ICM 49144

Team #49144

February 2, 2016

Abstract

The goal of an abstract is to give an overall description of the paper. We should illustrate these points:

1)A little background of the problem.
2)Our basic model
3)Our new model, namely what is the improvement of this model compared with those previous ones. Some descriptions about the weakness of previous models(which are what we are trying to solve).
4)The good results we get from our new model.

In general, the abstract must illustrate what we have done in this paper and the significance of these improvements(usually by showing those good results we have).

Understanding the process of information getting around in society network has attracted increasing attention. Information flow in complex society's network is of paramount importance when trying to solve the problem of the spreading of news, whether for analyzing historic events or for future prediction. In the light of the development of social media and social ties, and based on the previous raised models, we bring up a model here so as to represent the society's information spreading process in the network, as well as the historical trend of information network's development.

The requirement is to build a model to simulate the , with which we have to obtain the following results: build a virtual network to model the spreading of information, and validate it with data from past and now; use it for prediction in year 2050; use the controllable and deterministic factors of information, network topology, links and nodes, to influence and predict the spreading speed and the number of influenced individuals. With more taken into consideration, we obtained more freedom in the model, by far improved the simulation result of real world event, and also coincide with the historical development of social network.

1 Introduction

News is packaged information about current events happening somewhere else[1]. It flows from individual to individual, from region to region. News could move through many different media, based on word of mouth, printing, broadcasting, and electronic communication. Thanks to the development of media and our tech-connected social network, the flow of information becomes such easy and wide-ranging as it is today. Meanwhile, the contents of news and people's bias have also changed hugely. Therefore, it is intriguing to investigate how social network structure and information value could influence the spread of information and public opinion. In this paper, we bring up a network model to simulate the information spreading process and its influence upon public interests, as well as predicting how the network will develop in the future.

We organize our paper in this following structure.

- **Introduction** includes background, restatement of problem and literature review.
- **Model** includes illustrations of two models. We first build a basic model based on previous studies, and then devise a new model incorporating our observations.
- **Result** includes mathematical formalization of each task , simulation result for each task and further explanations of some phenomena.
- **Sensitivity Analysis**
- **Further Discussion** includes what we could have improved and what may have potential meaning and worth detailed experiment.
- **Strengths and Weakness**

1.1 Restatement of Problem and Solutions

We are required to build a mathematical model to analyze the relationship between speed/flow of information vs inherent value of information. More specifically, we have five tasks; we categorize them into two sub-problems:

- Build a model to simulate the spread of information, and analyze the interaction between network structure and information flows.This includes (a),(d) and (e)
- Build a model to simulate the development of social network, and use data from the past to predict the future. This includes (b) and (c).

In this paper, we present a network model inspired by variant of SIR model [2]. We use nodes to represent individuals and edges to represent communication links between them. Given parameters, which depict structure of the network, we could simulate the spreading of news by updating network variables. Furthermore, we could change those parameters so as to mimic the development of our social network. This method solves first sub-problem satisfactory. As for the second sub-problem, we draw an analogy between the development of social network and news spreading procedure. We then derive some reasonable trending curves which accord with historical data. Moreover, our model could be used to explain diffusion of innovations and data explosion.

$liangdianliangdianliangdianliangdianliangdian$

1.2 literature review

In an very early stage, researchers believed that there are many similarities between spreading of news and spreading of epidemics. Thus, it is quite common to analyze spreading of news by an epidemic model, such as susceptible-infected-recovered (SIR) model. There are also studies based on advanced methodology, i.e. SIHR [3]. However, SIR is a analytical model; this lead to an inevitable problem that we can not take too many considerations into account. Therefore, more and more studies are focused on a cellular automata variant of SIR[2]. This new class of model would enable us to consider a much finer network structure.

In this cellular automata variant of SIR, each node has four states, namely 'UNKNOWN', 'KNOWN', 'APPROVED' and 'EXHAUSTED'. When a node receives a message from its neighbor, it transfers to KNOWN state and remains. Every time it receives this news, it has a probability to shift to APPROVED and spread message to its neighbors. Thereafter, this node will remain at EXHAUSTED and do nothing. Since this is a discrete-time network model, one can investigate further by manipulating network structure.

In recent years, researchers focus heavily on the effect of network structure. Liu et al. propose a dynamic network which could rewire its links during the procedure[4]. Ou et al. discuss the effect of heterogeneous connections upon news spreading[5]. Kwak et al. crawl real data from Twitter and build a network to simulate the reality[6].

However, despite the fact that there are lots of studies on different structure, few people have compared social network in different periods. In the light of previous works, we devise our new network model. On the one hand, our model takes many important factors into consideration, such as probabilistic states transition, memory effect, directed information flow, heterogeneous structure, news value and public bias, etc. On the other hand, we incorporate a few network structure parameters which would vary depending on different periods. Thus, we could simulate and predict the development of social network for decades.

2 Assumptions and Justifications

This is a very important section. In this part, you should give all the basic assumption under which you build your model. Every model has its applicable range, so you must figure out what are those assumption at the beginning of building a model. Those assumption should be inspired and concordant with our daily intuition. The form of assumption and justification should be more or less like this:

- **A sentence of assumption** A sentence of justification. Usually, this is a narrative of an intuition from which you derive your assumption. You should depict this intuition and its connection with the assumption.

3 Notations

All the variables and constants used in this paper are listed in **Table 1** and **Table 2**.

Table 1 Symbol Table–Parameters

Symbol	Definition	Type
	Parameters	
λ	The probability a node will APPROVED when it first hears the news	Local
λ_1	The inherent value of the news	Global
b	The degree of trust in a community (network)	Global
c_i	The authority of the i^{th}node	Local
w_{ij}	The weight of edge e_{ij}	Local
t	Time-step	Global
v	The spreading speed of news from media	Global
α	The step-size of feedback enhancement	Global
β	The step-size of decay	Global
a_M	The likelihood an individual get information from media	Local
a_F	The likelihood an individual get information from famous people	Local
a_N	The likelihood an individual get information from normal people	Local

Table 2 Symbol Table–Variables

Symbol	Definition	Type
	Variables	
$m_i(t)$	The number of times the i^{th} node has heard the news until time t The confidence of i^{th} node towards the news until time t	Local
UNKNOWN	The node has not yet heard the news, waiting for news	Local
KNOWN	The node has heard the news at least once and has a probability to approve	Local
APPROVED	The node will spread the news to its neighbor and then shift to UNCONCERNED state	Local
UNCONCERNED	The node will not participate in any procedure	Local
λ_2	An individual's initial bias towards a certain piece of news	Local
T	Response variable, the final time the network takes to be steady	Global
n	Response variable, the total number of APPROVED individuals	Global
u	Response variable, the speed of information getting around defined by $u = \frac{n}{T}$	Global
$\Delta\lambda_2$	Response variable, the fluctuation of λ_2	Local

4 Model

4.1 Basic Model

4.1.1 Overview

This is a fundamental model which will give us some insight into the network. Inspired by [2], we use N nodes to simulate N identical users in reality and each node stochastically links to other k nodes. When a given event occur at a node in this graph, we use a dynamic probabilistic information propagation model (DPIP), discussed in details as follows, to simulate the spreading of informations. This model only depicts the case that all users are the same, with no existence of celebrities or huge media company.

$Ourmodelconsistingoffourstates, describethetransferringofstatestomakeitclear$!!!!$ieafterKnon$

4.1.2 Assumption and Justification

- **All nodes are identical** In this model, we do not consider celebrities or huge media company, therefore, all nodes equally important in the sense that they all have k edges connecting with them.
- **Each node has four states** Since we use a node to represent a person, a node has four states corresponding with a person has four attitudes

towards an information, UNKNOWN, KNOWN, APPROVED and UNCONCERNED.

- **Transferring from KNOWN to APPROVED is a probabilistic event** Once a node heard a piece of news from its neighbor, it has a probability to transfer from KNOWN to APPROVED. The more times it hears the same news, the higher probability it has to approve this news.

4.1.3 Methodology

Our model is a time-discrete network model. We have a set of laws which update the parameters of the model every time-step. Those laws are based on previous assumptions and observations. In this part, we formalize those laws in a mathematical way, and explain how they could update parameters from time t to time $t+1$.

Step 1: **Initializing the network**

(a) Randomly constructing the network

Our network is a randomly connecting network. It has N nodes. Then we randomly connecting nodes with edges and make sure that, for each node, there is always k edges connecting with it.

(b) Initializing the states

To begin with, every nodes should be UNKNOWN to a piece of news except the node which originates this news. Therefore, we set every nodes in our network to the UNKNOWN state.

(c) Generating news

We randomly choose a node to be the news producer, transferring its state from UNKNOWN to APPROVED. This indicates that this node approves this news and is willing to tell its neighbors. The other nodes remain in the UNKNOWN state.

Step 2: **Propagating information**

(a) Spreading news

We traverse all nodes and find out those nodes with state APPROVED. For each APPROVED node, it is willing to spreading this news to its neighbors. If the neighbor is UNKNOWN about this news, it shift to KNOWN states. If the neighbor is already in KNOWN state but not APPROVED state, it will memorize this event and has higher probability to transfer to APPROVED state in later time.

Formalizing this step mathematically, we denote $m_i(t)$ to be how many times the i^{th} node has heard this news from its neighbor until time t. For obvious reason, this $m_i(t)$ could also be explained as belief. For each APPROVED node A in our network, we manipulate the states of all its neighbors. Suppose node B is one of its neighbor. If B is in the state UNKNOWN, we set B to the state KNOWN and let

$$m_i(t+1) = 1.$$

If B is already in the state KNOWN, we remain it at KNOWN and let

$$m_i(t+1) = m_i(t+1) + 1.(1)$$

If B is in the state UNCONCERNED, we do not make any change to B. After doing this to all neighbors of A, we change its state to UNCONCERNED.

(b) Transferring from KNOWN to APPROVED

Based on our observation, once an individual hears a piece of news, he will choose to believe in or not depending on how many times he has heard it. The more times it is, the more willingly he is to trust it. Therefore, we use a probability model to describe this phenomenon.

For each KNOWN node, if it hears the news from its neighbor at time t, it has a probability to approve this news. Denoting $P(m)$ as the probability of a node transferring from KNOWN to APPROVED, we have

$$P(m) = (\lambda - 1)e^{-b(m-1)} + 1,$$

where m is $m_i(t)$ and λ, b are network parameters.Once a node transfers to APPROVED, it will spread the news at next time-step, namely, at time $t+1$. We could also observe that the larger m is, the larger $P(m)$ is; this is in accordance with our previous observation.

We need to mention that if a node does not hear news from its neighbors at time t, it will definitely not transfer to APPROVED, even though it may have heard the news once before $t-1$. This is in accordance with our intuition that one will not abruptly change his mind as long as he does not hear anything new.

(c) Iterating

In previous steps, we have updated states and $m_i(t+1)$ for each nodes. By iterating those previous two steps, we could simulate the propagation of information.

4.2 Improved Model

4.2.1 Overview

Our new model is a hierarchy, 3-layered, directed, weighted network, with belief cumulated and probabilistic transferring mechanism.

4.2.2 Modifications

In this section, we compare our new model with the previous one and illustrate the modifications explicitly.

- **Constructing the network as a directed graph.**In the basic model, we only construct a undirected graph; information could flow either from A to B or B to A. However, to consider a finer structure, we need to use a directed graph. Followers on Twitter and receivers of media are both good examples to support our choice.

- **The effect of celebrities and mass media.** In the basic model, we treat every individual equally. Nevertheless, in reality, there are celebrities and mass media which are more significant to the social network, i.e. a person may have a large number of followers on Twitter and hence are more influential. In this new model, we allow a fraction of nodes to represent those significant celebrities and one node to represent mass media.
  As for the edge structure, there are a huge amount of edges directed from each celebrity node to the normal users, meaning that they have lots of followers and news could flow from them to lots of people. Similarly, there are edges from media node to every other nodes in the network, namely media permeates in everyone's life. There are also edges from every other nodes directing to media, meaning that media could gather information from all over the world.All those nodes have four states, and it uses the same probabilistic transferring mechanism as other nodes.

- **Every node has a authority value.** We introduce the concept of authority. In real world, we tend to believe the informations telling by a reliable person or a authoritative media company. We use c_i to denote the degree of authority of i^{th} node. Once the i^{th} node is in state of APPROVED and spread news to its neighbor j, the formula (1) is replaced by

  $$m_j(t+1) = m_j(t) + c_i.(2)$$

- **Every edge has a weight value.** As the development of technology, we use different methods to receive news, such as newspaper, radio and internet. Hence, we can contact with more people; we have higher speed to spread news and we gradually change our source of informations. To simulate this evolution, we use a weighted edge graph. Every edge has a weight and this weight will influence how fast the belief $m_j(t)$ cumulates. Mathematically, formula (2) is replaced by

  $$m_j(t+1) = m_j(t) + w_{ij} * c_i, (3)$$

  where $w_{ij} \in [0, 1]$. Although each edge has a weight, the meaning of weights for different types of edges are different. We illustrate this as follows.

1. If the edge is connecting between two individuals(not including the media node), the weight represents the strength of social tie. With the help of technologies, we are able to communicate much more people compared with early stage, so it is reasonable to have more connections in our network model. However, in order to sustain the stability, we use a weight increasing with respect to time rather than increasing edges. From a probabilistic view, the expectation number of ones connections is $w_{ij}k$. If we increase w_{ij} by time, we could simulate the effect that one has more and more friends in his social network.
2. If the edge is directing from media node to individual, the weight represents to what extent people are exposed to social media. In early time, only a few people were able to buy newspapers. As the invention of radio and TV, more and more people are able to gather information from all kinds of media, and nowadays, the majority of human beings are exposed to media in daily life, whether they are willing or not. Thus, this weight is also a increasing parameter as time goes by.
3. if the edge is directing from individual to media, the weight represents to what extent media pay attention to this individual. In our model, media node could also gather information from individual, and, after it approves this news, the media node spread this news to other nodes. Since a media company will only believe credible people, we give more weights to celebrities compared with normal people.

- **A node has an appetite of information source.** A node has different likelihood to gather informations from its neighbors depending on its neighbors' types. For example, in 1900s, people could only know news from newspaper. Now, some people are more likely to hear news from celebrities' Twitter rather than from traditional media, even though they are exposed to traditional media everyday. We have used the weight of edge to depict the extent of exposure to media, now we use $[a_M, a_F, a_N]$ to depict people's willingness to get information from media, celebrities and normal people respectively. Furthermore, we assume

$$a_M + a_F + a_N = 1$$

  since they also represents the ratio of information people get from different source.

- **The edges from media have speed attribute.**In different periods, the speed of media is different, varying from a few days to a few hours. This will significant affect the speed of information propagation. Therefore, to observe this effect, we denote v as the speed and the delivering of message from media node to an individual needs $\frac{1}{v}$ time steps. We calculate v as the weighted average of speed of different media, discussing in later section.

- **Each information has different inherent value.**As a matter of fact, people are more willing to spread valuable news. In our model, λ depicts this willingness, since it is the probability a node transferring to APPROVED when it first hears the news. We use $Model$ to evaluate the inherent value λ_1 of a piece of news and then let

$$\lambda = \lambda_1.(3)$$

- **People have different bias.** In different periods, different people have different affection towards different types of news. We use $Model$ to evaluate a node's bias λ_2 towards a piece of news. Then, combining with previous modification, we replace (3) by

$$\lambda = \lambda_1 \lambda_2$$

- **Feedback and decay of people's bias** In reality, the information we receive could also have influence upon us. For example, if one's friend talks about basketball all day, he might also develop an interest in basketball. On the contrary, if a person is interested in basketball but unable to find a friend to chat with, he might lose his interest gradually. In the light of this observation, we introduce the mechanism of feedback and decay of λ_2 into our model.
  Every piece of news will disappear ultimately. We could make an analogy between this life-span and the time T our network model required to achieve the steady state. During T, if a node transfers to APPROVED, we aggregate its λ_2 by this formula,

$$\Delta\lambda_2 = \alpha \frac{\partial f(x)}{\partial x} \mid_{x=\lambda_2},$$

  where $f(x) = \frac{1}{1+e^{-x}}$ and α is the step-size parameter. This is used to simulate the feedback effect. The sigmoid function $f(x)$ could help to regulate $\lambda_2 \in [0, 1]$. On the contrary, after T, which means the storm of this news has already vanished, we need to decrease λ_2 at every node by this formula,

$$\Delta\lambda_2 = -\beta,$$

  where β is also a step-size parameter. This linear reduction could be used to represent the effect of memory decay.

5 Result

Generally speaking, we are utilizing a network model to simulate the information spreading process in real world society, during which, information's content and network structure show their influences upon information's spreading and public interests. Formalizing the question mathematically, we define T as the the number of time-steps required to obtain the steady result, n as the final number of APPROVED nodes, and u as the number of APPROVED ones divided by the steady time: $u = \frac{n}{T}$. n and u could be viewed as the breadth and speed of the information flow respectively. Furthermore, in order to predict the developing trend of social network, we collected data of some media. The developing trend

5.1 Requirement A

This task asks us explore the flow of information and find what qualifies as news. To give a quantificational description of news , we evaluate the three properties of news[1]: dissemination, timeliness and inherent value. The inherent value is new's initial attribute depicted by λ_1(see datails in Model Section and the Appendix), while the dissemination and timeliness describe the flow of information. In the following part, we use the number of nodes approving news, n, to quantify dissemination and use how fast the news spreads, u, to quantify timeliness.

With the increase of λ_1,

1. n also increases following a 'S-curve' pattern (Discussed in details in the latter section), indicating that when the value of news increases, the number of people APPROVED(n) also increases. When news value increases, in its first stage, the influence of the news increases slowly; and then in second stage its influence increases quickly, and finally the influence become relatively steady.

2. T first increases and then decreases. This means that a worthless piece of news dies down quickly, and very valuable news gets spread over quickly, while the news of medium value takes a little longer time to spread.

3. u increases monotonously which implies that the spreading speed of news increases with respect to the value. When the news is very valuable, its spreading speed increases very fast.

In accordance with the three characters of news, we can give a mathematical definition: a piece of information that satisfies the following conditions could be defined as news: has a spreading speed over certain level, and spreads over to at least some number of people.

From the

5.2 Requirement B

In this part we give a time-based description of the social media involved in the model, so that we can predict the information communication situation for today. Based on our model, the diffusion process of a newly invented social medium is actually the same as diffusion of news. Imagine what happens when a new kind of medium is invented: if it is not attractive enough, it dies down; otherwise if it is, it will spread over the the social network like a piece of valuable news. This intuition can be partly illustrated by the graph describing the development of mobile phone and smart phone:

Hence we take another look of the result for Requirement A. The relation between λ, n is similar to a sigmoid function, making a mathematical analyse necessary: further research implies that spreading rage of news(n) is can be simulated by a logistic function ($y = \frac{1}{1+e^{-x}}$) of λ:

The logistic function , which could be modified as $n = \frac{a}{1+e^{-\lambda}b} + c$, fit with the actual data perfectly. Use the same function on the data of subscribers of cellular telecommunications in US, the result is quite satisfactory. The development of new medium could be resembled by the spreading of news, both of which could be described by the function $n = \frac{a}{1+e^{-\lambda}b} + c$. And the declining of medium user is similarly explained: other social media replace its position, and the declining trend is described by a symmetric sigmoid function. With the formula we have we can calculate the predicted usage of cellular telecommunication of 2014:

the bias of which is reasonable. On the other hand, we obtain the result for newspaper:

5.3 Requirement C

The formula we derived from the previous data enables the prediction of social network capacity and relationship. The capacity of a communication network could be described as the volume of data inside. After acquiring statistical data, we witness a mode for the

The pattern of data volume increasing was considered exponential, however, the experimental result shows that sigmoid function fits better than that of exponential model.

5.4 Requirement D

When describing the influence of a piece of news on people's 'interest and opinion', we propose n and $\Delta\lambda_2$ to describe the degree to which people's own opinion are influenced. The greater n, the more people approve the news and the greater $\Delta\lambda_2$, the more people change their mind.

5.5 Requirement E

This task requires us to determine how different factors could be used to spread information and influence public opinion. Those factors include information value, peoples initial opinion and bias, form of the message or its source, and the topology or strength of the information network in a region. Since the effect upon public opinion is to make them approve a piece of news, we use the number of approved nodes, n, to resemble the ultimate influence on public opinion. To make this problem more realistic, we propose such a scenario: Suppose we are government officers and we want to spread a piece of news. Obviously, we hope this news could influence as large as possible. Therefore, our task is to test each factor as well as evaluate the result, and ultimately find a method to spread our information as influential as possible.

- **Information value** We have showed this result in the section for requirement A.
- **People's initial opinion and bias**
- **Source of message** As a matter of fact, mass media, celebrities and normal people have different importance upon the flow of information. In our model, they are distinguished by their degrees, their values of authorities and people's appetites of information source. Therefore, choice of the starting point of news might have significant influence on the result. We use our model to simulate the effect of different choice, based on estimations of parameters which are showed in table 3.

Table 3 The effect of source of information–Parameters

Type	Number	Degrees	Authority	Appetite of source	Media's interests
mass media	1	500	3	0.5	-
celebrities	25	120	1	0.4	0.9
normal people	475	6	0.5	0.1	0.1

We illustrate the result in figure......... From this graph, we could easily distinguish those curves. The leftmost one is result from averaged result of sending news from mass media, the middle one is from celebrities and the rightmost one is from normal people. It is no doubt that mass media is our best choice. However, we could observe that as the news' value becomes larger, the curve of mass media and the curve of celebrities tend to be identical. This implies that we could also choose celebrities as a relatively good choice.

- **Topology and strength of network** We investigate the spreading of news with respect to its value in two different topology of network, ordered network and random network. For both network, we use 500 nodes to build the network and each node has 6 edges. However, in the ordered network, we order those nodes evenly in a circle and require each node only connect to its nearest 6 nodes, whereas in the random network we connect them randomly. We show the results of n versus λ_1 in $figure$...........

It is obvious that random network has a significant advantage of spreading news compared with ordered network when λ_1 is relatively large.We also calculate the corresponding diffusing speed and its result also prefer random network in most cases. However, we note that the situation is reversed when λ_1

The above mentioned parameter λ_1,λ_1, b could correspondingly describe the information value, peoples initial opinion and bias, and strength of the information network; we can use the seed of information to resemble the source of the message; and accordingly we can vary the network structure to simulate the topology of information network.Spreading of information and influence on public opinion is measured by u,n, and the $\Delta\lambda_2$

6 Sensitivity Analysis

This is actually the concern about whether our model is statistical robust to some outliers. Those outliers could be change in data, could be change in assumptions, could be change in parameter value. We should try our best to test under different outlier conditions, what are the behaviors of our model and compare them in details.

7 Further Discussions

I think this part is more adaptable to the problem requirement, namely we could find what the problem expect us to discuss in this part. Nevertheless, except those specific topic, there are some general things we should discuss no matter what the problem is. They are,
1)Strengthness and weakness.
2)What else can we do in the future.

8 Further Discussion

In the measurement against inherent value of news, there do exist some defective part, for example the Analytic Hierarchy Process (AHP) could be used to give a quantized measure of news value. However, due to lack of space and time, we are not expounding it in details.

On the other hand, we realized that our discussion on historical change of media is not convincible enough. Without enough time and professinal insight, we found it hard to obtain the first-hand data for the media. The flawed model we used can, to some degree reflect the fact for social network, but still requires some modification on the media related parameters. The topology structure here is another point to be discussed.

The memory decay and new edge establishment is deliberately neglected here.

The numerical description of topology could possibly be acquired. Here we restrict the discussion within the categories of topology but did not give a more theoretical description.

References

[1] Stephens, M., *A history of news*, Oxford University Press, USA, 2007.

[2] Lü, L., Chen, D.-B., and Zhou, T., "The small world yields the most effective information spreading," *New Journal of Physics*, Vol. 13, No. 12, 2011, pp. 123005.

[3] Zhao, L., Wang, J., Chen, Y., Wang, Q., Cheng, J., and Cui, H., "SIHR rumor spreading model in social networks," *Physica A: Statistical Mechanics and its Applications*, Vol. 391, No. 7, 2012, pp. 2444–2453.

[4] Liu, C. and Zhang, Z.-K., "Information spreading on dynamic social networks," *Communications in Nonlinear Science and Numerical Simulation*, Vol. 19, No. 4, 2014, pp. 896–904.

[5] Ou, C., Jin, X., Wang, Y., Chen, W., and Cheng, X., "The Impact of Heterogeneous Spreading Abilities of Network Ties on Information Spreading," *Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), 2015 IEEE International Conference on*, IEEE, 2015, pp. 454–460.

[6] Kwak, H., Lee, C., Park, H., and Moon, S., "What is Twitter, a Social Network or a News Media?" *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, ACM, New York, NY, USA, 2010, pp. 591–600.