Rumor detection and prevention in Social Networks

The Psychology of Fake News[1]

Fake news

- misinformation and disinformation
- hyperpartisan news
- yellow journalism

- conspiracy belief
- Superstition
- bullshit receptivity
- misperceptions

Why Do People Fall for Fake News?

- Political Motivations
    - Identity protective cognition
- Reasoning
    - dual-process theories
        - implicit (automatic) or unconscious
        - explicit (controlled) or conscious
- Heuristics
    - familiarity

Why Do People Fall for Fake News?

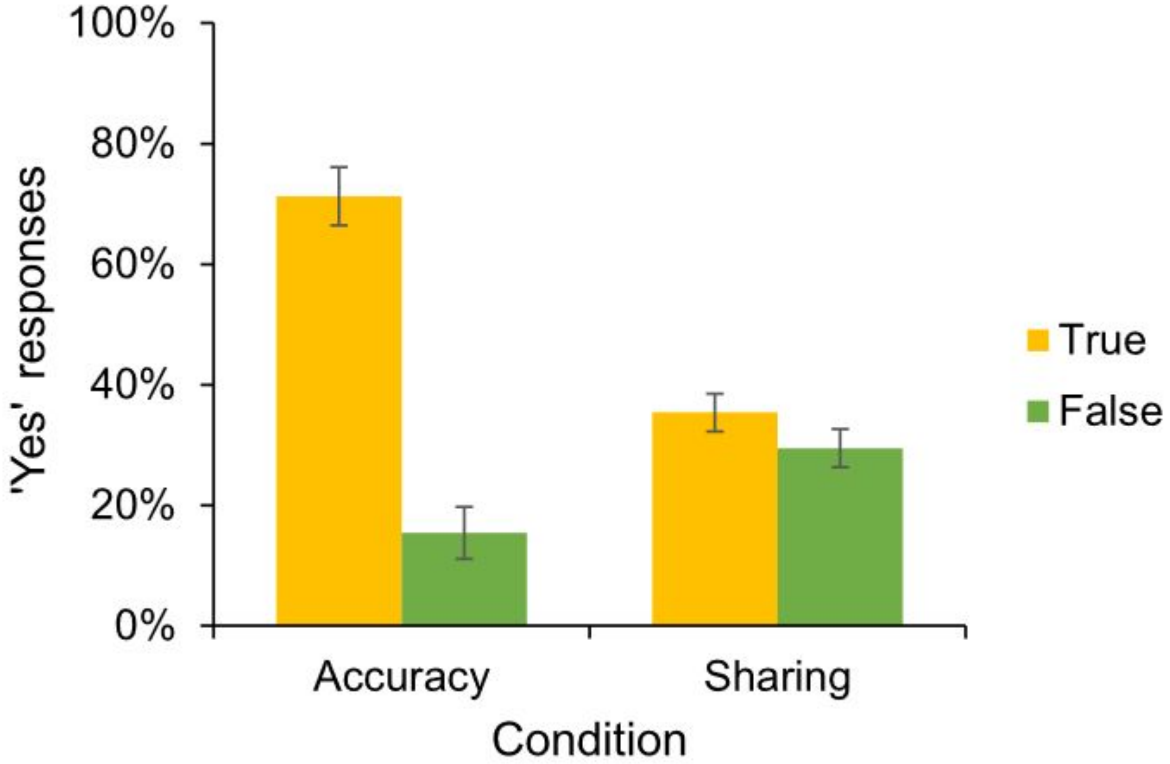
- Political Motivations
    - Identity protective cognition
- Reasoning
    - dual-process theories
        - implicit (automatic) or unconscious
        - explicit (controlled) or conscious
- Heuristics
    - familiarity

Cognitive Reflection Test

'A bat and ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?'

'If you're running a race and you pass the person in second place, what place are you in?'[2]

Social Media Sharing Does Not Necessarily Imply Belief



In literature, the commonly used methods to control the rumors can be divided into three categories: Removing associations between users to block rumors ; Blocking influential users ; Spreading truth to clarify rumors

Blockchain-based rumor detection approach for COVID-19[5]

- Biased graph-based social media network
- Deep Learning Models
- Blockchain Network

Social media network

- GN =(N, C)
  - M malicious nodes
- degree distribution scheme
  - $F(k)=e^{-kc} * ((kc^{k}) / k!)$
- Messages:
  - Correct
  - Rumor
- epidemic routing protocol

blockchain

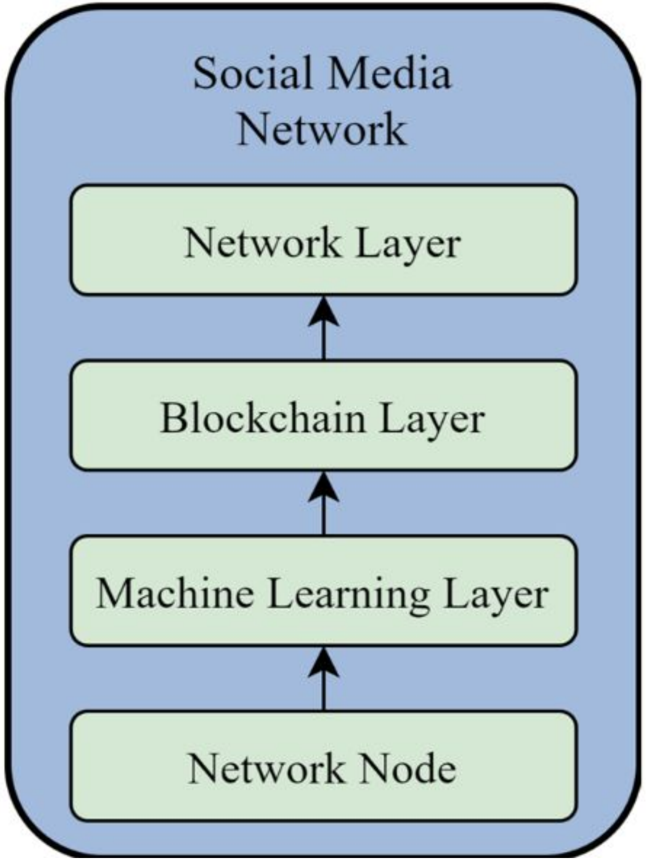
- a joint PoW-PoS strategy

machine learning layer

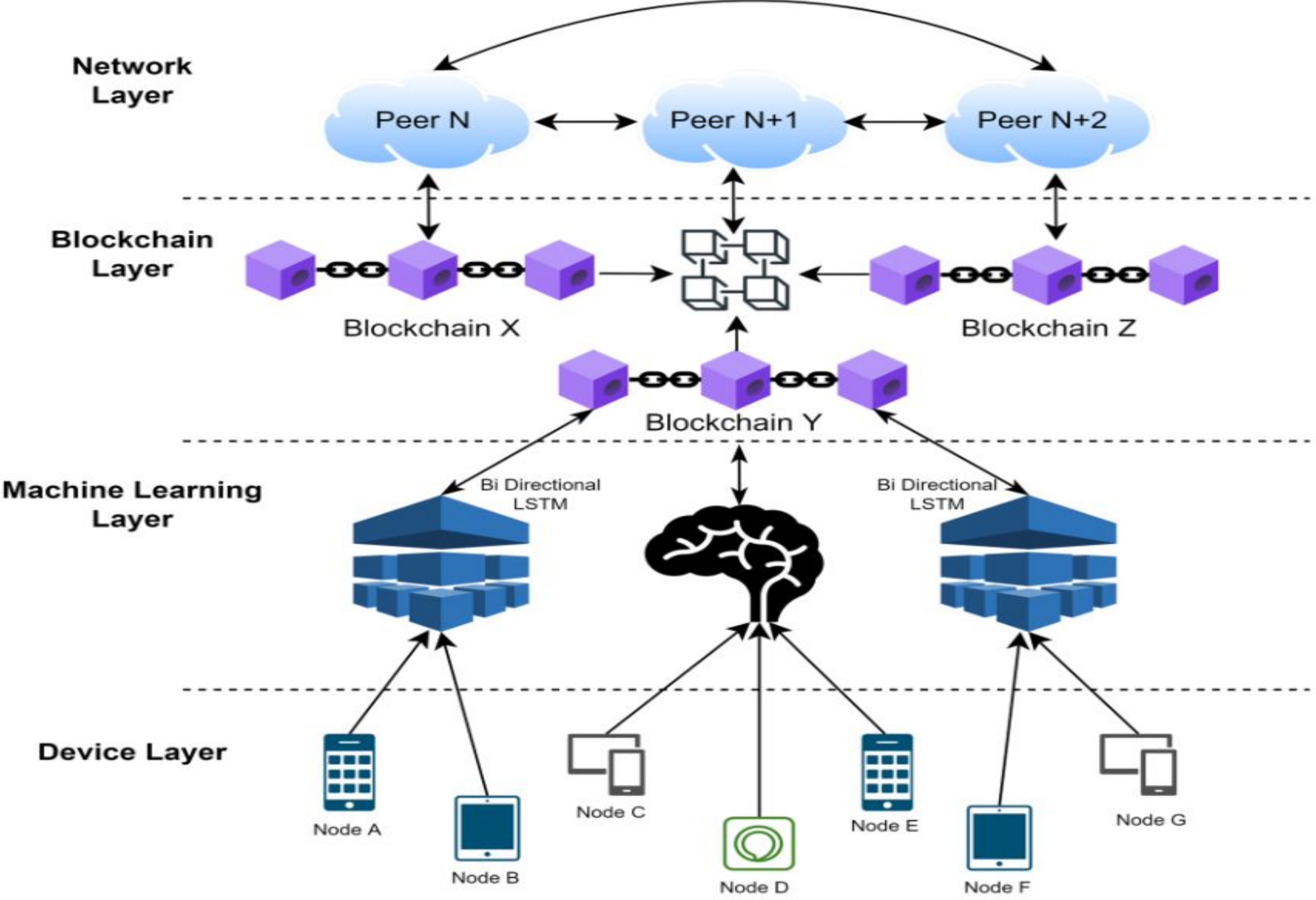
- Bi-LSTM
  - two LSTM models are trained simultaneously

malicious node

- credibility rating
  - $C(X)=\lambda(C(X))old +(1-\lambda)\omega$



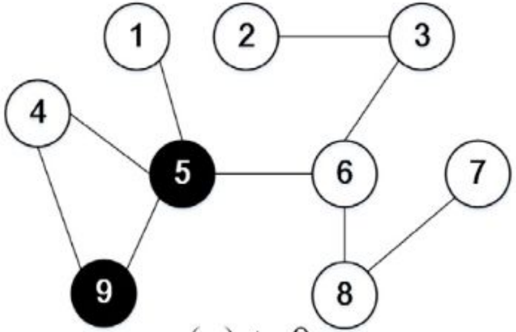
Message flow in network


Network Layer
Peer N
Peer N+1
Peer N+2
Blockchain Layer
Blockchain X
Blockchain Z
Blockchain Y
Machine Learning Layer
Bi Directional LSTM
Bi Directional LSTM
Device Layer
Node A
Node B
Node C
Node D
Node E
Node F
Node G

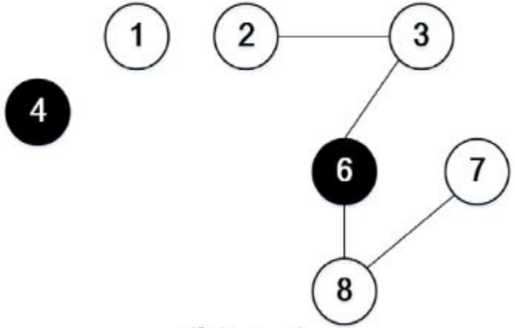
Fast controlling of rumors with limited cost in social networks[4]

classify the users into different groups with different controlling measures.

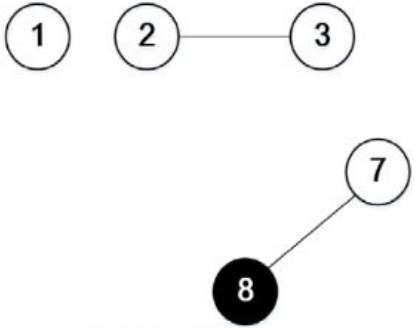
- $H1$
  - taking no action
- $H2$
  - tagging the user
- $H3$
  - blocking access to information
- $H4$
  - spreading the truth
- $H5$
  - deleting the user's account



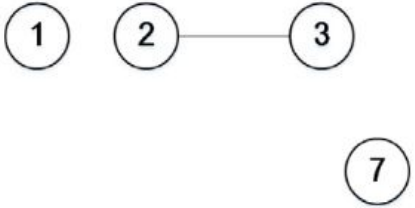
(a) t=0



(b) t=1



(c) t=2



(d) t=3

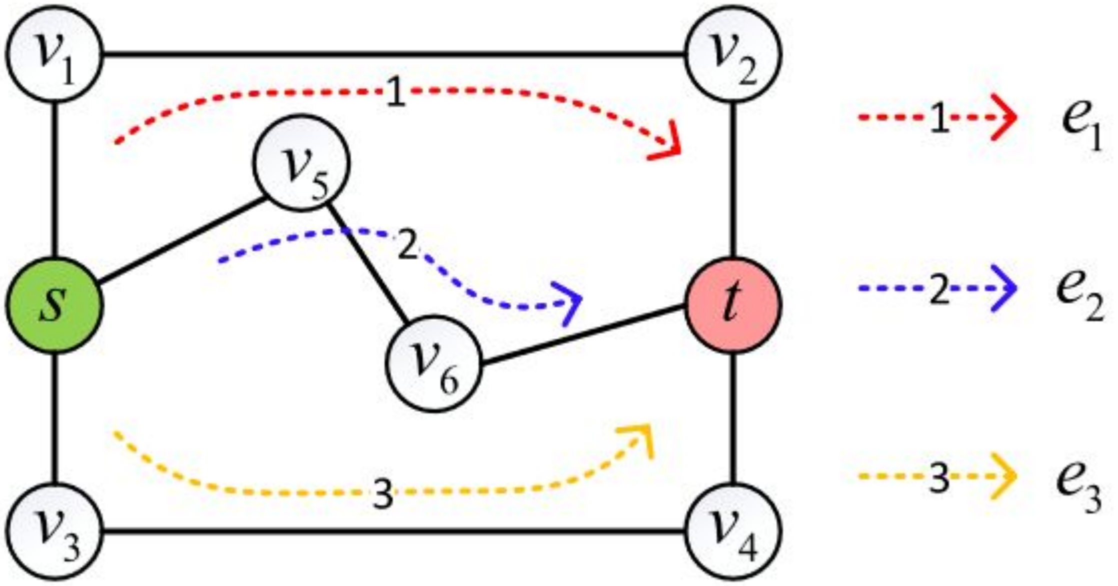
Rumor Dissemination Interruption for Target Recipients in Social Networks[3]

- untrue emergence saying of COVID-19 in an area
- many very sensitive individuals referred to as target recipients
  - Many users may know the truth or do not care about rumors
  - Only those ones who received the rumors and took irrational behaviors

TARGET INFORMATION DISSEMINATING(TID) Model

- network G = (V, E)
- target recipients is denoted by T
- information source nodes is denoted by S
- R_{ST} is path between s and t

Single-Source-Single-Target Problem



$$H_{st} = 1 - \prod_{e \in R_{st}} \left(1 - p_{st}^{e}\right) \quad (1)$$

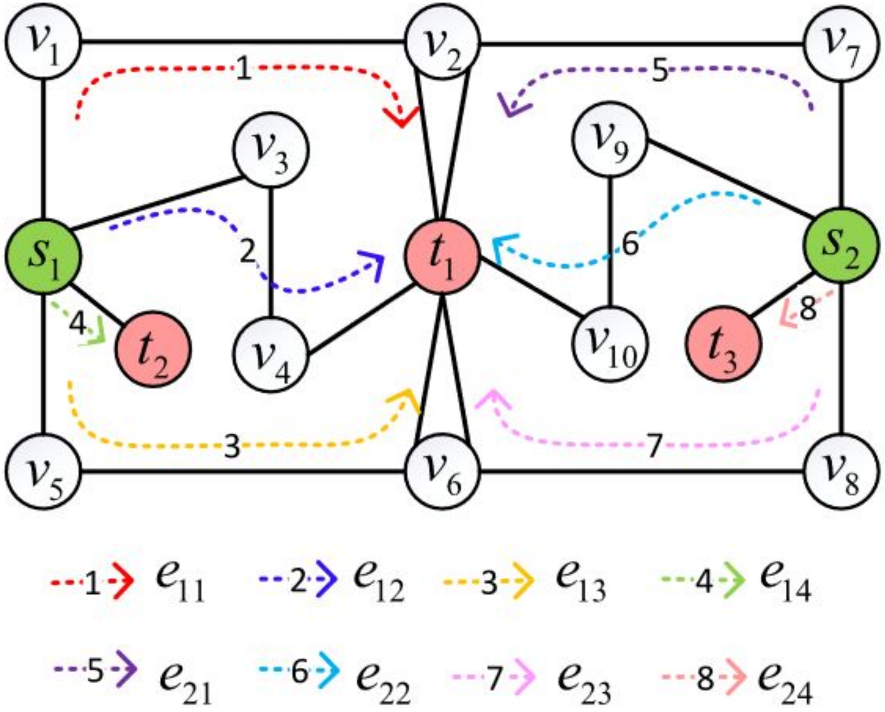
$$F_{st} = \prod_{e \in R_{st}} \left(1 - p_{st}^{e}\right). \quad (2)$$

Greedy Solution for the SS-MinTID

$$\begin{aligned}
G_x^{\#} &= F_{st}^{\#}(E^- \cup \{x\}) - F_{st}^{\#}(E^-) \\
&= \frac{\prod_{e \in R_{st}}(1 - p_{st}^e)}{\prod_{e \in \left\{R_{(E^- \cup \{x\})} \cap R_{st}\right\}}(1 - p_{st}^e)} - \frac{\prod_{e \in R_{st}}(1 - p_{st}^e)}{\prod_{e \in \{R_{E^-} \cap R_{st}\}}(1 - p_{st}^e)} \\
&= \frac{F_{st}}{\prod_{e \in \{R_{E^-} \cap R_{st}\}}(1 - p_{st}^e)} \\
&\quad \times \left[\frac{1}{\prod_{e \in \{R_x \cap (R_{st} \setminus (R_{st} \cap R_{E^-}))\}}(1 - p_{st}^e)} - 1\right]. \qquad (6)
\end{aligned}$$

$$\max \quad (G_x^{\#}) \sim \min\left\{\prod_{e \in \{R_x \cap (R_{st} \setminus (R_{st} \cap R_{E^-}))\}}(1 - p_{st}^e)\right\}. \qquad (7)$$

Multisources–Multitargets Problem



$$F_{ST} = \prod_{i=1}^{n} \prod_{j=1}^{m} \prod_{e \in R_{s_i t_j}} \left(1 - p^{e}_{s_i t_j}\right) = F_{s_1 t_1} * \cdots * F_{s_n t_m}$$

$$H_{ST} = 1 - \prod_{i=1}^{n} \prod_{j=1}^{m} \prod_{e \in R_{s_i t_j}} \left(1 - p^{e}_{s_i t_j}\right) = 1 - F_{ST}.$$

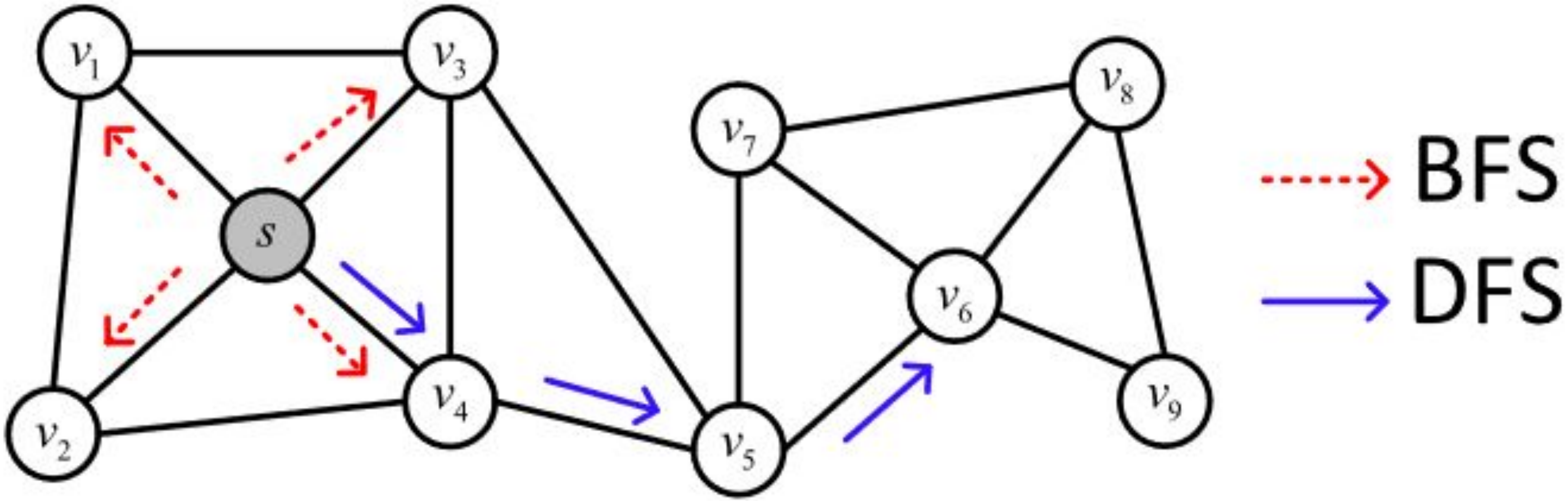
Greedy Solution for the MM-MinTID

$$
\begin{aligned}
G_x^{\#} &= F_{ST}^{\#}\left(E^{-} \cup \{x\}\right) - F_{ST}^{\#}\left(E^{-}\right) \\
&= \frac{F_{ST}}{\prod_{i=1}^{n} \prod_{j=1}^{m} \prod_{e \in \left\{R_{\left(E^{-} \cup \{x\}\right)} \cap R_{s_i t_j}\right\}} \left(1 - p_{s_i t_j}^{e}\right)} \\
&\quad - \frac{F_{ST}}{\prod_{i=1}^{n} \prod_{j=1}^{m} \prod_{e \in \left\{R_{E^{-}} \cap R_{s_i t_j}\right\}} \left(1 - p_{s_i t_j}^{e}\right)} \\
&= \frac{F_{ST}}{\prod_{i=1}^{n} \prod_{j=1}^{m} \prod_{e \in \left\{R_{E^{-}} \cap R_{s_i t_j}\right\}} \left(1 - p_{s_i t_j}^{e}\right)} \\
&\quad \times \frac{1 - \prod_{i=1}^{n} \prod_{j=1}^{m} \prod_{e \in \left\{R_x \cap \left(R_{s_i t_j} \setminus \left(R_{s_i t_j} \cap R_{E^{-}}\right)\right)\right\}} \left(1 - p_{s_i t_j}^{e}\right)}{\prod_{i=1}^{n} \prod_{j=1}^{m} \prod_{e \in \left\{R_x \cap \left(R_{s_i t_j} \setminus \left(R_{s_i t_j} \cap R_{E^{-}}\right)\right)\right\}} \left(1 - p_{s_i t_j}^{e}\right)}.
\end{aligned}
$$

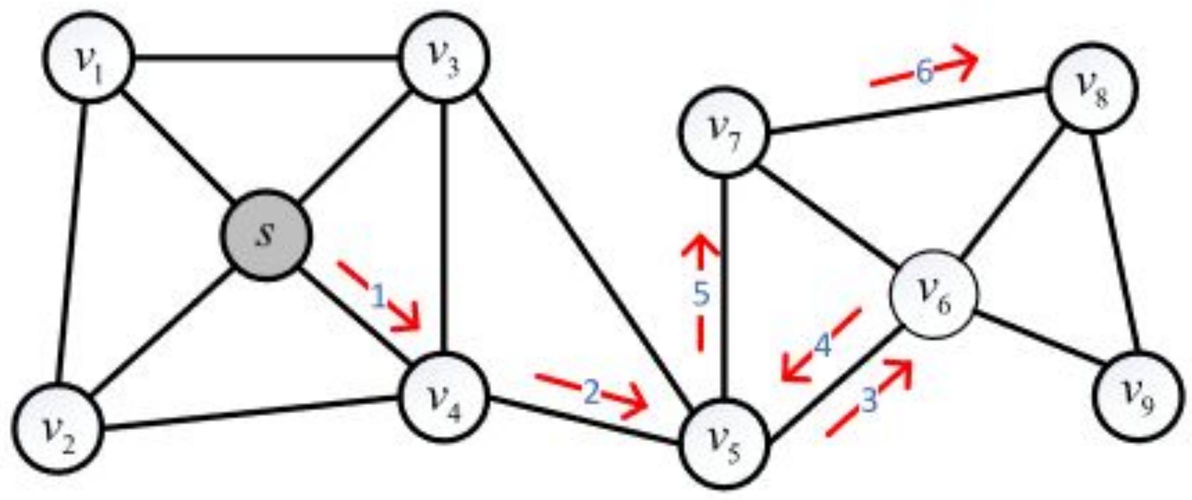
Greedy Solution for the MM-MinTID(con't)

$$\max \left(G_x^{\#}\right) \sim \min\left[\prod_{i=1}^{n}\prod_{j=1}^{m}\prod_{e\in\left\{R_x\cap\left(R_{s_it_j}\setminus\left(R_{s_it_j}\cap R_{E^-}\right)\right)\right\}}\left(1-p_{s_it_j}^{e}\right)\right].$$

Path Sampling



Path Sampling(con't)



Schematic of RW algorithm

heuristical strategy-based rumor influence decay mechanism

- RumorDecay k Hop Nearest Neighbor: RumorDecaykHNN
- RumorDecay k Hop Random Walk: RumorDecaykHRW

An entropy-based method to control COVID-19 rumors in online social networks using opinion leaders[9]

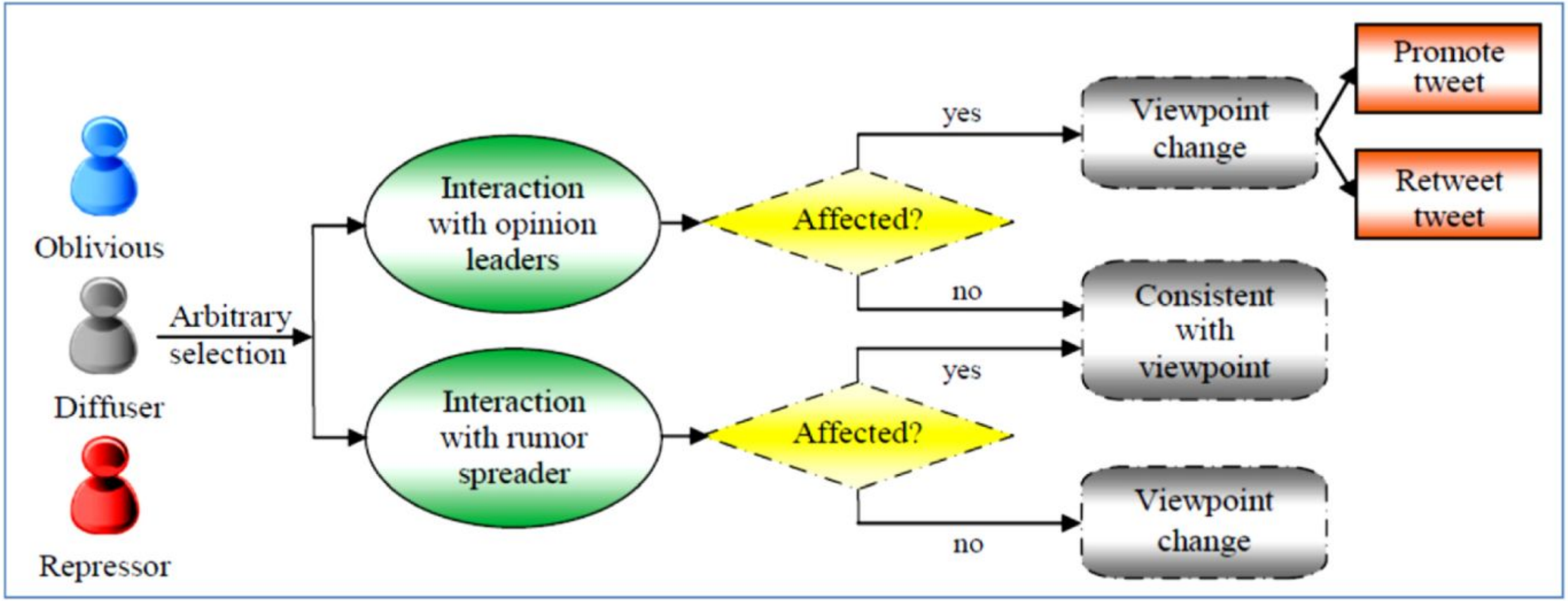
Trust

- provides the opportunity for the user with whom a user can share and accept information

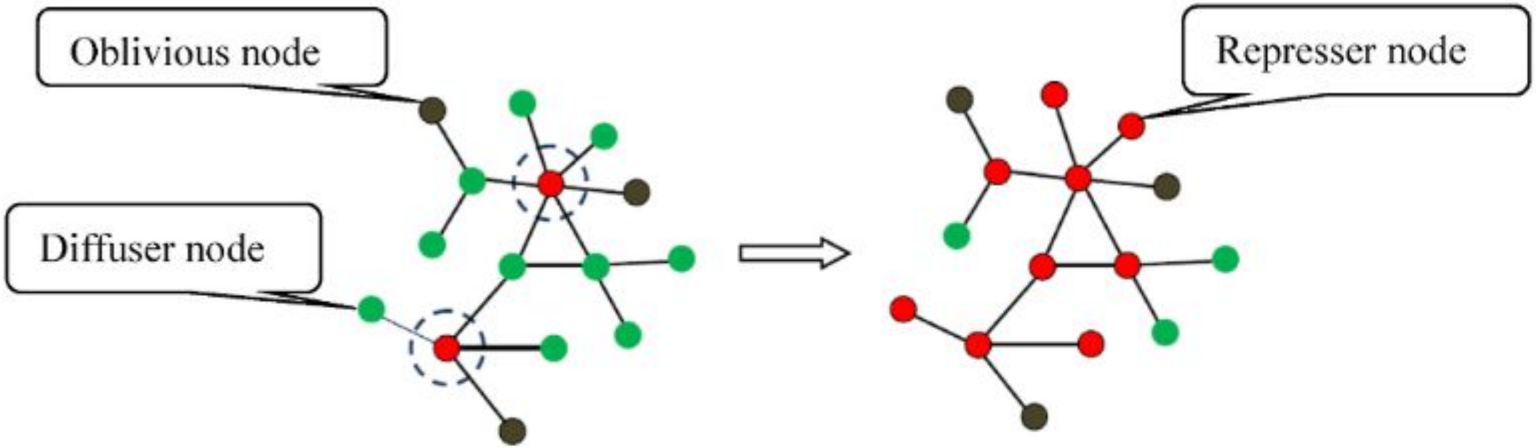
Opinion Leader(OL)

- an OL is an individual or coalition of people who significantly impact the social network's human acceptance process
- OL behaves like a represser who can control spreading rumors as much as possible

Human-influenced viewpoint representation



Network representation with diffuser, represser, and oblivious node



pinion Leader-based Rumor Detection (OLRD) Algorithm

Input; 1. Rumor threshold λ

2. Total m number of tweets posted by n number of users

Output:Decision about rumor spreading

Steps:

1. Apply the ROLI algorithm to find the top-T OLs.
2. For $\forall$ t in m do

Preprocess t by removing iterating characters, hashtags, hashtags, and URLs.

3. End for;
4. For $\forall$ j in n do

Calculate the polarity-based reputation

Compute the degree of trust T_j.

5. End for;

pinion Leader-based Rumor Detection (OLRD) Algorithm

6. For ∀ t in m do

Measure the entropy Et (x_i) of each tweet.

if (Et (x_i) < λ)

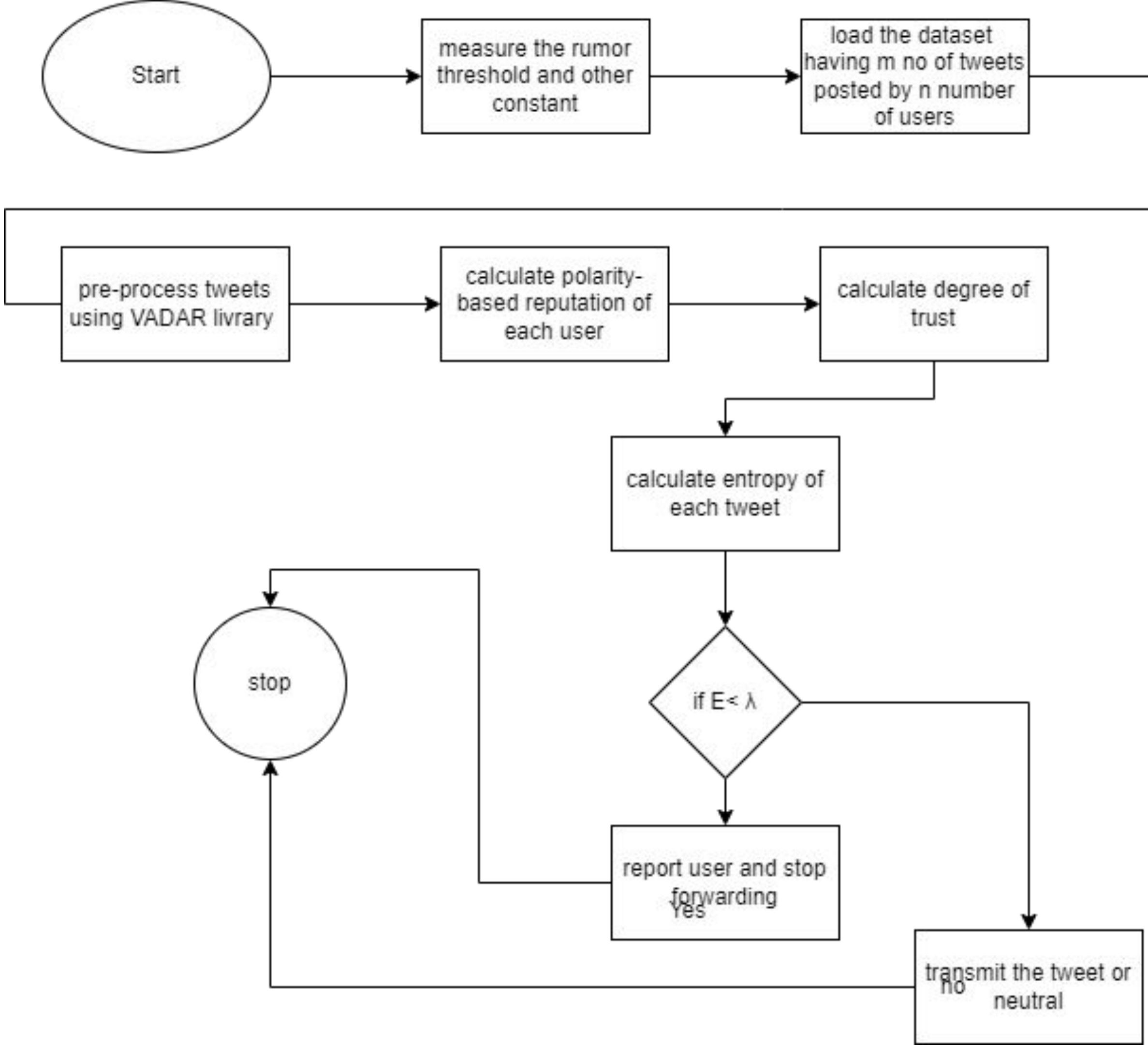
Report the tweet as a rumor and discontinue the rumor spreading.

Else

Transmit the tweet typically. $r_j(t)$.

7. End if;

8. End for;


Start
measure the rumor threshold and other constant
load the dataset having m no of tweets posted by n number of users
find top-N OIs using ROLI algorithms
pre-process tweets using VADAR livrary
calculate polarity-based reputation of each user
calculate degree of trust
calculate entropy of each tweet
if E< λ
Yes
report user and stop forwarding
no
transmit the tweet or neutral
stop

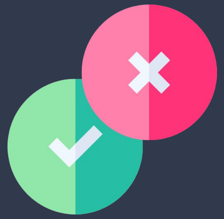
Soft rumor control in social networks: Modeling and analysis[6]

Soft rumor control model

Trusted friends

Reputable authorities

Pros and cons



- Pros:

1. Using an evolutionary game approach to model the spread and control of rumor
2. utilize two sources of providing anti-rumor, i.e. reputable authorities and trusted friends to control rumor
3. The proposed rumor control mechanism can be more effective than hard rumor control mechanisms

Pros and cons

- Cons:
  - The proposed rumor control model is sensitive to the society conditions
  - the proposed soft rumor control mechanism will not control rumor appropriately if:

1- rate of persisting malicious human users who try to deceive people are high

2- social network users are malicious bots

- When a user hears a rumor, generally three different actions are possible:

1- Rumor spreading

2- Anti-rumor spreading

3- Ignoring the rumor

Therefore

Three strategies for users are considered:
rumor spreader (RS), anti-rumor spreader (ARS), and ignorant (I)

The proposed soft rumor control model

- we have two tangible sources for provision of awareness in social network:

-RAs

-TFs

Models:

- Trust model
- Enhanced evolutionary game model

RA: Reputable authorities
TF: Trusted Friends

Trust model

- An important part of the trust model for rumor control is to select trusted friends.

Trust Model

- Interest
- Social intimacy and popularity
- Selecting trusted consultants

Interest

- Acquiring the user’s expertise is a complicated task

we roughly assume that those who are interested in a certain field, are more likely to be expert in that area too

$$IN(f,q) = \sum_{r_i \in R_f} \sum_{t_j \in q} \frac{tf(t_j, r_i) . irf(t_j)}{|R_f|}$$

Social intimacy and popularity

- Social intimacy measures the degree of social similarity between the requester user and his/her friend according to their common relationships

More intimacy between a user and his/her friend

=

More likely the friend will pay attention to the user's queries and respond to them

$$SI(u, f_i) = \frac{\left(Sim\left(ol(u, A), ol\left(f_i, A\right)\right) + Sim\left(il(u, A), il\left(f_i, A\right)\right)\right)}{2}$$

Social intimacy and popularity

- Social popularity of a user measures the degree of responsiveness of the user

More popular = More responsive

$$SP\left(u, f_i\right) = \frac{\sum_{k=1}^{n}\left(\frac{Retweet(t_{ik})}{max\,Retweet(A)} + \frac{Reply(t_{ik})}{max\,Reply(A)}\right)}{2}$$

Selecting trusted consultants

- trusted friends are friends who are interested in the rumor context and have high social intimacy and high social popularity toward the requester user

$$\boxed{Trust(u, f_i, q) = \alpha * IN(f_i, q) + \beta * SI(u, f_i) + \gamma * SP(u, f_i)}$$

Enhanced evolutionary game model

1- Soft rumor control using RAs
2- Soft rumor control using TFs

Soft rumor control using RAs

- One approach of rumor control is to receive anti-rumor messages from RAs
- RAs in different situations have different strengths to convince people not to send rumors
- if we consider a celebrity as an RA, the more the celebrity has published correct and precise anti-rumors and establishes closer relationship with the people and, as a result, people trust the celebrity.

Soft rumor control using TFs

- The effect of consulting TFs on rumor control
- Each user calculates the finesses of his/her friends and selects one of his/her friends proportional
- The user selects his/her friends with highest fitness to imitate their strategy

A social immunity based approach to suppress rumors in online social networks[7]

- In this paper, an anti-rumor information spreading approach is proposed to contain rumors collectively by following a bio-inspired immunization method called social immunity

Rumor containment - a social immunity approach

1 -Identify the most influential persons in every community who can enable the herding/cascading process

2 -The gateway influential spreaders are identified to increase the true information reach across communities

3 -True information is spread to reduce the susceptibility of the individuals in the network

Rumor containment - a social immunity approach

- Rumor intensity describes the severity of the rumor transmission in the network.
- The intensity of rumor affection in the network is calculated by tracking the amount of population affected by the rumor.

$$r_d(t) = \frac{Number\ of\ rumor\ affected\ nodes}{Number\ of\ total\ nodes}$$

Processes

- Herding process
- Cascading process

Herding influencer

- OSNs* follow power-law degree distribution and having a higher cluster coefficient
- The higher cluster coefficient implies that the people from the same community share almost the same interest.
- An influential person from the same community has the ability to impact more numbers of people in the community.

*OSN: Online social networks

Gateway influencer

- A bridge connector between two clusters will be influencing both the clusters.
- This person can introduce the information from one cluster to the other
- This person can send true information faster between communities

Algorithms

Algorithm 1 Herding Influence Finder

INPUT: $G = (V, E, B)$, a n set of communities $C = \{C_1, C_2, C_3, \dots, C_n\}$, rumor depth $r_d(t)$
OUTPUT: Herding Influencer set S.
$S \leftarrow \theta, U' \leftarrow V - S,$
foreach $Comm$ in C do:
    $Seed_{Comm} = Round(rd(t) * \gamma(|Pros(t) + Ign(t)|))$
    for $range(Seed_{Comm})$):
        $v = \arg max_{u \in Comm/S}\{(BC_u)|u \in [Prosocial\ , Ignorant]\};$
        $S = S \cup v;$
    endfor
endforeach
return set S

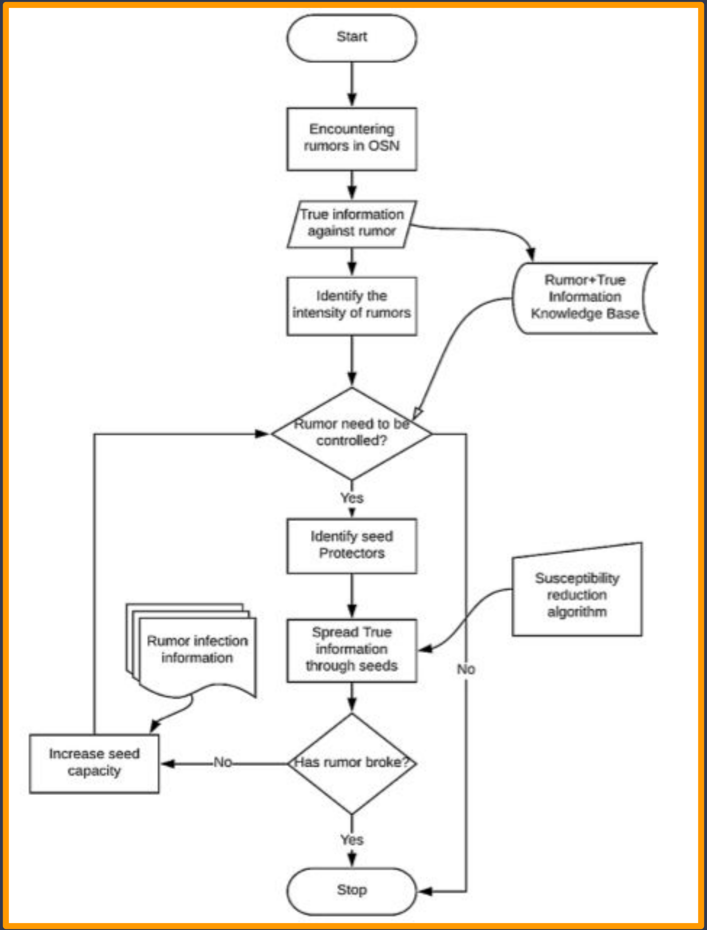
Algorithm 2 Gateway Influencer Finder

INPUT: $G = (V, E, B)$, a n set of communities $C = \{C_1, C_2, C_3, \dots, C_n\}$,
OUTPUT: Gateway influence set GI
$GI \leftarrow \theta, U' \leftarrow V - S,$
foreach $Comm$ in C do
    foreach node i in $Comm$:
        if $\exists\ e \mid (i, u)\ \&\ u \nexists\ in\ Comm$:
            $PN = PN \cup i;$
        endif
    endforeach
    $w = \arg max_{u \in PN}\{(BC_u)|u \in [Prosocial\ , Ignorant]\};$
    $GI = GI \cup w;$
endforeach
return set GI

Rumor containment - Susceptibility reduction

- Susceptibility of individuals is reduced by increasing the immunized population in the network
- The true information should reach a maximum number of individuals in the network faster than rumor to ensure the rumor is suppressed

Social immunity for susceptibility reduction (rumor control) in OSNs



COVID-19 Rumor Detection Using Psycho-Linguistic Features[8]

- The study utilizes a publicly available and labelled Twitter data set and proposes a novel feature space, which can detect rumor instances with high accuracy
- The proposed feature space not only includes content-based features, but also includes psycho-linguistic features

The COVID-19 rumor dataset

- Tweets
- Comments
- Sentiment

The authors also cross-checked the labelled sentiment with online sentiment analysis tool, MonkeyLearn

CONTENT-BASED AND CONTEXTUAL FEATURES

- The content-based features are the features that are directly generated from the contents of the tweet themselves.

5X  4X  50 Words 

- Contextual features on the other hand are features that add more context information to the content in these microblog environments

20X  10X  5X 

PSYCHO-LINGUISTIC FEATURES

- Psycho-linguistic features refer to features that are generated by analysing the language used in the content, with the goal of identifying the psychological and emotional profile of the user, who is responsible for creating the content

- For Example: Use of first-person and second-person pronouns are indicative features for imaginative writing

Summary of feature space

Feature Class	Features
Content-based Features	noOfSmiliesInTweet
	noOfQuestionMarksInTweet
	lengthOfTweet
	noOfSmiliesInComments
	noOfMentionsInComments
	noOfHashTagsInComments
	presenceOfURLsInComments
	noOfQuestionMarksInComments
	lengthOfComments
Contextual Features	noOfLikesTweet
	noOfCommentsTweet
	noOfRetweets
	noOfRepliesTweet
	noOfLikesComments
	noOfRetweetsOfComments
	noOfRepliesOfComments
	noOfSupportStanceOfTweet
	noOfDenyStanceOfTweet
	noOfCommentStanceOfTweet
	noOfQueryStanceOfTweet
Psycho-linguistic Features	hasSwearWords
	hasModalAdverb
	hasActionAdverb
	has1stPersonPronoun
	has2ndPersonPronoun
	hasMannerAdverb
	hasHedgeWords
	hasSuperlatives
	hasComparatives
	sentimentScore
	emotionScore

Experimental results across different setups

Algorithms	Experiment A (Psycho-linguistic Features Only)				Experiment B (Content-based and Contextual Features)				Experiment C (Content-based + Contextual + Psycho-linguistic features)			
	Accuracy(%)	Precision	Recall	F-Measure	Accuracy(%)	Precision	Recall	F-Measure	Accuracy(%)	Precision	Recall	F-Measure
J48	63.77	0.692	0.502	0.582	76.34	0.767	0.760	0.763	77.80	0.768	0.800	0.784
JRip	62.18	0.664	0.499	0.570	73.98	0.731	0.761	0.746	75.50	0.740	0.787	0.763
Random Forest	66.65	0.738	0.520	0.610	80.86	0.775	0.871	0.820	81.00	0.775	0.871	0.820
Naive Bayes	62.41	0.674	0.487	0.565	61.30	0.676	0.441	0.534	66.59	0.705	0.575	0.634

References

1. Pennycook, Gordon, and David G. Rand. “The Psychology of Fake News.” *Trends in Cognitive Sciences* 25, no. 5 (May 1, 2021): 388–402. https://doi.org/10.1016/j.tics.2021.02.007.
2. Mosleh, Mohsen, Gordon Pennycook, Antonio A. Arechar, and David G. Rand. “Cognitive Reflection Correlates with Behavior on Twitter.” *Nature Communications* 12, no. 1 (February 10, 2021): 921. https://doi.org/10.1038/s41467-020-20043-0.
3. Jiang, Zhongyuan, Xianyu Chen, Jianfeng Ma, and Philip S. Yu. “RumorDecay: Rumor Dissemination Interruption for Target Recipients in Social Networks.” IEEE Transactions on Systems, Man, and Cybernetics: Systems 52, no. 10 (October 2022): 6383–95. https://doi.org/10.1109/TSMC.2022.3144141.
4. Yao, Xiaopeng, Yue Gu, Chonglin Gu, and Hejiao Huang. “Fast Controlling of Rumors with Limited Cost in Social Networks.” Computer Communications 182 (January 15, 2022): 41–51. https://doi.org/10.1016/j.comcom.2021.10.041.
5. Syed Mahbub, Eric Pardede, A. S. M. Kayes. “COVID-19 Rumor Detection Using Psycho-Linguistic Features.” *IEEE Access* (10 November 2022). https://doi.org/10.1109/ACCESS.2022.3220369

6. Askarizadeh, Mojgan, and Behrouz Tork Ladani. “Soft Rumor Control in Social Networks: Modeling and Analysis.” Engineering Applications of Artificial Intelligence 100 (April 1, 2021): 104198. https://doi.org/10.1016/j.engappai.2021.104198.
7. Srinivasan, Santhoshkumar, and Dhinesh Babu L D. “A Social Immunity Based Approach to Suppress Rumors in Online Social Networks.” International Journal of Machine Learning and Cybernetics 12, no. 5 (May 1, 2021): 1281–96. https://doi.org/10.1007/s13042-020-01233-0.
8. Mahbub, Syed, Eric Pardede, and A. S. M. Kayes. “COVID-19 Rumor Detection Using Psycho-Linguistic Features.” IEEE Access 10 (2022): 117530–43. https://doi.org/10.1109/ACCESS.2022.3220369.
9. Jain, Lokesh. “An Entropy-Based Method to Control COVID-19 Rumors in Online Social Networks Using Opinion Leaders.” Technology in Society 70 (August 1, 2022): 102048. https://doi.org/10.1016/j.techsoc.2022.102048.