# Online Social Behavior in Twitter: A Literature Review

Olav Aarts*, Peter-Paul van Maanen*†, Tanneke Ouboter* and Jan Maarten Schraagen*‡
*Netherlands Organisation for Applied Scientific Research (TNO), The Netherlands
Email: {olav.aarts,peter-paul.vanmaanen,tanneke.ouboter,jan_maarten.schraagen}@tno.nl
†Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
‡Twente University, Enschede, The Netherlands

***Abstract*—This literature review is aimed at examining state of the art research in the field of online social networks. The goal is to identify the current challenges within this area of research, given the questions raised in society. In this review we pay attention to three aspects of social networks: actor, message, and network characteristics. We further limit our review to research based on Twitter data, because this online social network is the most widely used by researchers in the field.**

***Keywords*-Twitter; Social network services;**

## I. Introduction

Social networks are around as long as humans organized themselves into groups. The digital age introducing social media gave social networks an unprecedented impulse to flourish. Meanwhile, quite a few studies have been carried out on social media and social networks. In order to get a comprehensive view of the literature, we limit our focus on the most widely used online social network for scientific analysis, Twitter. Most of the research in this field is based on data gathered from Twitter. Although research has been done based on other social networks and media like Facebook and YouTube, this work has not led to new insights, new approaches or novel contributions, as compared to Twitter. Hence, the research question we address in this paper is: what is the current state-of-art in both describing and explaining online behavior within Twitter and, based on this, what future research do we expect to be most promising? To answer this research question we have organized the literature on online behavior within Twitter according to three types of characteristics, namely actor, message, and network characteristics.

First and foremost, *actor characteristics* determine to what extent individuals are part of the network and how much influence the individual actor has to spread the information further. Evidently, actors form the building blocks of a social network. It is between these actors that communication or information exchange may or may not occur.

Second, *message characteristics* form a part of the explanation to the extent to which information circulates through the social network. Message characteristics play a role in the distribution and propagation of information. An attractive or stimulating message is more likely to spread to various individuals within the network. Unappealing or uninformative messages are likely to die out quickly.

Third, *network characteristics* give us insight in the relational features between actors instead of features confined to one actor only. The extent to which information spreads is not solely determined by either the characteristics of the actor or the message characteristics. The social context or network (who is connected to whom, and in which way) to a large extent determines the flow and diffusion of information. This relational approach is a distinctive feature of a social network approach.

This paper is divided into sections where we review the literature on actor, message and network characteristics and their effects on online social behavior, respectively. In the final section, we draw conclusions on all of these topics and discuss the open issues suitable for future research.

## II. Actor characteristics

Two aspects of actor characteristics are widely examined, individual informational influence and conformity to norm or social pressure. Individual informational influence is to a large extent determined by the relationship one has to the informant. Informational influence is stronger to the extent that it is exerted by others who are perceived as similar to the self on key dimensions as defined by Turner [1]. The relative anonymity on online media that users can create for themselves, allows for greater freedom from social norms and accountability.

Mainstream theories of social influence are largely inclined to suggest that social influence in groups is either due to rational individual information processing or conformity pressure. However, neither of these arguments can account for crowd behavior. Emergent behavior can arise without any clear normative framework and without accountability, which rules out the notion that conformity can sufficiently explain crowd behavior [2]. Additionally, some crowd behavior appears to be nonrational and it is thought that a crowd is able to prevent its members from engaging in thorough and systematic information processing. Hence, a classical solution has been to suggest that crowds behave as they do because normal individual processing is suspended. Stimuli found in crowd behavior lack normal cognitive functions, making the crowd members revert back to instinctive behavior normally held in check (now more salient due to relative anonymity). The reason for the apparent uniformity of emergent behavior is that we all want more or less the same thing [2].

There are two sides to anonymity: anonymity of others to oneself, and anonymity to others. A basic social psychological model is based on 1) the social context, 2) anonymity and its corresponding process and 3) the resulting outcome [3]. The model starts with social context determining whether group or personal identity is salient. Then moving on to the degree of anonymity and other de-individuating factors, the process of either increased or reduced group salience takes hold. The outcome of this being adherence to either group or personal norms and standards.

In the literature, several models have been proposed that try to predict the influence of actors in a network based on actor characteristics. Influence can be described by the amount or sizes of cascades in a network caused by a certain 'seed', i.e., a certain event, activation or intervention by a certain actor. A cascade is a sequence of activations generated by a contagion process, in which nodes (actors) cause connected nodes to be activated with some probability. Bakshy, Hofman, Mason & Watts [4] show a predictive model of influence, in which cascade sizes of posted URLs are predicted using the individuals' attributes and average size of past cascades. The following features are included as predictors:

Seed user attributes:

- number of followers
- number of friends
- number of tweets
- date of joining

Past influence of seed users:

- average, minimum, and maximum total influence
- average, minimum, and maximum local influence

The type of influence Bakshy *et al.* try to predict is of a rather narrow kind: being influenced to pass along a particular piece of information. Moreover, influencing another individual to pass along a piece of information does not necessarily imply any other kind of influence, such as influencing their purchasing behavior, or political opinion. The term 'influencer' should in the context of their study be interpreted as applying only very narrowly to the ability to consistently seed cascades that spread further than others.

In light of the emphasis placed on prominent individuals as optimal vehicles for disseminating information, the possibility that 'ordinary influencers'- individuals who exert average, or even less-than-average influence are under many circumstances more cost-effective, is intriguing. Correspondingly, Bakshy *et al.*'s conclusion that word-of-mouth information spreads via many small cascades, mostly triggered by ordinary individuals, is also likely to apply generally. They found that predictions of which particular user or URL will generate large cascades are relatively unreliable. Bakshy *et al.*

mainly use statistical analysis-based models and a potential disadvantage of their method is that it may mistakenly attribute influence to what is in reality a sequence of independent events.

Mehta, Mehta, Chheda, Shah & Chawan [5] have also modeled influence on Twitter using actor characteristics. Their study discusses how Twitter data is used as a corpus for analysis by the application of sentiment analysis and a study of different algorithms and methods that help to track influence and impact of a particular user/brand active on the social network. They showed that the number of followers an actor has, has much more predictive value with respect to estimating the amount of influence an actor has, as compared to the number of retweets and lists. No combined predictive value was reported.

## III. Message characteristics

Message characteristics, the second of three characteristics, deals about social transmission. Social transmission is everywhere. Friends talk about almost every topic. Interpersonal communication affects everything from decision making and well-being to the spread of ideas, the persistence of stereotypes, and the diffusion of culture. But although it is clear that social transmission is both frequent and important, what drives people to share? Why are some stories and information shared more than others? Traditionally, researchers have argued that rumors spread in the 3 C's (conflict, crisis, and catastrophe) and the major explanation for this phenomenon has been generalized anxiety (i.e., apprehension about negative outcomes). Such theories can explain why rumors flourish in times of panic, but they are less useful in explaining the prevalence of rumors in positive situations, such as a festival or other social event. Although recent work on the social sharing of emotion suggests that positive emotion may also increase transmission, why emotions drive sharing and why some emotions boost sharing more than others remains unclear [6].

Berger [6] suggests that transmission is driven in part by arousal. Physiological arousal boosts sharing. This hypothesis not only suggests why content that evokes more of certain emotions (e.g., disgust) may be shared more than other content, but also suggests a more precise prediction. Emotions characterized by high arousal, such as anxiety or amusement, will boost sharing more than emotions characterized by low arousal, such as sadness or contentment.

Naveed, Gottron, Kunegis & Alhadi [7] studied Twitter by examining retweets, thereby determining what the Twitter community considers interesting on a global scale. They used this to construct a function of interestingness to generate a model to describe the content-based characteristics of retweets. They trained a prediction model to predict for a given tweet its likelihood of being retweeted based on its contents. From the parameters learned by the model they deduced what the influential content features were that contributed to the likelihood of a retweet (e.g., hashtag, username(@), URL, questionmark, arousal, dominance, negative emoticon). They found that general topics affecting many users (e.g., Christmas) are more likely to be retweeted than narrow, personal topics. Also, messages with hashtags, usernames, URLs, question marks, and exciting and intense tweets are more likely to be retweeted than messages with exclamation marks or positive emoticons.

Also looking at retweeting as the key mechanism for information diffusion in Twitter, Suh, Hong, Pirolli & Chi [8] gathered content and contextual features from 74M tweets and used this data set to identify factors that were significantly associated with retweet rate. They found that, amongst content features, URLs and hashtags have strong relationships with retweetability. Amongst contextual features, the number of followers and followees as well as the age of the account seemed to affect retweetability, while, interestingly, the number of past tweets did not predict retweetability. Furthermore, they state that retweets have quite different content characteristics from normal tweets: 56.7% of retweets have URLs in them while only 19.0% of regular tweets have URLs. We believe that this research would inform the design of sense making and analytics tools for social media streams. A possible drawback of the method of Suh *et al.* is that it does not study the causal relation between, for instance, URLs and the probability a tweet is going to be retweeted: the presence of URLs can be a result of a third variable that was not measured.

Ye & Wu [9] analyzed the propagation patterns of general messages and show how breaking news (Michael Jackson's death) spreads through Twitter. They evaluated different social influences by examining their stabilities, assessments, and correlations, thereby aiming to characterize information propagation and social influence. Their results show that in general, a tweet is retweeted quickly and a significant portion of messages propagate far away from the originator, and the discussions are not restricted to his/her followers. Though this latter finding is rather polarized, either a message (about half) is propagated one hub only or four hubs or more (37%). A message is replied quickly and a message flow does not last long (less than one hour).

Bakshy *et al.* [4] also studied the effects of content. They found that content that is rated as more interesting tends to generate larger cascades on average, as does content that elicits more positive feelings (confirming Naveed *et al.* [7]). Most explanations of success tend to focus only on observed successes, which invariably represent a small and biased sample of the total population of events. When the much larger number of non-successes are also included, it becomes difficult to identify content-based attributes that are consistently able to differentiate success from failure at the level of individual events. Bakshy *et al.* mention that many more experimental studies are needed to solve this issue.

Lehmann, Gonçalves, Ramasco & Cattuto [10] examined how users respond to an incoming stimulus, i.e., a tweet (message) from a friend, and showed that the 'principle of least effort' combined with limited attention plays a dominant role in retweeting behavior. Specifically, they observed that users retweet information when it is most visible, such as when it is near the top of their Twitter stream. Moreover, their measurements quantify how a user's limited attention is divided among incoming tweets, providing novel evidence that highly connected individuals are less likely to propagate an arbitrary tweet. Their study indicated that the finite ability to process incoming information constrains social contagion, and they concluded that rapid decay of visibility is the primary barrier to information propagation online.

## IV. Network characteristics

Lerman & Ghosh [11] show that the social network structure behind news sites such as Twitter and Digg plays a crucial role in the spread of information within these sites. Differences between Twitter and Digg are primarily due to Digg's use of their front page which promotes articles and a relatively small number of top users that promote articles. Popularity of news stories and blog posts on Twitter grows smoothly until saturation, via spreading through the follower graph. The evolution of a story on Digg shows two distinct phases: an upcoming phase and a promoted phase (showing a tipping point), driven by Digg's promotion mechanism. On both sites, it takes a day or less for the number of votes/retweets to saturate to their final values. On Twitter, many URLs never spread beyond the seed and its followers. A handful of URLs spread more than ten hops from the seed. On Twitter, far fewer URLs spread within a community than on Digg. Each retweeter is most likely to follow only one previous retweeter. On Twitter, cascades are more tree-like because the follower graph does not have significant community structure. On both networks, though information cascades spread fast enough for one seed to infect thousands of users, they end up affecting less than 1% of the follower graph.

In the literature, different effects are studied that limit the final size of epidemics on social network platforms. Weng, Flammini, Vespignani & Menczer [12] examined whether the diversity of information we are exposed to, and the fading of our collective interests for specific topics affect the popularity of different memes. People share messages on a social network but can only pay attention to a portion of the information they receive. In the emerging dynamics of information diffusion, a few memes go viral while most do not. Weng *et al.* constructed an Agent-based Model (ABM) that can explain the massive heterogeneity in the popularity and persistence of memes as deriving from a combination of the competition for our limited attention and the structure of the social network, without the need to assume different intrinsic values among ideas. However, even in the simplified settings of social media platforms, it is hard to disentangle the effects of limited attention from many concurrent factors, such as the structure of the underlying social network, the activity of users and the size of their potential audience, the different degrees of influence of information spreaders, the intrinsic quality of the information they spread, the persistence of topics, and homophily.

In another study, Weng, Lim, Jiang & He [13] have shown that network characteristics of mutual connectivity between nodes on Twitter (friend) can be explained by 'homophily' (i.e., a twitterer follows a friend, and he or she follows back because they share a topic of interest).

Gonçalves, Perra & Vespignani [14] found that users can entertain a maximum of 100–200 stable relationships. They conducted a large survey of online exchanges or conversations on Twitter, collected across six months involving 1.7 million individuals. They tested the theoretical cognitive limit on the number of stable social relationships, known as Dunbar's number. On the basis of this empirical evidence, they proposed a simple dynamical mechanism, based on finite priority queuing and time resources, that reproduces the observed social behavior. Even in the online world, cognitive and biological constraints hold as predicted by Dunbar's theory, limiting users' social activities. This simple model offers a basic explanation of a seemingly complex phenomenon observed in the empirical patterns on Twitter data and offers support to Dunbar's hypothesis of a biological limit to the number of relationships.

In relation to the network's structure, Ver Steeg, Ghosh & Lerman [15] observed a difference in the dynamics of the epidemic threshold versus the final epidemic size (popularity of a topic). On the one hand, it is easier for a story to take off within a smaller, more tightly connected community, thereby lowering the epidemic threshold. On the other hand, for cascades to grow very large, it is better to have a more homogeneous link structure to reach all parts of the graph quickly. Clusters have the effect of marginally decreasing the size of cascades by sequestering an infection in one part of the graph.

According to Ardon, Bagchi, Mahanti, Ruhela, Seth, Tripathy & Triukose. [16], topics become popular when disjoint clusters of users discussing them begin to merge and form one giant component. Less popular topics generally exist in highly disconnected clusters. Topics that are going to become very popular witness intense discussion within communities at first. When the level of intensity rises then the users who bridge communities enter the discussion in a big way causing a merging of what were earlier disjoint discussions. Twitter is a partially democratic medium in the sense that popular topics are generally started by users with high numbers of followers (called 'celebrities'); however, for a topic to become popular it must be taken up by non-celebrity users.

Grabowicz, Ramasco, Moro, Pujol & Eguiluz [17] pay attention to the fact that different kinds of (intimacy) relations in the network have an influence on the information contagion. Strong ties refer to relations with close friends or relatives, while weak ties represent links with distant acquaintances. Grabowicz *et al.* observed that weak ties act as bridges between groups and are important for the diffusion of new information across the network, while strong ties are usually located at the interior of the groups. Weak ties are exemplified in Twitter by retweets, whereas mentions are exemplary for strong ties (internal links).

As described above, the sample network influences the reliability and the extent of generalizing results. Kwak, Lee, Park & Moon [18] conducted the first quantitative study on the entire Twittersphere and information diffusion on it (by crawling the Twitter site they obtained 41.7 million user profiles, 1.47 billion social relations, 4762 trending topics, and 106 million tweets). In its follower-following topology analysis they found a non-power-law follower distribution, a short effective diameter (average path length of 4.12), and low reciprocity (only 22,1% of the linked pairs have a reciprocal relationship). They found chains on Twitter to be at most of length ten, with the spread having a long tail distribution ranging up to hundreds of nodes. They showed that the majority (over 85%) of topics were headline news or persistent news in nature. Any retweeted tweet is to reach an average of 1,000 users no matter what the number of followers is of the original tweet. Once retweeted, a tweet gets retweeted almost instantly on next hops, signifying fast diffusion of information after the 1st retweet.

In his bachelor thesis, focusing solely on celebrities, Rosenman [19] described several aspects of the network of Twitter. Only about 22% of Twitter relationships were mutual. Users often connected with individuals with little intention of actively communicating with them. Most tweets received very little attention, but a handful received a very large amount of attention, leading to a long-tailed power-law distribution.

Rosenman states that influence is not a simple matter of follower count. Actual activity needs to be taken into account, e.g., number of retweets (influence) or mentions (buzz). He found that, across different types of influence, the degree to which a celebrity is discussed on Twitter is an extremely useful predictor, while follower counts are comparatively less predictive. His results showed that on the sample of 60 most followed celebrities:

- High buzz (the degree to which a celebrity is talked about on twitter) is associated with a large number of retweets (Spearman 0.70)
- Tweets sent and retweet count are positively correlated (Spearman 0.44)
- Retweets are significantly correlated with follower counts (Spearman 0.58)
- No correlation between audience size and tweeting frequency (followers are not attracted to an abundance of content generation)
- No correlation between reference count and tweeting frequency (amount of buzz is largely independent of content generation; e.g., Beyonce never sent a single tweet but attracted a large number of followers and generated a lot of buzz)

Hence, both buzz and audience size are strong indicators of influence, but they do not tell the whole story. For instance, there is extremely high variance in the distribution of retweet counts, and there is no certainty that any celebrity tweet would generate retweets. Hence, retweet prediction is extremely difficult for individual tweets. Nonetheless, according to Rosenman, such a model would be incredibly useful in allowing marketers to project the actual number of impressions to be generated on a tweet-by-tweet basis for prominent celebrity tweeters.

A large body of work has been devoted to defining and identifying clusters or communities in social and information networks, i.e., in graphs in which the nodes represent underlying social entities and the edges represent some sort of interaction between pairs of nodes. Most such research begins with the premise that a community or a cluster should be thought of as a set of nodes that has more and/or better connections between its members than to the remainder of the network. Leskovec, Lang, Dasgupta & Mahoney [20] explored from a novel perspective several questions related to identifying meaningful communities in large social and information networks, and came to several striking conclusions.

Rather than defining a procedure to extract sets of nodes from a graph and then attempt to interpret these sets as 'real' communities, they employed approximation algorithms for the graph partitioning problem to characterize as a function of size the statistical and structural properties of partitions of graphs that could plausibly be interpreted as communities. In particular, they defined the network community profile plot, which characterizes the 'best' possible communityaccording to the conductance measure (the ratio between outgoing edges/internal edges over a wide range of size scales). They studied over 100 large real-world networks, ranging from traditional and on-line social networks, to technological and information networks and web graphs, and ranging in size from thousands up to tens of millions of nodes.

Leskovec *et al.* observed tight communities that were barely connected to the rest of the network at very small size scales (up to 100 nodes); and communities of size scale beyond 100 nodes that gradually 'blended into' the core of the network and thus became less 'community-like', with a roughly inverse relationship between community size and optimal community quality. This observation agrees well with the so-called Dunbar number which gives a limit to the size of a well-functioning community.

To Castellano, Fortunato & Loreto [21] computer simulations play an important role in the study of social dynamics since they parallel more traditional approaches of theoretical physics aimed at describing a system in terms of a set of equations, to be later solved numerically and/or, whenever possible, analytically. One of the most successful methodologies used in social dynamics is agent-based modeling. The idea is to construct the computational devices (known as agents with some properties) and then simulate them in parallel to model the real phenomena. In physics this technique can be traced back to molecular dynamics [22] and Metropolis and Monte Carlo simulations [23]. The goal is to address the problem of the emergence from the lower (micro) level of the social system to the higher (macro) level. Their models of crowd behavior [23, p. 36] seems a promising venue to pursue.

## V. Conclusion

In answer to the first question addressed in this review (what is the current state-of-art in both describing and explaining online behavior within Twitter?), we can conclude with the following.

*Actor characteristics* are widely used to model social influence within Twitter and research on this topic is mainly focused on two aspects: individual information influence and conformity to social pressure. For individual behavior, several attributes are defined that try to predict the influence of an individual actor in a network. It is also shown that when actors become part of a group (crowd behavior) their behavior becomes more uniform, and individual attributes will lose their predictive value.

Explanatory and predictive models on *message characteristics* are in development. An important predictor for message transmission is related to the emotion of arousal: high arousal evoked by the content will boost sharing of information. Considering specific Twitter data, URL's and hashtags are also shown to have strong relations with retweetability. We conclude from the literature reviewed that influence metrics cannot be studied in isolation and that a broad notion of influence is required.

*Network characteristics* and relational aspects have received the least attention in the literature. An interesting but persistent finding related to these aspects is that 'viral' processes of online social influence simply do not happen, despite the widespread adherence to the idea. Information online spreads in short chains whereby some highly influential seeds (i.e., nodes) infect a vast number of others. We believe that an interesting issue for further research is why the chains of social influence remain so short in online situations.

Finally, we believe that, although actor, message and network characteristics were somewhat artificially isolated in the current review, as well as in much of the current research, these characteristics should actually be combined. Using cognitively-inspired agent-based modeling techniques in conjunction with system-level network parameters provides a promising venue for further research.

## References


[1] J. C. Turner, *Social influence*. Milton Keynes: Open University Press, 1991.

[2] T. Postmes, "The psychological dimensions of collective action, online," in *The Oxford handbook of internet psycology*, A. Joinson, K. McKenna, T. Postmes, and U. Reips, Eds. Oxford University Press, 2007, ch. 12, pp. 165–184.

[3] R. Spears, M. Lea, and T. Postmen, "Computer-mediated communication and social identity," in *The Oxford handbook of internet psycology*, A. Joinson, K. McKenna, T. Postmes, and U. Reips, Eds. Oxford University Press, 2007.

[4] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts, "Everyone's an influencer: Quantifying influence on twitter," in *Proceedings of the fourth ACM international conference on Web search and data mining*. ACM New York, NY, USA, 2011, pp. 65–74.

[5] R. Mehta, D. Mehta, D. Chheda, C. Shah, and P. M. Chawan, "Sentiment analysis and influence tracking using twitter," *International Journal of Advanced Research in Computer Science and Electronics Engineering*, vol. 1, no. 2, pp. 72–79, 2012.

[6] J. Berger, "Arousal increases social transmission of information," *Psychological science*, vol. 22, no. 7, pp. 891–893, 2011.

[7] N. Naveed, T. Gottron, J. Kunegis, and A. C. Alhadi, "Bad news travel fast: a content-based analysis of interestingness on twitter," in *WebSci '11: Proceedings of the 3rd International Conference on WebScience*, 2011.

[8] B. Suh, L. Hong, P. Pirolli, and E. H. Chi, "Want to be retweeted? large scale analytics on factors impacting retweet in twitter network," in *Proceedings of Social Computing / IEEE International Conference on Privacy, Security, Risk and Trust*, 2010.

[9] S. Ye and F. Wu, "Measuring message propagation and social influence on twitter.com," *Social informatics*, vol. 6430, pp. 216–231, 2010.

[10] J. Lehmann, B. Gonçalves, J. J. Ramasco, and C. Cattuto, "Dynamical classes of collective attention in twitter," in *Proceedings of the 21st International World Wide Web Conference (WWW)*, 2012.

[11] K. Lerman and R. Ghosh, "Information contagion: An empirical study of the spread of news on digg and twitter social networks," in *Proceedings of the Fourth International Conference on Weblogs and Social Media*, Association for the Advancement of Artificial Intelligence. AAAI Press, Menlo Park, California, 2010.

[12] L. Weng, A. Flammini, A. Vespignani, and F. Menczer, "Competition among memes in a world with limited attention," *Scientific Reports*, vol. 2, 2012, 29 March.

[13] J. Weng, E. P. Lim, J. Jiang, and Q. He, "Twitterrank: Finding topic-sensitive influential twitterers," in *Proceedings of the third ACM international conference on Web search and data mining*. ACM New York, NY, USA, 2010, pp. 261–270.

[14] B. Gonçalves, N. Perra, and A. Vespignani, "Modeling users activity on twitter networks: Validation of dunbars number," *PLoS ONE*, vol. 6, no. 8, 2011, e22656.

[15] G. Ver Steeg, R. Ghosh, and K. Lerman, "What stops social epidemics?" in *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, Association for the Advancement of Artificial Intelligence. AAAI Press, Menlo Park, California, 2011, pp. 377–384.

[16] S. Ardon, A. Bagchi, A. Mahanti, A. Ruhela, A. Seth, R. M. Tripathy, and S. Triukose. (2011, November) Spatio-temporal analysis of topic popularity in twitter. [Online]. Available: http://arxiv.org/abs/1111.2904

[17] P. A. Grabowicz, J. J. Ramasco, E. Moro, J. M. Pujol, and V. M. Eguiluz, "Social features of online networks: The strength of intermediary ties in online social media," *PLoS ONE*, vol. 7, no. 1, 2012, e29358.

[18] 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 World-Wide Web Conference (WWW)*, 2010.

[19] E. T. R. Rosenman, "Retweets — but not just retweets: Quantifying and predicting influence on twitter," 2012, bachelor's thesis, applied mathematics, Harvard College, Cambridge.

[20] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney, "Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters," *Internet Mathematics*, vol. 6, no. 1, pp. 29–123, 2009.

[21] C. F. Castellano, S. Fortunato, and V. Loreto, "Statistical physics of social dynamics," *Review of Modern Physics*, vol. 81, no. 2, pp. 591–646, 2009.

[22] B. J. Alder and T. E. Wainwright, "Phase transition for a hard sphere system," *Journal of Chemical Physics*, vol. 27, no. 5, pp. 1208–1209, 1957.

[23] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. Teller, and E. Teller, "Equation of state calculations by fast computing machine," *Journal of Chemical Physics*, vol. 21, no. 6, pp. 1087–1092, 1953.