Crying "Wolf" in a Network Structure: The Influence of Node-generated Signals

Tomer Tuchner[1] and Gail Gilboa-Freedman[2]

[1] Efi Arazi School of Computer Science, IDC Herzliya, P.O. Box 167, Herzliya, Israel, 4610101
tomertuc@gmail.com
[2] Adelson School of Entrepreneurship, IDC Herzliya, P.O. Box 167, Herzliya, Israel, 4610101
gail.gilboa@idc.ac.il

Abstract. Research into rumor spreading in a social network has largely assumed that information may only originate from an external content provider. In today's age individual nodes may also be content providers. The propagation of a signal generated by a node in the network may contribute or diminish the efforts of information diffusion, as signals become an imprecise indication of a node's knowledge.
We present a model that allows for incorporating node-generated information into the well-studied area of modeling rumor spread in a network. We capture this by a stochastic information transmission mechanism at each node, with a positive probability to spread the rumor without holding its value. Simulations are performed using synthetic Watts-Strogatz networks, along with a real-world Facebook sample graph. Using decision trees as a descriptive tool, we examine the effects of the rate in which internal non-informed nodes generate information on the properties of the rumor spread process.
As our main results we show that: increasing the rate of information generated by non-informed nodes may have monotonous or non-monotonous influence on the rumor spread time, in dependency with whether the network is sparse on not. We also identify that a strategy of increasing external communication in order to gain higher pureness level tends to be effective only for a medium level range of this generation rate and only in sparse networks.

Keywords: rumor spread, advertising, word of mouth, social networks, decision tress, predictive models

1 Introduction

With the spread of new social media technologies, which guide more and more of our access to news [9], our responses to everything from natural disasters [22] [10] to terrorist attacks [28], are increasingly disrupted by the spread of unreliable rumors online. The source of these unreliable rumors is often internal, with claims potentially generated by a single non-informed user (for example via a Twitter post) and presented as news. Recipients of this information may then knowingly or unknowingly spread signals that are false.

To examine rumor spread behavior in a network, and how it is affected by unreliable information, let us recall of the fable about the boy who cried "wolf". The boy amuses himself by crying "wolf" to see the panic he causes in the community, but consequently fails to get assistance when a real threat appears. The current study extends the "wolf" story and investigates the influence of "wolf" cries in a network structure, which are equivalent to generating information not originating from an external source of information. This information flow results with suspicion towards received data.

Unlike existing models that describe the transmission of unreliable information [18], our model assumes not only that nodes transfer unreliable information, but also that information may be generated by nodes which have not received any information regarding the rumor from an external source.

The model assumes that curiosity arises about the value of some variable (for example, how many causalities were in an earthquake), and there is some external trustworthy source of information (for example, a news channel) that spreads the real value of the variable (for example, that 23 people were killed in the earthquake). It also assumes that there is an internal diffusion of the information inside the network, from one person to another, and the internal diffusion can be of the real value (23) or of false data (other values, invented by non-informed individuals).

The fact that there is an internal diffusion of data, generated by nodes that have not yet received any information about the rumor, has potentially two opposing impacts on the rate of propagation. On the one hand, spontaneous generation of information in the social network, may increase aggregate growth of informed population. On the other hand, it results with suspicion towards information, which may cause a slowdown in the data spread rate.

The spread of the rumor involves a large number of actions taken by a large number of entities which interact with each other, generating aggregated patterns which are hard to predict, and often impossible to analyze analytically [30]. For this reason, we take a numerical approach, running simulations of rumor spread processes using combinations of the model parameters. Given that some of the model parameters have a non-linear effect, we use decision trees to analyze the simulation results [23]. We highlight this approach as of potential value for other numerical studies on complex networks that depend on large number of variables, with complex relationships between variables and research targets.

We examine the rate in which internal non-informed nodes generate information, and reach two interesting results that focus on the effects this rate has on: 1) the rumor spread time; and 2) the level of pureness for informed nodes, i.e. nodes receiving information originating from an external source.

Concerning the first effect, we examine when does the fact that there is a high rate of faked signals (i.e., generated by a node that doesn't hold a value of the rumor), can harm the rumor diffusion in terms of spread time. We show that the answer is tightly bounded to whether the network is sparse or not.

Concerning the second effect, we identify that a stategy of increasing external communication in order to gain higher pureness level tends to be effective for medium level range of this generation rate, and only in sparse networks.

2 Related Work

In today's "post-truth" age [14], there is increasing interest in the concept of "fake news" – i.e., false stories – in the scientific literature, specifically their epidemiology [17], detection [3], and impact. This phenomenon has the potential to influence attitudes toward journalistic objectivity [21], and may impose real costs on society and politics [1]. From a corporate point of view, false stories have the potential to damage a firm's or brand's image and propel firms into financial disaster [11]. Of course, false stories have always existed – but the ability of social media platforms to spread such narratives rapidly and aggressively gives the question new importance. One recent study focusing on the social network Twitter found that fake news "diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information" [29].

Reliability in general, and rumor reliability in particular, is a central concept in theories of decision-making [25] [27] [33], cooperation [6], communication [26], viral marketing [13] [16], and markets [2]. It is a vast research topic spanning multiple disciplines. The study of reliability commonly draws on network models, which often use the term reliability for the probability that the proportion of informed individuals exceeds a certain value. Some network models associate reliability with social cohesion [31]. Modern network models define reliability as the probability of data transmit from one element to another [7], which is the probability that a given node will be informed.

Our model follows previous work in probability theory on interacting particle systems [20]. We formulate the simplest extension of an independent cascade model, a type of model that has been investigated in the context of marketing and word-of-mouth processes [16] [8]. The dynamic at the node level follows a predefined scheme of response probabilities and is a function of the state of the nodes with which it interacts, as in [12]. Our contribution is in considering a richer dynamic for these interactions, specifically the possibility of activation by non-informed nodes.

We simulate the rumor spread process on a Facebook graph sample taken from the SNAP project [19], on a random graph, and on a series of synthetic Watts-Strogatz networks [32]. Like other studies in the literature on rumor spreading online [4], we also consider a mid-sized network (500 nodes). The intuition behind this is that opinions and rumors often spread within a particular online community which is not that large – for example, people contributing to an online forum or people tweeting and re-tweeting some hashtag.

We analyze the simulation results, by organizing the results in a decision tree for each phenomenon of the rumor spread process we aim to understand. Decision trees are widely used in Machine Learning [15], with the purpose of predicting a target value (a class) from some input features. To build the decision tree, a modeler uses a data set that includes a list of measurable properties (features), one of which is the target. Decision trees are considered to be one of the most popular predictive models (see [24] for survey). They are also used as descriptive tools.[5]

3 Useful Definitions

Below are several definitions for terms used in the article. The purpose of the current section is to help the reader understanding the model description, which contains some new technical terms. Reader may skip to Section 4, and return to this section when there are terms that require further review.

1. **External Communication** – Transmission of information from an external advertiser (source) to any node in the graph.
2. **Internal Communication** – Transmission of information between neighboring nodes in the graph.
3. **Informed/Non-informed Node** – At each iteration, each node can be in one of two states: Informed or Non-informed. Informed nodes are nodes that hold information about the rumor. Non-informed nodes are nodes that do not hold such information.
4. **Pure/Non-pure Node** – Each Informed node is in either a Pure or Non-pure state, and thus there are Pure Informed Nodes or Non-pure Informed Nodes. The pureness state of a node is determined when it is activated (becomes Informed). A Non-informed node that has been activated by an external source, or by a Pure node, is Pure. Otherwise, it is Non-pure. (This is a recursive definition.)
5. **Faked/Held Signal** – Produced when a Non-informed/Informed node (respectively) transmits information to its neighbors.
6. **Reliable/Unreliable signal** – Each signal may be Reliable or Unreliable. Faked signals are unreliable. Held signals are reliable/unreliable when held by pure/non-pure nodes respectively.
7. **Suspicion Factor** – The probability that Non-informed node which receives information from a neighboring node chooses to accept that information and become Informed (rather than to reject it and stay Non-informed).

4 Model

We introduce a model for rumor spread over a network. The network is represented by a graph with a finite set of nodes and a set of undirected edges. We define three states of nodes: *Non-informed* (a node that does not hold information; see def. 3), *Pure-informed* (a node that holds information originating from a reliable source; see def. 4) and *Non-pure informed* (a node that holds information originating from an unreliable source; see def. 4). The novelty of our model is in the fact that non-informed nodes have a positive probability for generating and spreading information on each iteration.

On each iteration, every Non-informed node spreads information to its neighbors with probability $\mathbf{Q}$ (rate of faked signals – always unreliable; see def. 5 and 6), whereas informed nodes spread information to their neighbors with probability $\mathbf{P}$ (rate of held signals – either reliable or not; see def. 5). Every Non-informed node can be activated (become Informed), either by an external source (see def. 1) – with probability $\boldsymbol{\alpha}$, separately for each node – or by the internal communication with its neighbors (see def. 2), where the probability for it to adopt the information from its neighbors is multiplied by the *suspicion factor* $\mathbf{P}/(\mathbf{P}+\mathbf{Q})$ (see def. 7). Nodes that become informed, cannot be deactivated (return to be Non-informed) in later stages of the process. The intuition behind the suspicion factor is that, the possibility of generating signals by non-informed nodes, may hold nodes from adopting any rumor that they get. Without any prior knowledge on whether a node is informed or not, it would be intuitive for its neighbor to consider the likelihood of a signal to be generated by an informed node, and become informed with probability that increases with this rate.

Process is fully described in Algorithm 1. Regarding the question of whether a node becomes pure or non-pure following an activation, we have to consider the possibility that in a single iteration it may choose to adopt received signals from several sources at once. For consistency of the pureness concept, we define that if one chooses to adopt some unreliable signal in some iteration, then it becomes non-pure, or in other words – non-pure activation is dominant.

In Table 1 we show a numerical example for the signal distribution depending on a node's state. Every row in the matrix represents the state of a node, and every column represents the conditional probability for the signal it may transmit, depending on its state. In the given example, $P = 0.8$ and $Q = 0.6$. Please note two important points: 1) A non-informed node spreads unreliable signals (as for not holding the advertised value); and 2) An informed node spreads reliable or unreliable signals in agreement with its state – pure or non-pure (as for holding a value that is sourced at an external source or not).

Table 1. Numerical example for the distribution of the signal transmitted by a node in a specific iteration, depending on the state of the node. In this example, $P = 0.8$ and $Q = 0.6$.

	Unreliable signal	Silence	Reliable signal
Non-informed	0.6	0.4	0
Pure informed	0	0.2	0.8
Non-pure informed	0.8	0.2	0

5 Example: Facebook Sample Graph

We start by simulating our model on a Facebook graph sample, and on a *Erdős-Rényi Random-Graph* with the same number of nodes (4039) and the same average degree (approx. 44). The Facebook graph sample was taken from the SNAP project [19]. The *random-graph* was generated with a computer program in Matlab.

Figure 1 displays the number of iterations needed to reach 95% of informed nodes against Q, where P and α are fixed at 0.01, for the two networks (i.e., the Facebook sample and the random setup).

Before we talk about whether Q strengthens or harms rumor spread in different settings, we wish to examine whether it actually has a two-sided effect on the spread process. As we see in this example of the Facebook network, spread of the rumor speeds up as Q increases, but only at the far left of the graph, where Q is low. At a certain point, increasing Q further once again slows the speed at which the rumor spreads. Figure 1 demonstrates that the ability to speed up the spread of the rumor by increasing the rate of faked signals, may or may not exist, in dependence with network topology.

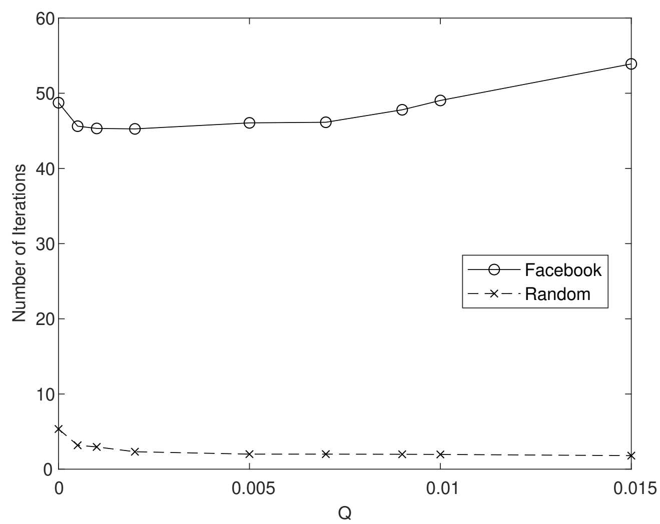

Fig. 1. Extent of rumor spread as a function of the probability of faked signals. The Y-axis is the number of iterations needed to reach 95% of the nodes being informed, and the X-axis is Q, where P and α are fixed at 0.01

6 Methods

We simulate the rumor spread process as described by the model for a variety of model parameters and network structures. We then analyze the simulation results, using decision trees for classifying the simulations by their properties in decision tree structures.

6.1 Preparation of Networks

We wrote a Matlab program for generating synthetic networks using an algorithm from Watts and Strogatz [32]. These networks have 500 nodes and vary on the parameters K and β, which describe respectively the ratio between edges and nodes in the graph, and how close the graph is to a random network, where $\beta = 1$ implies a random network and $\beta < 1$ displays properties of a small-world structure (closer to a social network).

6.2 Numerical Simulations

Combinations of model parameters (α, P, Q) and network parameters (K,β) were considered in a full factorial design experiment. Consistent with previous literature [8], we set the advertising rate α lower than P (the probability of an Informed node to generate a signal). We also set Q – the probability of a faked signal (generated by a Non-informed node) – to be lower than P, under the reasonable assumption that being informed about a rumor should not reduce the probability of spreading it.

Each of the five input variable parameters was manipulated to produce a variety of spread process simulations. For each set of parameters we ran 20 simulations and calculated their mean results. In total we examined $6,720$ sets of parameters (no $\alpha = Q = 0$), in each case running the rumor spread process until 95% of the network was informed.

The parameter ranges were set as follows:

1.	K – Half the average degree	3,10,15,20
2.	β – Randomness of the graph	0.5,0.8,1
3.	α – Advertising rate	0–0.01 (steps of 0.00125)
4.	P – Held signal rate	0.01–0.07 (steps of 0.01)
5.	Q – Faked signal rate	0–P (9 values, constant step)

6.3 Analysis: Decision Trees

After generating all possible outcomes over the set of simulation parameters, we analyzed the influence of the parameters on the behavior of the rumor spread. For this purpose, we took the decision trees approach to elicit relevant observations. Results served as input data for training a decision tree – i.e., a set of rules organized in a hierarchical structure that could serve as a predictive model for the relevant measure. This tool can reveal observations that are non-intuitive and that therefore would not likely be advanced and tested by a "human" learning approach. We wrote a computer program in Python to build the trees.

The results obtained from the decision trees are of rigorous nature, because each observation spots a node or leaf in the tree, that represents a hypothesis with high significance according to traditional methods.

Each decision tree specifies the properties of its nodes. Specifically, for each leaf (which is a node with no arrow coming out from it), the tree specifies: the number of samples that were sorted into this leaf over the categories of the target; their distribution in terms of how many samples fall in each category; and the prediction assigned to this leaf. For each internal node (not a leaf), the graph also specifies the splitting criteria.

We examined the classifications of the simulation results in the trees, and derived our observations based on values in the same class. We are most interested in classes that are homogeneous in terms of the target values of the simulations that fall into them.

7 Results

Firstly, we examine the two-sided impact of information generated by non-informed nodes on the time for a rumor spread. Increasing the rate of generating such information, increases both transmission rate and suspicion. When there exists a value of Q for which lower values are associated with acceleration in the rumor spread and higher values with a slowdown, we say that the network exhibits a **turning point**. We saw an example for this in figure 1 for the Facebook graph. If a turning point exists, then we are not sure that it is beneficial to increase internal communication of faked signals, in order to achieve a faster rumor spread. We get a simple decision tree (see Figure 2), showing that for a sparse network ($k = 3$), it is not always beneficial to increase internal communication (the probability for turning point is 0.95, while for all simulations it is 0.58), and for a highly-connected dense network ($k > 3$), it is beneficial to increase internal communication (the probability for no turning point is 0.82, while for all simulations it is 0.42). Intuitively, when the graph is highly-connected, signals are being spread to many nodes in each iteration, overcoming suspicion.

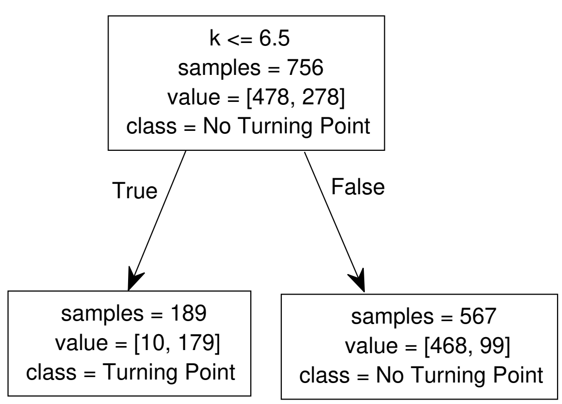

Fig. 2. Decision tree to predict existence of a *turning point*

Our second interest focus lies in the question of – how many nodes received their information from a reliable source (external source or pure node)? We wish to identify which set of parameters brings the level of pure nodes at the end of the process, to a **high pureness** versus low pureness (higher or lower purness than mean observed pureness). The decision tree is displayed in Figure 3. We can see that for medium levels of Q ($0.0025 \leq Q \leq 0.02$, presented in the tree as $Q > 0.002$, $Q \leq 0.021$), if the graph is sparse ($k = 3$), we can increase the advertisement rate ($\alpha \geq 0.00375$), in order to get high pureness and avoid low pureness. When $\alpha \leq 0.0025$ then the probability for low pureness is 0.96, but if $\alpha \geq 0.00375$ then the probability for low pureness is 0.33. We see that for high or low values of Q, we don't need to increase advertisement rate to get high pureness, because there will be too much or not enough unreliable signals for the pureness to be high/low, accordingly.

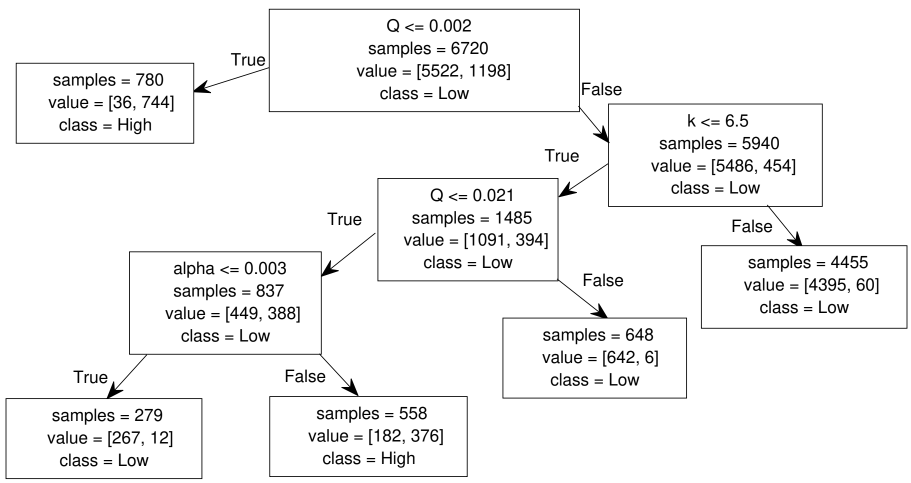

Fig. 3. Decision tree to predict *high* or *low pureness* at the end of the run

8 Conclusions

Motivated by the massive effect of unreliable transmissions in current information technologies, we show the significance of the probability that information will be created at the node level without being informed.

We show some practical implications for a manager controlling the flow of information inside a social network. First, we see that increasing the internal communication of faked signals, may cause a negative effect on the rumor spread time, but only when the network is sparsely connected. Furthermore, increasing the advertisement rate to gain a high pureness level of the nodes, may be only effective for a certain (medium level) range of Q values (faked signal rate), and only if the network is sparse.

Our analysis shows the strength of decision trees as a tool to gain insights in situations that are complex for an analytical solution and depend on many variables, with non-obvious effects on the questions of interest.

9 Acknowledgements

We thank Dr. Ilan Gronau from the Interdisciplinary Center Computer Science School for his knowledgeable and useful remarks.

References

1. Allcott, H., Matthew, G.: Social media and fake news in the 2016 election. Journal of Economic Perspectives 31, 211–236 (2017)
2. Bikhchandani, S., Hirshleifer, D., Welch, I.: A theory of fads, fashion, custom, and cultural change as informational cascades. Journal of Political Economy 100, 992–1026 (1992)
3. Conroy, N.J., Rubin, V.L., Chen, Y.: Automatic deception detection: Methods for finding fake news. In: Proceedings of the 78th ASIS&T Annual Meeting. p. 82 (2015)
4. Cuypere, E.D., Koen, D.T., Wittevrongel, S., Fiems, D.: Opinion propagation in bounded medium-sized populations. Performance Evaluation 99, 1–15 (2016)
5. Delen, D., Kuzey, C., Uyar, A.: Measuring firm performance using financial ratios: A decision tree approach. Expert Systems with Applications 40(10), 3970–3983 (2013)
6. Fehr, E., Fischbacher, U.: The nature of human altruism. Nature 425, 785–791 (2003)
7. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. In: Proceedings of the National Academy of Sciences. pp. 7821–7826 (2002)
8. Goldenberg, J., Libai, B., Muller, E.: Talk of the network: A complex systems look at the underlying process of word-of-mouth. Marketing Letters 12, 211–223 (2001)
9. Gottfried, J., Shearer, E.: News use across social media platforms 2016. Pew Research Center (2016), http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/
10. Gupta, A., Lamba, H., Kumaraguru, P., Joshi, A.: Faking sandy: Characterizing and identifying fake images on twitter during hurricane sandy. In: Proceedings of the 22nd International Conference on World Wide Web (WWW '13 Companion). pp. 729–736. ACM, New York (2013)
11. Hayes, B., Kotwica, K.: Crisis Management at the Speed of the Internet: Trend Report. Elsevier (2013)
12. Holland, J.H.: Hidden Order: How Adaptation Builds Complexity. Addison-Wesley, New York (1995)
13. Jurvetson, S.: What exactly is viral marketing. Red Herring 78, 110–112 (2000)
14. Keyes, R.: The Post-Truth Era: Dishonesty and Deception in Contemporary Life. St. Martin's Press, New York (2004)
15. Kotsiantis, S.B., Zaharakis, I., Pintelas, P.: Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering 160, 3–24 (2007)

16. Krider, R.E., Weinberg, C.B.: Competitive dynamics and the introduction of new products: The motion picture timing game. Journal of Marketing Research 35, 1–15 (1998)
17. Kucharski, A.: Post-truth: Study epidemiology of fake news. Nature 540, 525 (2016)
18. Lazer, D.M.J., Baum, M.A., Benkler, Y., Berinsky, A.J., Greenhill, K.M., Menczer, F., Metzger, M.J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S.A., Sunstein, C.R., Thorson, E.A., Watts, D.J., Zittrain, J.L.: The science of fake news. Science 359(6380), 1094–1096 (2018)
19. Leskovec, J.: Social circles: Facebook. SNAP Project, Stanford University (2012)
20. Liggett, T.M.: Interacting Particle Systems, vol. 276. Springer Science & Business Media (2012)
21. Marchi, R.: With facebook, blogs, and fake news, teens reject journalistic "objectivity". Journal of Communication Inquiry 36, 246–262 (2012)
22. Mendoza, M., Poblete, B., Castillo, C.: Twitter under crisis: Can we trust what we rt? In: Proceedings of the First Workshop on Social Media Analytics (SOMA '10). pp. 71–79. ACM, New York (2010)
23. Quinlan, J.: Induction of decision trees. Machine Learning 1(1), 81–106 (1986)
24. Safavian, S.R., Landgrebe, D.: A survey of decision tree classifier methodology. IEEE transactions on systems, man, and cybernetics 21(30), 660–674 (1991)
25. Savage, L.J.: The theory of statistical decision. Journal of the American Statistical Association 46, 55–67 (1951)
26. Shannon, C.E.: A mathematical theory of communication. The Bell System Technical Journal 27, 379–423 (1948)
27. Simon, H.A.: The New Science of Management Decision. Harper & Brothers, New York (1960)
28. Starbird, K., Maddock, J., Orand, M., Achterman, P., Mason, R.: Rumors, false flags, and digital vigilantes: Misinformation on twitter after the 2013 boston marathon bombing. In: iConference 2014 Proceedings. pp. 654–662 (2014)
29. Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. Science 359(6380), 1146–1151 (2018)
30. Waldrop, M.: Complexity: The Emerging Science at the Edge of Order and Chaos. Simon and Schuster (1993)
31. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications, vol. 8. Cambridge University Press (1994)
32. Watts, D.J., Strogatz, S.H.: Collective dynamics of "small-world" networks. Nature 393, 440–442 (1998)
33. Wedgwood, R.: The aim of belief. Nous 36(s16), 267–297 (2002)

Algorithm 1 Rumor spread process with faked signals

```
Input:
  – P, Q, α - Held signal, faked signal and advertising rates
  – Network G with N nodes and adjacency matrix A
Output:
  – State - array with final state for each node (0 for non-informed, 1 for pure
    informed, 2 for non-pure informed)
  – Time - array with activation time for each node
ρ = P/(P + Q) // Suspicion factor
State = Time = zeros[1, N]
i = 0, informed = 0
while informed < 0.95 · N do
    i = i + 1
    // Randomizing transmission for this iteration
    Transmit = zeros[1, N]
    for n = 1 to N do
        if State[n] > 0 then // Informed
            Transmit[n] = 1 w.p. P
        else  // Non-informed
            Transmit[n] = 1 w.p. Q
        end if
    end for
    // Calculating new states
    NewState = zeros[1, N]
    for n = 1 to N do
        if State[n] = 0 then // Non-informed
            ActivatedByPure = 0
            for j = 1 to N do
                if AND(Transmit[j] = 1, State[j] = 1, A[n, j] = 1) then
                    // Transmitting neighbor is pure
                    ActivatedByPure = 1 w.p. ρ // Adopted with suspicion
                end if
            end for
            if ActivatedByPure = 0 then
                ActivatedByPure = 1 w.p. α // Activated by external
            end if
            if ActivatedByPure = 1 then
                NewState[n] = 1 // Pure
                Time[n] = i
            end if
            ActivatedByNonpure = 0
            for j = 1 to N do
                if AND(Transmit[j] = 1, State[j] ≠ 1, A[n, j] = 1) then
                    // Transmitting neighbor is non-pure or non-informed
                    ActivatedByNonpure = 1 w.p. ρ // Adopted with suspicion
                end if
            end for
            if ActivatedByNonpure = 1 then // Non-pure activation is dominant
                NewState[n] = 2 // Non-pure
                Time[n] = i
            end if
            if NewState[n] > 0 then // Activated
                informed = informed + 1
            end if
        end if
    end for
    State = NewState
end while
Return State, Time
```