

# Information Dissemination in Heterogeneous-Intent Networks

Abhimanyu Das  
Google Inc.

Emre Kiciman  
Microsoft Research

Sreenivas Gollapudi  
Google Inc.

Onur Varol  
Indiana University

## ABSTRACT

Many qualitative studies of communication practices on social media have recognized that people's motivation for participating in social networks can vary greatly. Some people participate for fame and fortune, while others simply wish to chat with friends. In this paper, we study the implications of such heterogeneous intent for modeling information diffusion in social networks. We *experiment* with user-level perception of messages, *analyze* large-scale information cascades, and *model* information diffusion in heterogeneous-intent networks. We perform carefully designed user studies to establish the relationship between the intent and language style of a message sender. Style of the user appear to adapt their language to achieve different intents. We perform a large-scale data analysis on Twitter message cascades and confirm that message propagation through a network is correlated with historical representations of individuals' intents. Finally, we posit a simple analytical model of information diffusion in social networks that takes heterogeneous intents into account and find that this model is able to explain empirically observed properties of structural virality that are not explained by current models.

## CCS Concepts

•Information systems → *Web mining*; •Social and professional topics → User characteristics;

## Keywords

information dissemination, user modeling, topic modeling, social media

## 1. INTRODUCTION

People participate in social media for many different reasons. Some join social media with the intention to socialize with friends or to meet new people. Others participate to promote a cause, or to gain fame as an authority or expert on their topics of interest. Much prior research has

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

*WebSci '16, May 22 - 25, 2016, Hannover, Germany*

© 2016 Copyright held by the owner/author(s). Publication rights licensed to ACM. ISBN 978-1-4503-4208-7/16/05...\$15.00

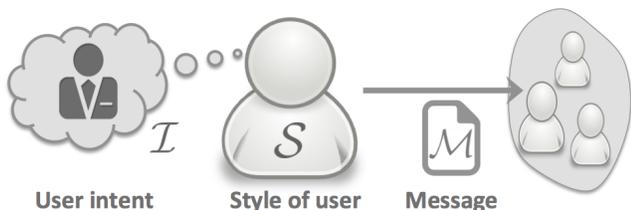
DOI: <http://dx.doi.org/10.1145/2908131.2908161>

documented the many reasons why people choose to participate on social networks such as communicating real-world friends or making new contacts [23, 20], connecting with colleagues and building professional relationships [10], and connecting with users that act as information providers [19]. While some social networks cater to populations with relatively homogeneous intentions (*e.g.*, online dating sites such as OkCupid focus on individuals interested in meeting others), the largest social networks, such as Facebook and Twitter, support a diverse population of users [5, 21, 32] with a large variety of intentions. Some of these social networks has also large participation of social bots to create social ties or promoting an orchestrated campaigns or advertisements [11, 29]. Such social networks are what we call *heterogeneous-intent networks*.

While there is no single unified theory of the relationship between an individual's intent and their communication behaviors, researchers in the fields of communication and media, political science, and related areas have demonstrated such behaviors do vary with a person's task or intent. For example, the emergence and evolution of social contracts studied in the evolutionary game-theory framework, where individuals choose messages considering their benefits and the state of the environment [27]. Thomas Schelling studied emergence of macro-behaviors through interactions among agents with micro-motives [26]. Other studies on such real-life interactions have also focused on social-contracts and diffusion of behaviors [2, 31]. However, to our knowledge, how such heterogeneity of intent manifests in individual communication behaviors within social media, and how such heterogeneous intents affect information diffusion in a social network has not been deeply investigated.

In this paper, we present a first study of the basic properties of communication in a heterogeneous-intent network and their implications. First, through both user studies and large-scale data analyses, we validate that individuals' specific intents affect their language style and, in turn, influence the specific messages individuals decide to send to their neighbors in the network (Figure 1). We find that, in the context of text messages in a social network (*e.g.*, status updates, tweets), language style does vary with intent, and that individuals with different styles do choose to send or forward different messages to their neighbors.

Secondly, we propose a simple structural model of information diffusion within a heterogeneous-intent network and present the results of experiments exploring its emergent properties. We find that, not only are the resultant diffusion trees broadly consistent with previous models and



**Figure 1: Schema of user communication.** A user’s intent affects the choice of style and the selected style influences the content of the message.

empirical observations, they are also able to reproduce empirically observed properties of diffusion trees that have not been captured by past models.

We address the following specific research questions in this paper:

**RQ1: Are different intents better fulfilled by some language styles rather than others?**

**RQ2: Do people with different intents select different language styles?**

**RQ3: Can we observe signs of heterogeneous selection of language styles in social network data at scale?**

**RQ4: Can we model the implications of the micro-processes of intent- influenced communication on macro scale information diffusion?**

## 1.1 Contributions

Through user studies, we establish the relationship between the intent of a message sender and the language style of messages, finding that not only are certain message styles more effective for certain intents, but also that people do strategically adapt or optimize their language to suit a specific intent. (§3) We perform two sets of user studies with crowd-sourced workers: (i) To answer RQ1, we validate that different message styles (such as emotional or logical argument styles) indeed have different affects on recipients (i.e., they fulfill different intents). And, (ii) to answer RQ2, we further analyze the relationship between message style and various intents of senders (such as having an intent to persuade a recipient, or an intent to appear likable or authoritative about a topic). We asked users to craft messages to fulfill a particular intent and evaluated their style preferences in the messages to study relation between style and a given intent.

To answer RQ3, we analyze Twitter message cascades and find that how a message propagation in a network is correlated with a historical representation of participating users’ message styles (§4). To understand how strategic local behavior may affect information diffusion through a network of individuals with heterogeneous intents, we perform a large-scale analysis of message cascades in Twitter. We use historical messages of individuals and language use as a representation of their styles and compare the styles of individuals who participate (or do not participate) in message cascades on Twitter. This way we explore how the style of the user relates messages and their participation to the cascades. We find that an individuals’ participation likelihood is correlated with the similarity between the individual’s style and the cascade message’s style,

even after conditioning on confounding factors, such as the distance from the cascade root. In other words, the diffusion of a message through a social network is related to how well the message matches the intents of individual users.

Finally, towards RQ4, we posit a simple analytical model of information diffusion in social networks that accounts for heterogeneous intents and find that it explains empirically observed properties of structural virality that are not explained by current models (§5 and §5.1). Based on the findings of our user studies and data analyses, we develop a structural model which assigns a  $d$ -dimensional vector representation of user intent to each node, and similarly assigns a vector in the same latent space to every message. Then, propagation of messages by a node is simply controlled by the distance between the message and the user intent vector of the node. We study information cascades generated with the above model under various configuration of parameters, and show that the resultant diffusion trees capture properties of structural virality that have been empirically observed in earlier work, but not explained by previous models.

## 2. RELATED WORK

Our work relates to several areas of literature. First, it relates to research characterizing people’s motivations for joining social networks and using it to achieve their goals. One of the most common motivations uncovered in these studies is the desire to re-establish ties with old friends and sustain existing relationships [16, 23, 20, 5, 28]. Closely related to this is the motivation to use social networks as a communication and coordination mechanism among friends [23]. Studies have found the creation of new relationships to be an important goal for some users, and indeed specialized social networking sites (dating sites) exist solely for that purpose [23, 5]. In the context of professional social networks, individuals use social networking to maintain relationships, advance their careers and advocate for their projects [10]. A significant use case for social networks is also information dissemination, and many people are motivated to be seen as experts or authorities for trusted information, while others are more interested in receiving information [19, 22]. Our work adds to this literature through quantified analysis of the interplay between the motivational and participatory intent of users and their communication behaviors.

To reach and interact with such a diverse population with heterogeneous intentions, users adopt their behavior and language according to their audiences and subjects of discussion. There have been several studies demonstrating how people adopt different language styles consciously or unconsciously to suit a specific situation. Examples of these effects observed on language mimicry in the context of power differentials between discussants [9] and prediction of message popularity [30]. In social psychology, there has been a large body of work on persuasion and social influence [6, 8, 33] that talks about various cognitive theories and psychological processes behind how people convince and persuade each other. In this work, we show that different styles do indeed have different effects in the context of a social network communications; and that people with different intents adopt different communication styles.

There has been a large body of work in the area of information diffusion through networks. Several early models for information diffusion were inspired from classical dis-

ease propagation models in epidemiology, such as SIR and SIS [3]. Other related diffusion models for product marketing included the Bass [4] model that is based on an S-shaped adoption curve. There has also been extensive work on modeling the adoption or spread of an idea, content or product in a social network. Well known classes of models in this domain include Threshold [17] and Cascade models [15], that specify how a node adopts a particular idea or product based on the adoption pattern prevalent in its neighborhood. Leskovec et al [24] proposed a stochastic model to study information diffusion in the context of product recommendation for ecommerce, and observed that cascade sizes follow a power-law distribution. Other papers [18, 1, 12] have also analyzed and characterized information propagation (URLs, topics, trends, etc) in blogs using well known models of information diffusion.

In recent work, Goel et. al. [13] proposes a formal measure, *structural virality*, of the degree to which a cascade reaches its audience through broadcast-like mechanisms vs. viral mechanisms. Using this measure, the authors conduct a large scale empirical study of a billion diffusion events for news, videos, images and petitions on Twitter, and observe a wide range of diverse cascading structures with varying structural virality, and show a low correlation between popularity and structural virality. Our work on modeling heterogeneous-intent networks introduce user specific properties that affect participation to the information cascades. In this model heterogeneity of user properties capture effect of micro-level interaction on macro-level cascades such as varying structural virality that have not been explained by previous models.

### 3. INTENT AND STYLE

In this section, we characterize the relationship between user intent, language styles, and the effectiveness of these styles at fulfilling different intents. We answer RQ1 and RQ2 through user studies that are carefully designed to enable us to assert the intent of a message author (by explicitly assigning them a specific task); quantify important aspects of message style (through crowdsourced judgments); and quantify the effect of messages (by questioning the subjects in our study).

#### 3.1 Methods

We want to explore the relationship between the intent of a user, the style in which they write, and the effectiveness of this style at achieving a given intent. Studies on user communication rely on observable variables such as messages or interactions among users. However, the motivation or intent underlying user behaviors is, in most contexts, hidden and unobservable directly. Crowd-sourced user studies allow us to assert this intent that would otherwise be invisible to us by explicitly assigning tasks to subjects. Of course, there are many possible intents that might underlie an individual’s participation in a social network, as well as many aspects of communication style that may be relevant to those many intents. In this study, we focus on 3 distinct and common intents of individuals and 2 aspects of language style in messages. The intents are:

1. The intent of being perceived as “likable”: a prerequisite perhaps to being sociable and friendly.

2. The intent of being perceived as an “authority”: important for people interested in gaining a positive reputation as a topical expert, etc.
3. The intent to persuade others: important for people interested in marketing, influence, and advocacy for a cause.

The two aspects of message style we measure are the use of logical and emotional tone in the text of a written message.

To provide a common ground for our analyses of the combinations of intent and style, we focus on messages where individuals express their opinion about some topic. For example, we ask a study participant to honestly write their opinion about a topic, at the same time asking them to write the message so that a recipient thinks that they are likable; or, alternatively, we deliver such an opinion message to a participant, and ask them to rate the perceived likability of its author. We perform these experiments for opinions about selected topics: products (Hyundai or Coca Cola); health & social issues (organ donation or organic food); and politicians (Barack Obama or George W. Bush) to analyze wide range of application domains.

To answer RQ1, we first study how measurable message styles affect recipients perception of author. We also studied how the style of the messages changes the initial opinion of receivers after presenting messages with opposite sentiment. To simplify the human labeling task, we focus on two high-level message styles: 1) whether a message contains a logical appeal or argument, such as facts and other detailed information; and, 2) whether a message contains emotional content or arguments, such as direct mentions of emotions (*e.g.*, love, hate), or indirect mentions of things people care about (*e.g.*, family, celebrities, memories).

We manually selected a set of 200 messages for each experiment topics. We annotated these messages on Amazon Mechanical Turks in two dimensions, *i.e.*, *logical* or *not logical* and *emotional* or *not emotional*. Note that messages may be both emotional and logical; or neither emotional nor logical. Each message was presented to 15 judges. We then slotted a message in one of the four quadrants resulting from the  $\{\textit{logical}, \textit{not logical}\}$  and  $\{\textit{emotional}, \textit{not emotional}\}$  labels based on the decision of a majority of the judges. The inter-annotator agreement for this step was 0.75 showing a large agreement between the judges. In addition, we also asked additional 15 judges to label each message as containing either a positive or negative sentiment.

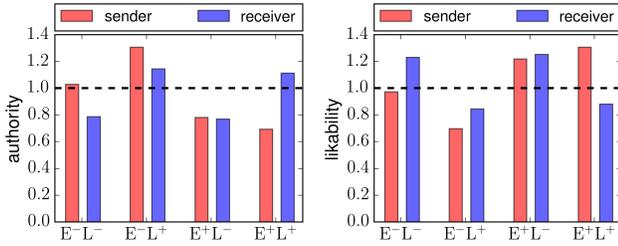
Analogous to the first experiments, we ran user studies to answer RQ2 by asking a set of 200 users to create messages on a particular topic with the goal of maximizing a specific effect. We created a “Human Intelligence Task” (HIT) on Amazon Mechanical Turk asking each judge to craft a message for a social platform to either appear likeable, authoritative, or to persuade friends to agree with her opinion. Table 1 shows a sample of the messages crafted by the sender and their subsequent categorization. Messages are then evaluated by 15 crowd-sourced workers based on different styles in the messages and how authors’ choice of message align with the most effective styles as measured in our first experiment on the perception of recipients.

#### 3.2 RQ1: Effects of Distinct Styles

First, we study how the logical and emotional content of messages affects recipients’ perception of authors. For

Message Text	Sentiment	Message Style
Hyundai cars just suck. Mine broke down right after their guarantee period ended.	Negative	Logical & Emotional
Hyundai cars are not giving good mileage compared to other car brands	Negative	Logical & Not Emotional
Hyundai should come up with some new designs for their cars	Negative	Not Logical & Emotional
Hyundai cars are unreliable and of cheap quality	Negative	Not Logical & Not Emotional
Hyundai offers a sweet, stylish ride for less money	Positive	Logical & Emotional
Hyundai cars are best comes with good mileage and pickup. Great value for money	Positive	Logical & Not Emotional
hyundai cars are very good to look and drive	Positive	Not Logical & Emotional
It looks good in design	Positive	Not Logical & Not Emotional

**Table 1: A sample of message texts written by study participants about *Hyundai Cars*, and labeled via crowd-sourcing.**

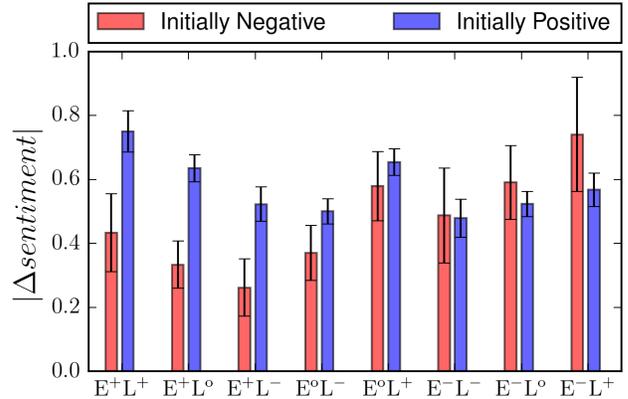


**Figure 2: Receiver columns show how perceived authority and likability vary with message style. Sender columns show how authors choose to write their messages to achieve a specified intent. Message styles are labeled as containing (+) or not containing (-) emotional (E) and logical (L) content.**

each topic we select 10 messages from randomly selected two quadrants of emotional or logical language styles and each evaluated by 15 crowd-source workers. Study participants are asked which set of messages are likely written by a more “likable” or “authoritative” person. In Figure 2, we present how language usage affects perceptions of likability and authority. Normalized values reflects the proportion of different message language use. Value greater than one indicates higher chance of choosing particular style to convey desired intent in the message. Logical messages are perceived to be written by more authoritative people while an absence of logical content in messages (esp. combined with emotional content) are more often associated with likable personalities.

Secondly, we aim to quantify the effect of a message’s emotional and logical components on its persuasive impact. We considered users with different initial positive or negative opinions about a topic. For this purpose, we carried out another study in which the user was initially asked to state her opinion, positive or negative, on a given topic. Depending on opinion measured in five-point scale (strongly negative=0 to strongly positive=4), she was presented 10 messages of the opposite sentiment, *all* chosen from a quadrant which was chosen uniformly at random. After being shown the 10 messages, she was asked for her opinion again.

We analyze results of opinion shift to determine which message styles are more effective at persuading recipients to change initial opinions. We remind the reader that in each treatment, we presented a set of messages with identically labeled message styles, promoting the opposite sentiment to that held by the study participant. Figure 3 summarizes the change in the sentiment as a result of particular treatment. Clearly, there is a change in the sentiment after



**Figure 3: Average change in shift from a sentiment after a particular treatment. Treatments are symbolized as Logical (L) and Emotional (E), where  $x^+$  and  $x^-$  stands for having only positive and only negative instances, respectively, and  $x^o$  for mixed.**

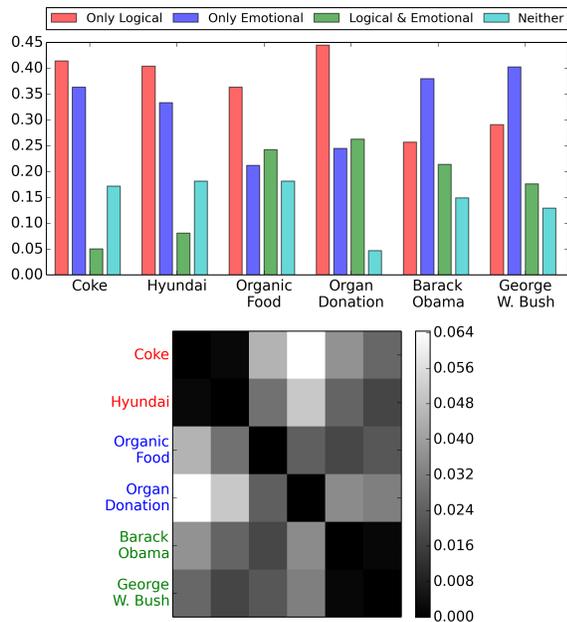
treatment. We observed that the persuasiveness of a message depends strongly on its style and the recipient’s initial sentiment about the topic. People having positive sentiment initially are more likely to be persuaded using emotional language. Similarly, logical arguments are more effective on users with negative initial sentiment.

Additionally, we computed the average shift *from* a prior sentiment which measures the overall drift away from that sentiment as a result of the treatment. For example, a change of 2 for a *strongly negative* sentiment implies that on the average, a user moved up two levels in her sentiment to around *positive*. We note that we also consider combinations of the quadrants in the set of treatments. For example, a treatment of *logical* implies all treatments that include *logical*, i.e., *logical* and *logical and emotional*.

These experiments demonstrate that whether messages use a logical and/or emotional style does indeed have different effects on recipients’ perceptions of the likability and authority of its authors as well as on the persuasive effect of the message. Next, we study whether message senders appear to strategically vary these styles in the messages they write.

### 3.3 RQ2: Adaptation of Style to Match Intent

In our second set of experiments, we ask people to write a message about a domain-specific topic with the specific purpose of either persuading or charming the recipient. We measure the degree of emotional and logical content chosen



**Figure 4: Message styles chosen by the senders are shown in histograms for each topics (a). Jensen-Shannon divergence of these distributions for topics shown as distance matrix and label colors indicates predefined categories (b). Similarities between topics are found to be consistent with categorical grouping.**

by the sender. In fact, on comparing the distribution of message attributes chosen by our senders, we find that, on average, senders do select the set of attributes that are best aligned for a specific domain and desired effect.

Figure 4(a) illustrates the distribution of message styles that the senders choose to send in respond to their intent. We observe that the senders showed a clear preference for a specific language style based on the topic, suggesting that the senders are indeed strategic in their choice of the message. Further, they exhibit a topic dependent behavior to create their persuasive messages. On products, emotional and logical are the only strategies that perform better. In the case of politics, emotional language plays a more dominant role in getting the message adopted while in the case of health & social issues, the sender benefits by selecting logical messages. We also cluster similarities between message style distributions by computing the Jensen-Shannon measure. In Figure 4(b) emerging clusters are consistent with the domain categorization.

## 4. OBSERVING HETEROGENEOUS INTENT AT SCALE

The previous section demonstrated, through controlled user studies, the relationship between a user’s intent and the language styles they choose to use. If we assume that social network participants act strategically, we would expect these people to select messages to propagate, reshare or retweet based on their own intents and corresponding language and content styles. That is, upon receiving a message, a person would be more likely to further propagate it

to their neighbors if its language and content matched the requirements of their own intents. Thus, we would expect the propagation path of a message through a set of users in a social network to be strongly influenced by the relationship between the message’s language and content and the users’ own intents.

### 4.1 Method

In this section, we test this hypothesis through analysis of a naturalistic, large-scale dataset consisting of 1 week of URL cascades, and supporting information, including the follower-follower network among users, and historical tweets by users who received or sent messages in the URL cascades. In our controlled user experiments, we are able to direct people to act in accordance with a specific intent. In observational studies on a large-scale dataset, however, we do not have the ability to directly measure (much less control) an individuals’ intent. Instead, we use the language of an individual’s past posts as a representation of their preferred language and content style that results from their intent.

We build a vector representation of the language and content style historically preferred by each of the users in our dataset. The experiments in this section analyze these “language vectors” in relation to cascade content and participation, and network neighborhoods. Our results show that, as measured by their language vectors, i) people participating in cascades are more similar to each other than to their neighbors who do not propagate messages; and ii) people participating in cascades are closer to the cascade message than people not participating. Furthermore, we report on the key topical dimensions of language vectors that most distinguish participants from non-participants in our studied cascades.

### 4.2 Data

To begin our data analysis of the relationship between individual intents and information dissemination in a social network, we collected URL cascades in the topics of *news* and *petitions* during the period of August 1-8, 2012. We focused our data collection on URLs related to 3 popular news websites and 3 popular petition web sites, shown in Table 2. Our data set consists of 1) 15,264 distinct URL cascades, including 217,600 messages sent by 121,726 users; 2) the follower/followee graph for users; 3) two weeks of historical tweets by all cascade participants; 4) two weeks of historical tweets of additional users who might have received a cascade message but who did not propagate it further. These tweets were collected via our organization’s access to an archive of the Twitter firehose.

We join these collected tweets with a high-fidelity crawl of the Twitter follower graph, collected through 2012, to recreate the propagation trees for our collected cascades. For each of tweet  $m$  in our dataset, we first identified the author and the users followed by the author (“followees”). If any of the author’s followees had already tweeted the URL contained in  $m$ , we declare the followee’s tweet to be the parent of  $m$ . If more than one of the author’s followees had previously tweeted the URL, we select the most recently received tweet as the parent of  $m$ . If none of the followees have previously tweeted the URL, we label  $m$  to be the root of a cascade tree.

Table 2 shows information about number of unique URLs for given domains and the distribution of cascade sizes in

	Topic	Site	$n_{url}$
1.	News	nytimes	6,030
		cnn	3,120
		usatoday	741
2.	Petition	twitition	2,411
		kickstarter	2,124
		change	838

**Table 2: Our study focused on cascades of URLs from news and petition web sites.**

our data set. In concordance with prior investigations of cascades on Twitter, we find that only a small percentage of cascades grow to a significant size [13, 7, 14].

Accordingly, for each of the users in our dataset, whether they participated in a cascade or simply received a message from the cascade without propagating it further, we collected two weeks of historical tweets. Non-participant users randomly selected from the friends of participating users—this way we can assume they observe cascade tweets from their friends and choose not to participate. To avoid confounding our analysis with messages that are directly related to messages in our 1 week sample, we collected our historical tweets from the period of June 17-30, 2012 (ending a full month before the collection of our cascade data), such that any language/content overlap between the historical behavior of users and our collected cascades will reflect primarily language style and broader topical interests. Note that we make the assumption that individuals’ primary intents underlying their participation in the social network are stable across the collection periods.

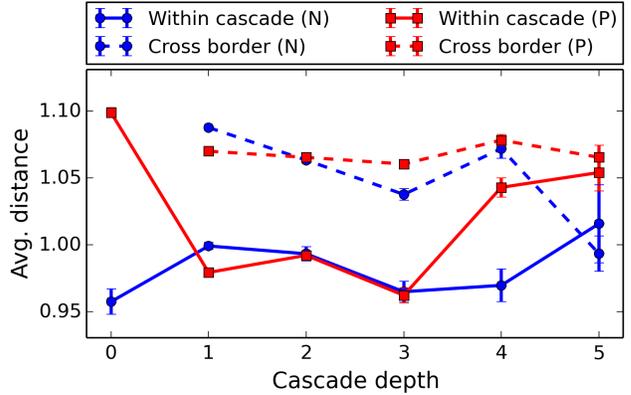
To convert historical tweets into a model of the language and content style preferences of users, we apply an LDA model of Twitter content that maps the words in tweets to a space of 2,610 learned topics. The training of the LDA model is described in detail in Ramage, Dumais and Liebling [25]. Briefly, it is a partially supervised (Labeled LDA) model, trained on approximately eight millions tweets collected during June, 2011. The top 200 most frequent topics were manually inspected and given descriptive names.

We apply this model to map every historical tweet to a 2,610-dimensional vector. We discard all users with fewer than 10 historical tweets. For the remaining users, we build a user model as the average of the vectors of their historical tweets.

### 4.3 RQ3: Heterogeneous message selection

Our goal is to determine whether our models of user language and content style capture any difference between people who choose to participate in a cascade, and those who do not.

Figure 5 presents the average distances among pairs of people who both participated in the same cascade, and between pairs of people where one person participated in a given cascade and one received the message but did not participate. We see that in total, distances between cascade participants and non-participants (dashed lines) is 8.7% greater than the distances among pairs of cascade participants (solid lines) in the news cascades and 7.3% greater in the petition cascades. These differences are statistically significant at depths 0 through 4. This trend, wherein co-participants’s language vectors are nearer each other than



**Figure 5: Pair-wise distances among cascade participants (solid lines) and pair-wise distances between cascade participants and non-participants (dashed lines), conditioned on depth in the propagation tree, for both news (N) and petition (P) data sets**

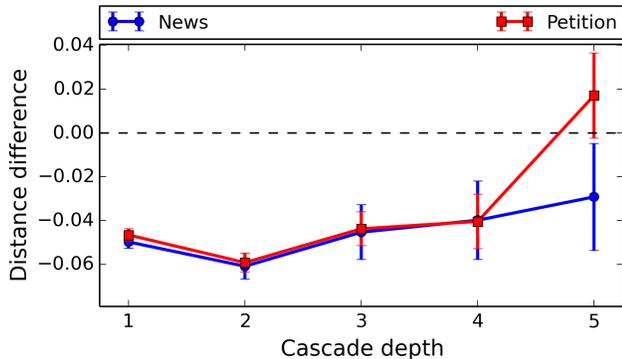
to non-participants holds on average, as well as when conditioned on cascade depth. From this, it seems that users participating in a cascade are more likely to share a similar language and content style, as compared to their neighbors who do not participate in the cascade.

Now that we see that there is a link between the vector representation of people’s language and content styles and their participation in a cascade, we want to see if such a relationship may also holds between a user’s language vector and the language vector of the cascade messages. A second measurement, shown in Figure 6, measures the distance between people who received a cascade message and the vector representation of the root message of the cascade. We find that the people who received the message but chose not to propagate it are further, on average, from the root message of the cascade vector than the people who received the message and chose to propagate it. This result tells us that the similarity of a message’s vector to a user’s vector is correlated with the user’s likelihood of propagating the message.

## 5. HETEROGENEOUS-INTENT NETWORK MODEL

In this section, we describe a simple model of the micro-level behaviors of senders and receivers that captures the observations learned from the experiments in Sections 3 and 4. Most notably, while most previous models of information dissemination assume that all nodes in a social network are equivalent, our model incorporates information about the heterogeneity of users into the dissemination of information. Interestingly, while many previous models incorporate a stochastic element to determine whether a node transmits a message to its neighbors, our nodes transmit messages in a deterministic fashion, while maintaining a stochastic element during network initialization. Despite these differences, our model is able to simulate realistic properties of diffusion cascades that are captured by previous models, and also recreate properties of structural virality that have not yet been explained by previous models.

We represent a message as a vector in  $d$  dimensional space, where every dimension corresponds to an attribute of the



**Figure 6:** For both news (blue) and petition (red) cascades, we show the difference between participant and non-participant distances to the root message vector, conditioned on depth in the propagation tree. Values less than zero mean that participants are closer to the cascade root than non-participants.

message such as logic, emotion etc. For simplicity, we assume that the range for each dimension lies in  $[0, 1]$ . Hence every message  $m$  is a vector in  $[0, 1]^d$

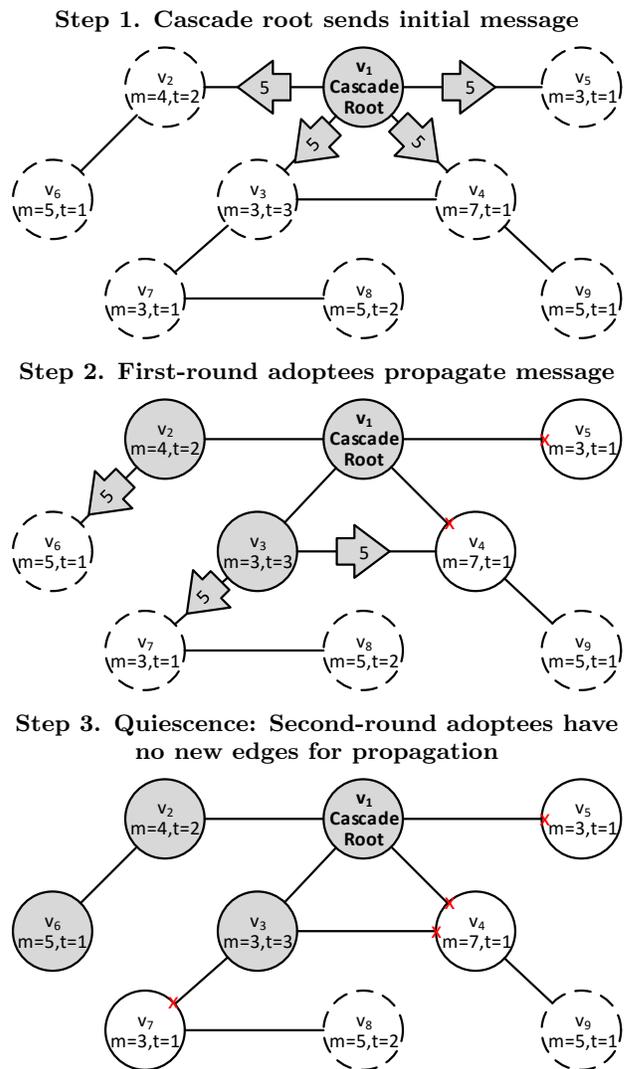
We consider an online social network graph  $G = (V, E)$  with nodes  $v_1, v_2, \dots, v_n \in V$ . For ease of notation, we will frequently interchange node  $v_i$  and index  $i$ . The vertices  $v_i$  correspond to individuals, and edges  $E = \{e_{ij}\}$ , denote social ties between the individuals. Every vertex  $v_i$  has a  $d$ -dimensional unit-length style vector  $m_i \in R^d$  and an acceptance threshold  $t_i \in R$ . Vertices in the graph can propagate messages along their edges to their neighbors. For a node  $v_i$ , its set of neighbors is denoted by  $N(i) = \{j : e_{ij} \in E\}$ , and its degree is denoted by  $d_i = |N_i|$ . We denote the total number of edges in the graph as  $|E|$ .

The dynamics of message propagation in the graph works as follows: Every vertex in the graph can act as a sender or a receiver of messages. A sender in the graph can compose a  $d$ -dimensional message vector  $m$  and broadcasts it to its neighboring vertices. A receiving vertex accepts an incoming message only if the incoming message is “close” (in some  $L_p$  norm) to its style vector. In particular, whenever a (receiver) vertex  $v_r$  receives a message  $m$  that it has not previously adopted, it adopts the message and propagates the same message to all other neighbors in  $N(r)$  iff  $\|m - m_r\|_p < t_r$ . Otherwise it drops the message, and does not adopt it. We typically assume  $p$  to be 1 (Manhattan distance) or 2 (Euclidean distance). Figure 7 illustrates an example of message propagation in a graph based on this model.

Note that unlike in the case of traditional information diffusion models such as Linear Threshold and Independent Cascade models, the adoption of a message composed by a sender node depends not only on the graph structure but also on the message vector itself. A natural question is then to understand what message a given sender should compose in order to persuade as many nodes in the graph to adopt the message. We leave this as an important area for future work.

## 5.1 Simulation Details

The purpose of our simulations is to validate that our model generates cascades that are consistent with past mod-



**Figure 7:** Example propagation of a message in our model. Each vertex  $v_i$  has 1-dimensional message vectors  $m_i$  and propagation threshold  $t_i$ . The filled vertices are nodes that have adopted the message at each step, unfilled vertices are nodes that have not adopted the message, and dashed vertices have not received the message.

els and, more importantly, with observations of naturalistic cascade events. We first present details of our simulation setup. Then, we characterize the relationships between key parameters of the model and the *virality rate* of information cascades. Finally, we study the *structural virality* of information cascades generated within our model, and find that our model captures properties of structural virality that have been empirically observed but not captured by previous models [13].

Using the model presented in previous section, our simulation tracks the propagation of a message represented as a  $d$ -dimensional vector through a social network. Given the network of individuals, each associated with its own  $d$ -dimensional style vector and threshold  $t$ , and an initial

starting node, a message’s propagation is deterministic.

In each of our simulation experiments, we build a 1 million node network and execute 100k information cascades. Our simulation model is governed by 2 parameters that control the creation of the network structure and the propagation of messages:  $d$  specifies the dimensionality of the latent message space and style vectors; and  $t$  the maximum threshold acceptance distance for re-broadcasting messages.

There are two key stages to building and running an information dissemination model. The first stage is the initialization of the network of connections among nodes. The nodes of the network are each randomly assigned a desired message vector and a random threshold smaller than  $t$ . The network is initialized as an asymmetric network. Every node’s in-degree is assigned according to a power law distribution parameterized by  $\alpha = 2.3$ .

The association of a style vector to nodes allows us to experiment with peer influence and homophily within our model. To do so, we add two additional parameters to control the network initialization. To represent the effect of homophily—that an individual is more likely to create a social tie to a user similar to herself, we replace the random selection of a neighbor with a preferential selection, where the closest neighbor in the  $d$ -space is selected from a random candidate pool of size  $n$ , and the candidate pool is fully replaced for each selection. While the distribution of outbound degrees is not directly controlled, this process does generate a heavy-tailed distribution.

To represent peer influence,  $s$  controls the smoothing of style vectors between a vertex in the graph and its neighbors. More formally, each node’s style vector  $m_i \leftarrow (1 - s) * m_i + s * avg(N_i)$  where  $N_i$  is the set of vertices neighboring  $i$ .

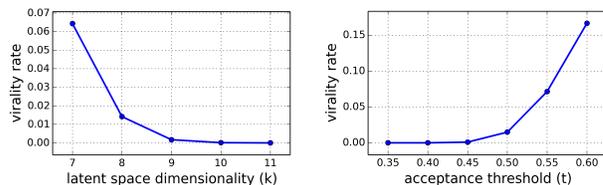
The second stage of the simulation—propagating a message through the nodes of this network—maps directly to the microbehavior of a sender and a recipient described by basic model. We represent each message as a vector in the same latent message space associated with the nodes in our network. Following the basic model, whether or not a node will re-broadcast a message it has received is determined by whether the message lies within the threshold distance of the node’s own desired message vector.

## 5.2 RQ4: Testing our Heterogeneous-Intent Network Model

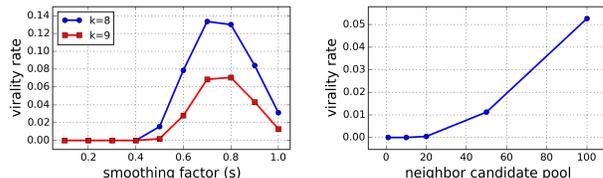
To determine whether or not our intent-influenced communication model correctly captures macro-scale information diffusion in heterogeneous-intent networks, we focus on two measures of information cascades: virality rates, and structural virality.

To compute virality rate, we simulated a million node network for 100k random initializations. We considered diffusion events with more than 100 adopter nodes as viral cascade. We define virality rate as the fraction of viral cascades among all simulated diffusions. Studying virality rates in our simulations, we find that for certain parameter configuration, namely,  $\alpha \approx 2.3$ ,  $t \approx 0.45$ ,  $s \approx 0.45$ , and  $8 \leq k \leq 10$ , our cascades fit previously empirically observed data. For example, we find that, consistent with past empirical observations, most cascades remain very small, and the rate of large-scale diffusion events (i.e., events reaching greater than 100 nodes) is roughly 1 in 1000 [13, 14].

Exploring the relationship between the four main parameters of our model, we confirm that as the space dimension-



(a) Relationship between dimensionality  $k$  and virality rate.  $n = 0.5$ ,  $t = 0.5$   
(b) Relationship between acceptance threshold and virality rate.  $k = 8$ ,  $t = 0.5$



(c) Relationship between smoothing factor and virality rate.  $k = 8$ ,  $t = 0.5$   
(d) Relationship between candidate neighbor pool size and virality rate.  $k = 7$ ,  $t = 0.5$

**Figure 8: Relationship between virality rate and model parameters**

ality  $d$  factor for the latent space increases, the virality rate decreases (Figure 8(a)). The intuitive explanation is that as  $d$  increases, the latent space becomes more sparse and the likelihood of a message satisfying the style vector and threshold of a sufficient number of connected vertices decreases. The effect of increasing the acceptance threshold  $t$  or the candidate pool size  $n$  and trivially increases virality (Figure 8(b) and 8(d)).

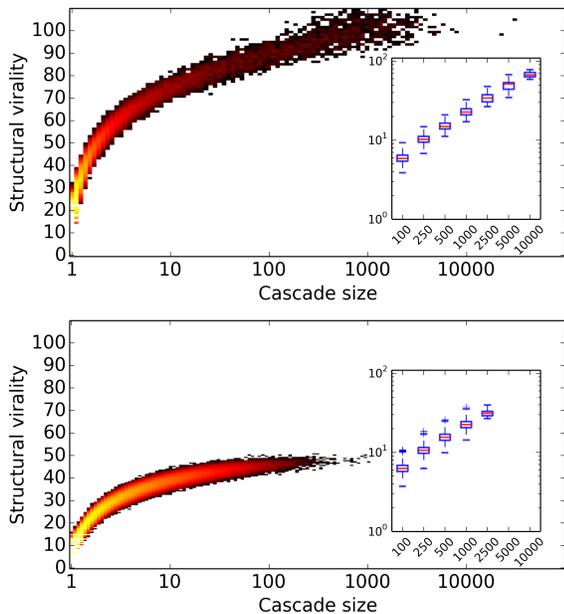
Structural virality is a measure [13] proposed to distinguish between information disseminations that occur primarily through broadcast mechanisms (one sender broadcasting a message to a large number of people, with relatively few or no independent decisions to rebroadcast or spread the message); and propagations where no one sender is responsible for most of the dissemination (i.e., multigenerational “viral” propagation). Specifically, structural virality  $v(T)$  over a diffusion tree  $T$  is defined to be the average distance between all pairs of nodes:

$$v(T) = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n d_{ij}$$

where distance  $d_{ij}$  is the shortest path length through the tree  $T$  between nodes  $i$  and  $j$ .

One of the elements of empirically observed information cascades that previous models have not captured is the variance in structural virality observed in information cascades. In their large-scale study of information diffusion events in Twitter, Goel et al. find that structural virality is weakly correlated (0.36) with the size of a cascade—smaller cascades are more likely to have a broadcast-like spread, and larger cascades are more likely to have a viral-like spread. However, as the correlation is weak, there are many large cascades that have a broadcast-like spread, and many small cascades that have a viral-like spread.

Figure 9 plots on a log-scale the structural virality of information cascades, binned by the size of the cascade under two information diffusion modes. Figure 9(b) shows structural



**Figure 9: Structural virality on different values of cascade sizes represented as heatmap and boxplot as an inset. Structural virality (a) under a persuasion model.  $k = 7, s = 0.45, t = 0.45$  and (b) simple contagion model with uniform infection probability.  $r = 0.7$**

virality under a simple contagion model, where the likelihood that any given node will propagate a message is  $\beta$ . We select  $\beta$  such that  $r = \beta \bar{d}$ ; where  $r = 0.7$ , and  $\bar{d}$  is the average degree of the network. As previously demonstrated in [13], this simple contagion model generates cascades unlike empirically observed cascades in that they are less likely to have a highly viral structure and, as cascade sizes increase, the variance of their structural virality decreases significantly.

In contrast, Figure 9(a) shows structural virality under our model of persuasion affects. Overall, we see that structural virality ranges from broadcast-like (low values) to very viral (high values). Within each bin of cascades, we also see a high variance in structural virality, successfully replicating the high variance in structural virality empirically observed in [13].

## 6. IMPLICATIONS AND FUTURE WORK

This first study of heterogeneous-intent networks has potentially broad implications for the study and understanding of communication behaviors and information dissemination in social networks; as well as the design of social media systems.

Explicit modeling of the heterogeneity of intents in a social network enables asking new questions, such as when and why the same message can lead to widely different cascades when initialized in different network locations. There are also potential implications for many tasks related to the analysis of information cascades, such as the early prediction of their size to maximizing their spread [14, 24, 18, 1], with broad applications motivated by a diverse set of public, social good and commercial purposes.

Of course, aspects of our proposed model, especially how

the intents of users and their network connections are initialized, require further exploration in future work. Similarly, validating the findings of our user studies and data analyses across additional varieties of intents and communication styles, in more controlled and/or diverse user populations, and across larger studies and data sets is also important future work. Incorporating additional factors into our model, such as peer influence, also remains for future work.

Better understanding of people’s intents and the behaviors they exhibit towards fulfilling them can lead to improved methods of helping people achieve their goals. Better understanding of how people learn what language styles are most effective may lead to methods for helping people learn to be effective more rapidly and easily. There may also be implications for how people adapt and learn to optimize their behaviors, and the influence of social network features on that learning with concordant implications for their individual success and viral spread of ideas and emergent community behaviors in other settings as well, both online and offline.

## 7. CONCLUSIONS

Individual’s motivations for participating in social networks can vary greatly, and this heterogeneity is reflected in the wide-variety of behaviors and language styles we see in large social networks. Through a mixed methods approach, incorporating user studies as well as large-scale data analyses, modeling and simulation, we present a first study of the implications of these heterogeneous intents. We first validate the interplay between a user’s intent and their language style, and their effectiveness in fulfilling their intents. Secondly, we propose a simple structural model of information diffusion within a heterogeneous-intent network and present the results of experiments exploring its emergent properties, finding that it recreates aspects of empirically observed cascades that are not captured in previous models.

Better understanding and recognition of the importance of the heterogeneity inherent in large-scale social networks—especially the heterogeneity of the underlying motivation and intent of the individual social network participants—has the potential to lead to significant improvements in our understanding of individual behaviors and emergent community behaviors, and our ability to understand and direct network phenomenon of social importance.

## 8. REFERENCES

- [1] E. Adar and L. A. Adamic. Tracking information epidemics in blogspace. In *Proc. of international conference on Web Intelligence*, pages 207–214. IEEE, 2005.
- [2] A. Alshamsi, F. Pianesi, B. Lepri, A. Pentland, and I. Rahwan. Beyond contagion: Reality mining reveals complex patterns of social influence. *PloS one*, 10(8):e0135740, 2015.
- [3] N. T. Bailey. *The mathematical theory of infectious diseases and its applications*. Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE., 1975.
- [4] F. M. Bass. A new product growth for model consumer durables. *Management Science*, 15(5), 1969.
- [5] D. Boyd and N. Ellison. Social network sites: Definition, history, and scholarship. *Journal of*

- Computer-Mediated Communication*, 13(1):210–230, 2007.
- [6] S. Chaiken, W. Wood, and A. H. Eagly. Principles of persuasion. 1996.
- [7] J. Cheng, L. Adamic, P. A. Dow, J. M. Kleinberg, and J. Leskovec. Can cascades be predicted? In *Proc. of the 23rd international conference on World wide web*, pages 925–936. International World Wide Web Conferences Steering Committee, 2014.
- [8] R. B. Cialdini. Influence: The psychology of persuasion. 1993.
- [9] C. Danescu-Niculescu-Mizil, L. Lee, B. Pang, and J. Kleinberg. Echoes of power: Language effects and power differences in social interaction. In *Proc. of the 21st international conference on World Wide Web*, pages 699–708. ACM, 2012.
- [10] J. DiMicco, D. R. Millen, W. Geyer, C. Dugan, B. Brownholtz, and M. Muller. Motivations for social networking at work. In *Proc. of the ACM conference on Computer supported cooperative work*, pages 711–720. ACM, 2008.
- [11] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini. The rise of social bots. *arXiv preprint arXiv:1407.5225*, 2014.
- [12] E. Ferrara, O. Varol, F. Menczer, and A. Flammini. Traveling trends: social butterflies or frequent fliers? In *Proceedings of the first ACM conference on Online social networks*, pages 213–222. ACM, 2013.
- [13] S. Goel, A. Anderson, J. Hofman, and D. J. Watts. The structural virality of online diffusion. In *Preprint*, 2013.
- [14] S. Goel, D. J. Watts, and D. G. Goldstein. The structure of online diffusion networks. In *Proc. of the 13th Electronic Commerce*, pages 623–638. ACM, 2012.
- [15] J. Goldenberg, B. Libai, and E. Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing letters*, 12(3):211–223, 2001.
- [16] B. Gonçalves, N. Perra, and A. Vespignani. Modeling users’ activity on twitter networks: Validation of dunbar’s number. *PLoS one*, 6(8):e22656, 2011.
- [17] M. Granovetter. Threshold models of collective behavior. *American journal of sociology*, pages 1420–1443, 1978.
- [18] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins. Information diffusion through blogspace. In *Proc. of the 13th international conference on World Wide Web*, pages 491–501. ACM, 2004.
- [19] A. Java, X. Song, T. Finin, and B. Tseng. Why we twitter: understanding microblogging usage and communities. In *Proc. of the 9th WebKDD and 1st SNA-KDD workshop on Web mining and social network analysis*, pages 56–65. ACM, 2007.
- [20] A. N. Joinson. Looking at, looking up or keeping up with people?: motives and use of facebook. In *Proc. of the SIGCHI conference on Human Factors in Computing Systems*, pages 1027–1036. ACM, 2008.
- [21] J. Kulshrestha, M. B. Zafar, L. E. Noboa, K. P. Gummadi, and S. Ghosh. Characterizing information diets of social media users. In *Ninth International AAAI Conference on Web and Social Media*, 2015.
- [22] H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? In *Proc. of the 19th international conference on World wide web*, pages 591–600. ACM, 2010.
- [23] C. Lampe, N. Ellison, and C. Steinfield. A face (book) in the crowd: Social searching vs. social browsing. In *Proc. of the 20th anniversary conference on Computer supported cooperative work*, pages 167–170. ACM, 2006.
- [24] J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. *ACM Trans. Web*, 1(1), May 2007.
- [25] D. Ramage, S. T. Dumais, and D. J. Liebling. Characterizing microblogs with topic models. *ICWSM*, 10:1–1, 2010.
- [26] T. C. Schelling. *Micromotives and macrobehavior*. WW Norton & Company, 2006.
- [27] B. Skyrms. *Evolution of the social contract*. Cambridge University Press, 2014.
- [28] T. Spiliotopoulos and I. Oakley. Understanding motivations for facebook use: Usage metrics, network structure, and privacy. In *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3287–3296. ACM, 2013.
- [29] V. Subrahmanian, A. Azaria, S. Durst, V. Kagan, A. Galstyan, K. Lerman, L. Zhu, E. Ferrara, A. Flammini, F. Menczer, et al. The darpa twitter bot challenge. *arXiv preprint arXiv:1601.05140*, 2016.
- [30] C. Tan, L. Lee, and B. Pang. The effect of wording on message propagation: Topic- and author-controlled natural experiments on twitter. In *Proc. of the 52nd annual meeting of the Association for Computational Linguistics*. ACL, 2014.
- [31] A. Vinciarelli, M. Pantic, H. Bourlard, and A. Pentland. Social signals, their function, and automatic analysis: a survey. In *Proceedings of the 10th international conference on Multimodal interfaces*, pages 61–68. ACM, 2008.
- [32] C. Wagner, V. Liao, P. Pirolli, L. Nelson, and M. Strohmaier. It’s not in their tweets: Modeling topical expertise of twitter users. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Conferenece on Social Computing (SocialCom)*, pages 91–100. IEEE, 2012.
- [33] W. Wood. Attitude change: Persuasion and social influence. *Annual review of psychology*, 51(1):539–570, 2000.