric networks of seed users. Using the timestamps of tweets, we reconstruct the order of posts in each ego-network to identify the time of potential influence from followees for each item. This allows us to directly measure susceptibility and adoption thresholds in three aggregation levels: item, user, and hashtag topic, to learn more about the variance of adoption behaviour.

A. Susceptibility

Susceptibility is the ability of an individual to adopt an *item* while influenced by others [1]. It is defined as the fraction of total number of adoption of items in X and total number of introduction of any of these items to user u. It takes higher values for users who potentially easier adopt items from set X. We introduce susceptibility at different levels of aggregation.

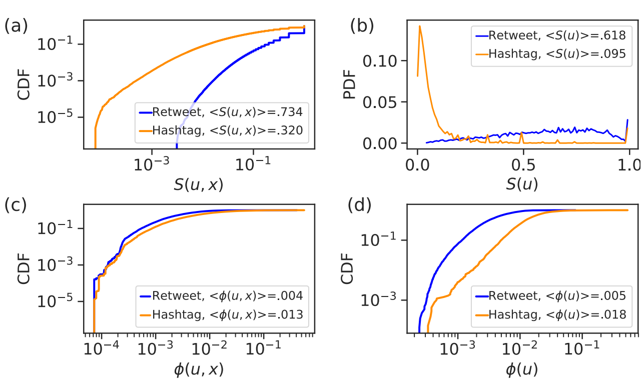

Fig. 1. Cumulative distributions of (a) item level susceptibility of retweets (blue) and hashtags (orange) adoptions; (b) Probability density functions of user level susceptibility; Cumulative distributions of (c) item level and (d) user level thresholds. Averages are shown in figure keys.

1) Item level susceptibility considers a user u to adopt a single item x and defined as $S(u,x) = 1/|X_I(u,x)|$. Although this definitions is somewhat trivial as the adopted item set is $X = \{x\}$, the distribution of this quantity appears to be very different between retweet and hashtag adoptions (Figure 1a). For retweets the average susceptibility $\langle S(u,x)\rangle = 0.734$ is high explained by the $\sim 80\%$ of retweet adoptions induced by single introduction, while on average 2.7 introduction is needed to convert a seed user. Meanwhile, for hashtags $\sim 82\%$ of adoptions preceded by more than one introduction and on average 40.8 introductions needed to make a seed user to adopt a hashtag. This is reflected by the significantly lower average susceptibility $\langle S(u,x)\rangle = 0.32$, suggesting stronger social influence necessary for the adoption of single hashtags.

2) User level susceptibility captures the overall susceptibility of a user to adopt any item introduced to her/him. It is defined as the $S(u) = |X_A(u)|/|X_I(u)|$ fraction of number of all adopted items and total number of introductions to user u.

Figure 1b shows striking differences between the adoption of retweets and hashtags. Most users have very small susceptibility to adopt hashtags with 39.6 introduction on average, reflected by the narrow distribution (orange) and its small average $\langle S(u)\rangle = 0.095$. Meanwhile, people are more susceptible for retweet adoption, doing it after 2.6 introductions on average. Interestingly, users are more homogeneous in terms of retweet adoption susceptibilities (blue) with a significantly larger average $\langle S(u)\rangle = 0.618$.

3) Topic level susceptibility takes a topical set of items X_i and it is defined as $S(u,X_i) = |X_A(u,X_i)|/|X_I(u,X_i)|$, i.e. the number of adoption of any topic $x \in X_i$ divided by the total number of their introductions to u. Corresponding distributions and their averages in Figure 2a indicate that people are almost twice more susceptible to adopt hashtags on *Tourism* or *Entertainment* as compared to topics on the *European Election* or social movements (*Gilets Jaunes*). However, this was somewhat expected as the topics with lower susceptibilities are the ones with more introductions.

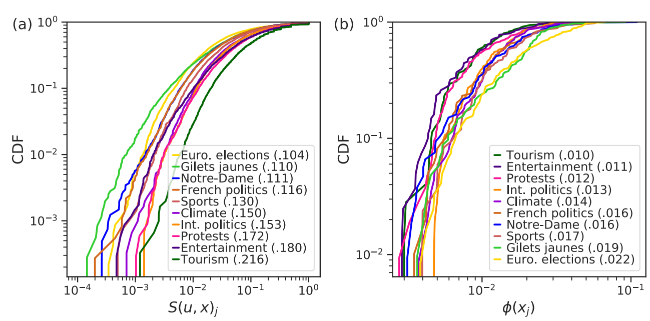

Fig. 2. (a) Cumulative distributions for (a) susceptibility and (b) threshold of hashtag topic groups. Keys are sorted by the corresponding average (a) susceptibility or (b) threshold values (in parenthesis).

B. Adoption threshold

The main hypothesis behind complex contagion is a cognitive threshold, which determines adoption behavior. Adoption thresholds were elegantly defined by Watts [5] for networks, what we extend to different aggregation levels here.

1) Item level thresholds, just like susceptibility, should reflect the strength of necessary social influence from different friends to adopt a single item. This can be formalized as the $\phi(u,x) = |F_A(u,x)|/|F(u)|$ fraction of number of followees who have adopted item x before u, and the total number of followees of u. To measure $\phi(u,x)$ we tracked the adoption order of seed users and their followees and obtained a heterogeneous distribution of item level thresholds with a broad tail (Figure 1c), in complete agreement with earlier observations [13]. Interestingly, retweet and hashtag adoption thresholds are distributed very similarly. However, since hashtag adoptions have three times larger average thresholds as retweets, this is in line with our earlier results suggesting easier adoption of retweets with smaller adoption thresholds.

2) User level adoption thresholds for users require certain aggregation. We define user level adoption threshold as $\phi(u) = avg(\phi(u,x))_x$, i.e. the average threshold over all adoptions of a user. This may capture differences between users (Figure 1d), while neglecting effects of item level heterogeneities. While user level thresholds are yet broadly distributed, they have more different retweet and hashtag adoption distributions as compared to item level. At the same time, consistently, thresholds of retweets appear four times smaller than hashtags.

3) Topic level thresholds, finally, are defined as the fraction $\phi(u,X_i) = |F_A(u, x \in X_i)|/|F(u)|$, which takes into account the adoption of any item x from a hashtag category X_i. Threshold distributions of each topic group (Figure 2b) appear relatively similar with broad tails and comparable average values (see figure keys for exact values). In line with our susceptibility results, the easiest adopted hashtags are related to *Tourism* or *Entertainment*, while the highest thresholds are again related to *Gilets Jaunes* or the *European Elections*.

C. Correlation of threshold and susceptibility

The question remains how the adoption threshold and susceptibility of people are correlated. By measuring the Pearson correlation coefficient between the user level susceptibilities to adopt retweets and hashtags of the same person we found a strong positive correlation with $r = 0.255$ ($p < 0.0001$). Users that are susceptible to adopt retweets, in many cases, have high susceptible values to adopt hashtags as well.

For the adoption thresholds of retweets and hashtags we found an even stronger correlation of $r = 0.391$ ($p < 0.0001$). Further we observed users with consistent adoption behavior as if they adopted easier hashtags they were more susceptibility for retweet adoptions too, with consistent threshold values.

We found strong negative correlations with $r = -0.218$ ($p < 0.0001$) between the hashtag susceptibility and threshold values of the same user, which is in agreement with earlier results. However, for retweet adoptions we observed a small positive correlation $r = 0.053$ ($p < 0.0001$). This is potentially because retweets can be introduced only once by a neighbor, making this measure more similarity to the threshold metric if each followee of a user is a prior adopter of many items.

D. Adoption due to reinforced or new influence

In this last section, we are interested in the role of reinforced or new influences that leads to the adoption of an ego. We concentrate exclusively on item level hashtag adoptions. First to quantify reinforced influence, we define two quantities: the inverse of item level susceptibility $S(u,x)^{-1} = |X_I(u,x)|$ capturing the number of *incoming influence* a user u received from his/her peers before adopting x; and the *number of unique neighbors* adopting earlier than u defined as $k_u\phi(u,x) = |F_A(u,x)|$.

The column-wise normalized $S(u,x)^{-1}$ distributions plotted for different $k_u\phi(u,x)$ values (Fig. 3a) show that the average number of reinforced influence (blue solid line in Fig. 3a) evolves surprisingly parallel with the theoretical minimum diagonal line, shifted with about 2.6 units up. This suggests that although one has only a few or many adopted friends, on average each of them needs to reinforce their influence 2.6 times to induce the adoption of the central ego. On one hand this suggests that reinforced influence is necessary to induce adoption, nevertheless this is a relatively small factor and more importantly seems to be largely independent from the number of influencing peers.

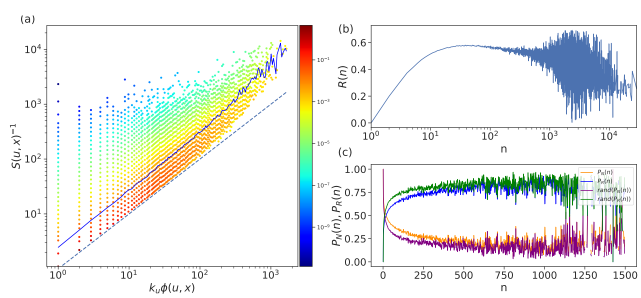

Fig. 3. (a) Column-wise normalized density plot of $k_u\phi(u,x)$ and $S(u,x)^-1$ value pairs of each hashtag adoption shown with averages (solid line) and theoretical minimum (dashed line). (b) Average entropy $R(n)$ of influence sequences with length n. (c) Last influence probabilities $P_N(n)$ and $P_R(n)$ of sequences with length n together with their shuffled references.

Finally, we focus on the sequence of influence events leading to the adoption of a hashtag x by u. This sequence consists of two types of events coding influence coming from a newly adopted neighbor (N) or a reinforced influence (R) from an already adopted peer. For each adoption this sequence necessarily starts with an N event but can continue with any series of R and N for any length n. We measure the average entropy $R(n)$ of such two-states sequences for each length n. This function starts from $R(n=1) = 0$ for sequences with a single N event and increases as more reinforcement R events appear until its maximum $R(n \simeq 50) \simeq 0.58$ were events are present with more equal probabilities. For longer n it decreases as reinforced events start to dominate the sequence. We can also easily measure the $P_N(n)$ probability that an n long sequence, followed immediately by the adoption of the ego, ends with an N event. Simultaneously, we can observe the $P_R(n)$ probabilities that it finishes with an R event. As a reference we can repeat the same measures after shuffling the labels of influence events, expect the very first one which is necessarily N. We found (Fig.3c) that for an extensive range of n, $P_N(n)$ probabilities are significantly larger than their randomized version, while the contrary is true for the $P_R(n)$. These results suggest that more sequence ending with a new neighbor influence than it would be expected by chance. Thus influence coming from a new neighbor may trigger adoptions with higher probability; a clear signature of threshold driven adoptions, where the number of different adopted neighbors matter more, while repeated influence coming from adopted peers may be less relevant to induce an adoption event.

VI. Conclusion

Built on the recent opportunities offered by online social systems we collected a very specific data set about the dynamical posting behavior of selected seed users and their complete ego-network on Twitter. By following the order of influence and adoption of retweets and hashtags, we observed two metrics: susceptibility and adoption thresholds, which can be associated with different social contagion mechanisms. We observed these measures at item, user, and topic group level of aggregations and empirically confirmed their heterogeneous distributions and correlations. We also observed significant differences between the adoption of retweets and hashtags, the former being easier to adopt. Finally, we observed that people need, on average, the same amount of reinforced influence from each of their neighbors, while a newly adopted peer can effectively increase the probability of adoption even after thousands of repeated influences from early adopters.

Our study has some limitations. Importantly, social influence might be weakly captured by our methodology, without any insights into the algorithmic solution of Twitter composing the post wall for each user. Nevertheless, our methodology can give an approximate idea about which tweet influenced users for adoption. The collected data set and the proposed analysis allows us to approach new research questions, like studying further the difference of repeated and diversified influence during a single adoption decision. This would lead us not only to better predictive models of adoption processes, but may also help us to identify ways to differentiate between simple and complex contagion mechanisms.

VII. Acknowledgments

This work was supported by CAPES, FAPEMIG (PPM-00253-18), and the STIC-AmSud Program (Project 18-STIC-07). MK was supported by the DataRedux (ANR-19-CE46-0008) and SoSweet (ANR-15-CE38-0011) ANR projects and the SoBigData++ (H2020-871042) project.

References

[1] T.-A. Hoang and E.-P. Lim, "Virality and susceptibility in information diffusions," in *Sixth international AAAI conference on weblogs and social media*, 2012.

[2] F. M. Bass, "A new product growth for model consumer durables," *Manage. sci.*, vol. 15, no. 5, pp. 215–227, 1969.

[3] E. M. Rogers, *Diffusion of innovations*. Simon and Schuster, 2010.

[4] T. C. Schelling, "Models of segregation," *Am. Econ. Rev.*, vol. 59, no. 2, pp. 488–493, 1969.

[5] D. J. Watts, "A simple model of global cascades on random networks," *Proc. Nat. Acad. Sci.*, vol. 99, no. 9, pp. 5766–5771, 2002.

[6] M. Granovetter and R. Soong, "Threshold models of diffusion and collective behavior," *J. Math. Sociol*, vol. 9, no. 3, pp. 165–179, 1983.

[7] D. J. Watts and P. S. Dodds, "Influentials, networks, and public opinion formation," *J. Consum. Res*, vol. 34, no. 4, pp. 441–458, 2007.

[8] S. Aral and D. Walker, "Identifying influential and susceptible members of social networks," *Science*, vol. 337, no. 6092, pp. 337–341, 2012.

[9] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts, "Identifying influencers on twitter," in *Fourth ACM International Conference on Web Seach and Data Mining (WSDM)*, 2011.

[10] Z. Ruan, G. Iniguez, M. Karsai, and J. Kertész, "Kinetics of social contagion," *Phys. Rev. Lett.*, vol. 115, no. 21, p. 218702, 2015.

[11] P. A. Dow, L. A. Adamic, and A. Friggeri, "The anatomy of large facebook cascades." *ICWSM*, vol. 1, no. 2, p. 12, 2013.

[12] J. Cheng, L. Adamic, P. A. Dow, J. M. Kleinberg, and J. Leskovec, "Can cascades be predicted?" in *Proceedings of the 23rd international conference on World wide web*. ACM, 2014, pp. 925–936.

[13] M. Karsai, G. Iñiguez, R. Kikas, K. Kaski, and J. Kertész, "Local cascades induced global contagion: How heterogeneous thresholds, exogenous effects, and unconcerned behaviour govern online adoption spreading," *Sci. Rep.*, vol. 6, p. 27178, 2016.

[14] W. O. Bearden, R. G. Netemeyer, and J. E. Teel, "Measurement of consumer susceptibility to interpersonal influence," *J. Consum. Res.*, vol. 15, no. 4, pp. 473–481, 1989.

[15] P. F. Bruning, B. J. Alge, and H.-C. Lin, "The embedding forces of network commitment: An examination of the psychological processes linking advice centrality and susceptibility to social influence," *Organ. Behav. Hum. Decis. Process.*, vol. 148, pp. 54–69, 2018.

[16] S. P. Tussyadiah, D. R. Kausar, and P. K. Soesilo, "The effect of engagement in online social network on susceptibility to influence," *J. Hosp. Tour.*, vol. 42, no. 2, pp. 201–223, 2018.

[17] C. Wagner, S. Mitter, C. Körner, and M. Strohmaier, "When social bots attack: Modeling susceptibility of users in online social networks." in *# MSM*, 2012, pp. 41–48.

[18] H. Mensah, L. Xiao, and S. Soundarajan, "Characterizing susceptible users on reddit's changemyview," 2019.

[19] B. E. Weeks, "Emotions, partisanship, and misperceptions: How anger and anxiety moderate the effect of partisan bias on susceptibility to political misinformation," *J. Commun.*, vol. 65, no. 4, pp. 699–719, 2015.

[20] T.-A. Hoang and E.-P. Lim, "Tracking virality and susceptibility in social media," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. ACM, 2016, pp. 1059–1068.

[21] R. K.-W. Lee and E.-P. Lim, "Measuring user influence, susceptibility and cynicalness in sentiment diffusion," in *European Conference on Information Retrieval*. Springer, 2015, pp. 411–422.

[22] L. Adamic *et al.*, "The diffusion of support in an online social movement: Evidence from the adoption of equal-sign profile pictures," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 2015, pp. 1741–1750.

[23] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, vol. 74, no. 1, p. 47, 2002.

[24] I. Twitter. Twitter developer. [Online]. Available: https://developer.twitter.com

[25] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv:1301.3781*, 2013.

[26] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE transactions on pattern analysis and machine intelligence*, no. 2, pp. 224–227, 1979.

[27] P. J. Rousseeuw, "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis," *J. Comp. App. Math.*, vol. 20, pp. 53–65, 1987.