Contents lists available at GrowingScience

International Journal of Data and Network Science

homepage: www.GrowingScience.com/ijds

Information diffusion model with homogeneous continuous time Markov chain on Indonesian Twitter users

Firdaniza[a*], Budi Nurani Ruchjana[a], Diah Chaerani[a] and Jaziar Radianti[b]

[a]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang, 45363, Indonesia*
[b]*Department of Information Systems, University of Agder, 4630, Kristiansand, Norway*

C H R O N I C L E

Article history:
Received: October 10, 2021
Received in revised format: November 20, 2021
Accepted: April 5, 2022
Available online: April 5 2022

Keywords:
Twitter
Information diffusion model
Influencer
Homogeneous continuous time Markov chain
KDD

A B S T R A C T

In this paper, a homogeneous continuous time Markov chain (CTMC) is used to model information diffusion or dissemination, also to determine influencers on Twitter dynamically. The tweeting process can be modeled with a homogeneous CTMC since the properties of Markov chains are fulfilled. In this case, the tweets that are received by followers only depend on the tweets from the previous followers. Knowledge Discovery in Database (KDD) in Data Mining is used to be research methodology including pre-processing, data mining process using homogeneous CTMC, and post-processing to get the influencers using visualization that predicts the number of affected users. We assume the number of affected users follows a logarithmic function. Our study examines the Indonesian Twitter data users with tweets about covid19 vaccination resulted in dynamic influencer rankings over time. From these results, it can also be seen that the users with the highest number of followers are not necessarily the top influencer.

1. Introduction

Studies on the model of information diffusion or information dissemination on social media such as Twitter have been carried out by many researchers. Various methods have been proposed and used in modeling information diffusion, such as the susceptible-infected recovered (SIR) model (Zheng et al., 2018), susceptible-infected-state (Kim & Seo, 2020), and susceptible-exposed-infected (SEI) (Kumar et al., 2020) which is basically representing an epidemic model. The user who receives the information is assumed to be affected by the user who submitted the information, resharing them to more people, working in almost the same way as a disease infects people in bigger groups. In addition to epidemic models, information diffusion has also been modeled using stochastic models (Kawamoto, 2013), continuous-time Markov chains (Li et al., 2014; Zhu et al., 2014; Firdaniza et al., 2021). In this case the information dissemination satisfies the Markov property, the future state depends only on the present state and is independent of the past (memoryless). In this case the information received by a user depends only on information from the previous user. Furthermore, information diffusion has been modeled as a regression model by Yoo et al. (2016). Yoo et al. suggest that the diffusion of information is influenced by the users who disseminate the information, the content delivered and the timing of the delivery of the information. Kwon et al. (2017) using Poisson regression in modeling the diffusion of information. In this work, information diffusion is influenced by information sources, group effects and cross-interactions. Information diffusion has also been modeled using the Evolutionary Game Theory model (Jiang et al. (2014), Bayesian network (Varshney et al., 2017), and Quantum q-attention model (Shuai et al., 2012). However, social media studies that apply a homogeneous continuous time Markov chain (CTMC) is still very limited, as indicated in the previous study (Firdaniza et al., 2022).

* Corresponding author.
E-mail address: firdaniza@unpad.ac.id (Firdaniza)

doi: 10.5267/j.ijdns.2022.4.006

The study of information diffusion models, especially on Twitter, is interesting to examine, as it offers many possibilities, approaches, methods, and theory development to investigate the relationships occurring in this specific social media. This study will use Indonesian Twitter content as a case study. Twitter is a microblogging service with a rapidly growing number of users, and one of the most popular research subjects when it comes to social media influence. As of December 2021, there are 15.7 million Twitter users in Indonesia that is, in fact, the sixth country with the most Twitter users in the world (statista.com, 2021). A Twitter user can follow other users, but not necessarily reciprocally, meaning users who are followed do not need to follow back (Kwak et al., 2010). If user B follows user A, then B is called a follower of A and A is called a follower of B (Firdaus et al., 2018). Twitter provides services for its users to create profiles, share stories and convey messages or information. Twitter users can convey messages (which are called tweets) that were originally only 140 characters, since 2017 it has been changed to a maximum of 280 characters (Fernando et al., 2019). Tweets can contain plain text, URLs, images, mentions of other users (starting with the "@" symbol), and hashtags ("#") i.e. words that you decide to highlight. The dissemination of information on Twitter has a distinctive pattern, where information flows through a network of follower-followee. If a tweet has been uploaded by a user, it will appear on the followers' timeline. If followers agree with the content of the tweet, then they can respond to the tweet by writing an RT (short for retweet) then "@" followed by the user ID. This indicates that information spreads from the first retweet (Kwak et al., 2010).

Relationships and interactions within individual groups will play an important role as a medium for disseminating information and influence among its members (Zhu et al., 2014). A user can be the most influential in conveying a tweet message if the message can spread to a large number of other Twitter users. In this case, people who have a strong influence on the Twitter network are referred to as influencers (Peng et al., 2018). In most diffusion models, influencers have a certain opportunity to influence their followers, and followers also have a certain threshold to be influenced (Li et al., 2014).

Measuring the most influential users on a social network is a conceptual problem, and there is no agreement on what methods, measure, and criteria to determine a certain Twitter profile is the influential user. Hence new influence measures may continue to emerge with different criteria (Riquelme & González-Cantergiani, 2016). To measure user influence, some researchers have proposed and applied proximity centrality, intermediate centrality, and degree centrality (Mittal et al., 2020; Tidke et al., 2020), PageRank (Zhang et al., 2020; Oo & Lwin, 2020; Alp & Öğüdücü, 2019; Bhowmick, 2019; Hamzehei et al., 2017), Buzz Rank (Simmie et al., 2014), T and HT (Qasem et al., 2017).

The previously studied information diffusion models are generally descriptive, not predictive. The influencer ranking is determined based on a static network, meaning that the influencer rating is determined only for the moment. In disseminating information, it is very important to know who the influencers are for a certain time in the coming period. From our previous research (Firdaniza et al., 2022), it was revealed that the study of the information diffusion model on Twitter with a continuous time Markov chain (CTMC) is still openly available to scrutinize, as there are very little studies approaching Twitter data analysis using the CTMC. Information diffusion on Twitter can be modeled with the Markov model because the dissemination of information fulfills the Markov property of memory less. Tweets received by followers only on tweets from one previous follower. Based on this linkage, the interesting thing that can be done is that we predict influencers for the desired time, such as within a day, a week or some other time.

In this paper, an information diffusion model using homogeneous CTMC is employed by using Indonesian Twitter data, regardless of users´ geographical locations. This model has advantages over static models, because it can answer the question of who has the most influence in disseminating information for a certain time in the future. This model refers to the research conducted by Li et al. (2014). The new novelty of this work lies in the procedure and method applied on the Twitter datasets to determine the number of affected users on the test data using a logarithmic function, in contrast to Li et al. (2014) which uses a linear prediction model. The logarithm function is used based on available real data, so we hypothesize that the determination of influencers will be close to the actual phenomenon.

2. Problem Description

Information diffusion on Twitter is characterized by tweet-retweet activity by Twitter users. If a user tweets about a topic at a certain time, and moments later another user uploads a tweet with the same topic, then this user is said to have been affected by the first user. Information is said to spread only if the users involved in discussing a topic are tied to each other in a network of followers-followees. The spread or diffusion of this information is modeled by homogeneous CTMC.

Referring to Ross (2010), the stochastic process $\{X(t), t \geq 0\}$ is said to be CTMC if it fulfills the Markov property, namely for all $s, t \geq 0$ and nonnegative integers $i, j, x(\gamma), 0 \leq \gamma < s$,

$$P\{X(t+s) = j | X(s) = i, X(\gamma) = x(\gamma),\ 0 \leq \gamma < s\} = P\{X(t+s) = j | X(s) = i\} \tag{1}$$

If $p_{ij}(t) = P\{X(t+s) = j | X(s) = i\}$ does not depend on s, then CTMC is said has a homogeneous transition probability and $p_{ij}(t)$ is called the transition probability from state *i* to state *j* in time *t*.

Users in the followers-followees network who are involved in discussing a topic are a state of CTMC. The questions in this research are:

1. How to estimate a transition probability matrix that describes the spread of information from one user to another on Twitter?
2. How to determine the top influencers dynamically over times?

To answer these research questions, this paper describes the research methodology as described in section 3.

3. Methodology

The methodology we use follows Knowledge Discovery in Database (KDD) (Han, et al., 2012) as shown in Fig. 1. We carry out the following processes: data retrieval, data preprocessing, data processing (data mining; using CTMC for the information diffusion model), then evaluation of model performance and post-processing; as the end of the process, namely getting influencers. To be more precise, in Section 3.1-3.4, some explanations on the KDD process are discussed.

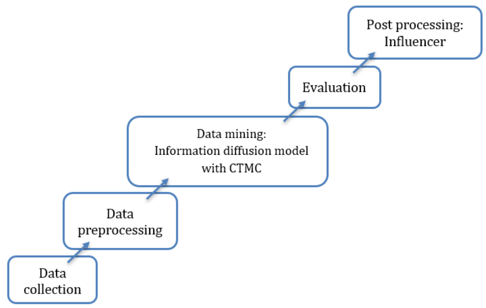


Fig.1. Research Methodology

3.1 Data collection

Twitter data was retrieved via netlytic.org using the keywords "VaksinasiCovid19" OR "vaksinasicovid19". We limit the tweets to be included that were written in Indonesian. Data collection was carried out from February 17, 2021 to February 27, 2021. We received 100,000 tweets discussing the topic covid19 vaccination.

3.2 Data preprocessing

At the data preprocessing stage, the following are carried out:

(i) Deleting Twitter users other than Indonesia, in this case, we remove Twitter users from Malaysia. From this data collection activity there were 38,787 active users that were involved in posting and resharing the tweets.

(ii) Searching the list of followers-followers of the user involved in the tweet using the Twitter API.
(iii) Creating a time network between tweets from followers and followers as a form of information dissemination that fulfills Markov's nature programmatically using Python. In this case the time data studied is 10 days (240 hours). We only processed 5,000 users involved due to limited data processing tools at our disposal. An illustration of the time network between tweets can be seen in Fig. 2.
(iv) Divide the data into 2 parts, namely 80% training data and 20% testing data. The training data is used to build the Information Diffusion model with CTMC, then the testing data is used to evaluate the model.

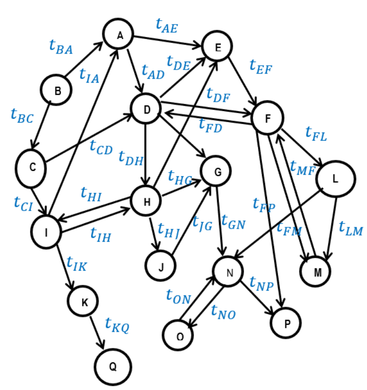


Fig. 2. Illustration of time network between tweets

In Fig. 2, t_{BA} represents the time between user B (follower A) tweeting the same topic after user A tweets.

3.3 Information Diffusion Model with CTMC

Information diffusion on Twitter begins with Twitter users uploading topics via tweets at time *t*, let's say $X(t)$. Users discussing topics depend only on previous users and not on the topic's spread history. In this case $\{X(t), t \geq 0\}$ forms CTMC like equation (1). Let i denote the Twitter user who is currently tweeting about a topic, j is the next user to tweet about the same topic after user i, and $x(\gamma)$ describes the topic's spread history before time *s*. The information diffusion model on Twitter with CTMC is to estimate the transition probability matrix $\mathbf{P}(t) = [p_{ij}(t)]$ that can be obtained through the transition rate matrix $\mathbf{Q}(t)$, namely the rate of information dissemination from one user to another. In this case, the rate of information dissemination is assumed to be constant or is called homogeneous CTMC.

The transition rate matrix $\mathbf{Q} = [q_{ij}]$ (Ross, 2010), represents the rate of topic spread from user *i* to user *j* where

$$p_{ij}(h) = q_{ij}h + o(h)\,,\ i \neq j \tag{2}$$

$$1 - p_{ii}(h) = v_i h + o(h) \tag{3}$$

where $o(h)$ is an infinitesimal that satisties $\lim_{h\to 0}\frac{o(h)}{h} = 0$ and the rate of topic spread from user *i* to any other user, $v_i = -q_{ii} = \sum_{j\neq i} q_{ij}$.

For any *i* and *j*, let

$$q_{ij} = v_i p_{ij} \tag{4}$$

Since v_i is the rate of spread of information from user *i* and p_{ij} is the probability that this information spreads to user *j*, then q_{ij} is called the instantaneous transition rate.

The transition rate matrix $\mathbf{Q}_{l\times l}$ satisfies

$$\mathbf{Q} = \begin{pmatrix} -\sum_{j\neq 1} q_{1j} & q_{12} & \cdots & q_{1l} \\ q_{21} & -\sum_{j\neq 2} q_{2j} & \cdots & q_{2l} \\ \vdots & \vdots & \ddots & \vdots \\ q_{l1} & q_{l2} & \cdots & -\sum_{j\neq l} q_{lj} \end{pmatrix} \tag{5}$$

If T_i is the topic spread time from user *i*, i. e. the time it takes for a topic to be accepted by other users, $P\{T_i > s + t | T_i > s\} = P\{T_i > t\}$ for every $s, t \geq 0$. Therefore, the random variable T_i is *memoryless* and has an exponential distribution with $E(T_i) = 1/v_i$ (Ross, 2010). In a stochastic process, a process that remains in state *i* and then moves to state *j* with probability p_{ij}, will be independent of time as long as it is in state *i*, and p_{ij} satisfies the property:

$$\begin{cases} p_{ii} = 0\,, \forall i \\ \sum_j p_{ij} = 1\,, \forall i \end{cases} \tag{6}$$

The probability of transition from state *i* to state *j* (Song et al., 2007) is

$$p_{ij} = v_i \exp\,(-v_i t_{ij}). \tag{7}$$

Based on Eq. (7) and Eq. (4), the value of q_{ij} can be determined by

$$q_{ij} = {v_i}^2 \exp\,(-v_i t_{ij}) \tag{8}$$

with t_{ij} is the time between user *i* uploading a topic and user *j* tweeting the same topic.

The transition probability matrix estimation is based on the Backward Kolmogorov equation (Ross, 2010),

$$p_{ij}'(t) = \sum_{k\neq i} q_{ik} p_{kj}(t) - v_i p_{ij}(t) \tag{9}$$

or it can be written,

$$p_{ij}'(t) = \sum_k q_{ik} p_{kj}(t). \tag{10}$$

Eq. (10) can be written in matrix form,

$$\mathbf{P}'(\boldsymbol{t}) = \mathbf{QP}(\boldsymbol{t}) \tag{11}$$

By integrating both sides of Eq. (11), we get

$$ln\mathbf{P}(t) = \mathbf{Q}(t) + C \text{ atau } \mathbf{P}(t) = C\exp(\mathbf{Q}t) \tag{12}$$

The solution of Eq. (11) with $\mathbf{P}(0) = \mathbf{I}$ (identity matrix) is

$$\mathbf{P}(t) = \exp\,(\mathbf{Q}t)\ . \tag{13}$$

Then by Taylor expansion, equation (13) can be written in power series, that is

$$\mathbf{P}(t) = \sum_{\boldsymbol{n}=\mathbf{0}}^{\infty} (\mathbf{Q}t)^n / n!\ . \tag{14}$$

For large enough *n*, the transition probability matrix of $\mathbf{P}(t)$ information diffusion model can be estimated by

$$\mathbf{P}(t) = \left(\mathbf{I} + \mathbf{Q}\frac{t}{n}\right)^n\ . \tag{15}$$

3.4 Diffusion Size

The Diffusion size is calculated to rank influencers. Based on the transition probability matrix that has been obtained, the size of the diffusion of user i for a period of time t is determined by multiplying the number of opportunities for information dissemination from user *i* to other users by the number of other users affected by user *i* using the formula

$$D_{i,t} = \sum_j p_{ij}(t).\, n_{ip} \tag{16}$$

where n_{ip} represents the predict of the number of users who tweeted the same topic after user *i* uploaded the tweet. In this paper, n_{ip} is determined by predicting a logarithmic function based on the resulting data, i.e

$$n_{ip} = a + b\log\,(n_i) \tag{17}$$

where a, b are constants and n_i represents the real data of the number of users who tweeted the same topic after user *i* uploaded the tweet. The user with the largest diffusion size value is the top influencer.

3.5 Research flowchart

The steps we carried out to rank influencers using an information diffusion model with a homogeneous time CTMC we depicted in a flowchart as shown in Fig. 3.

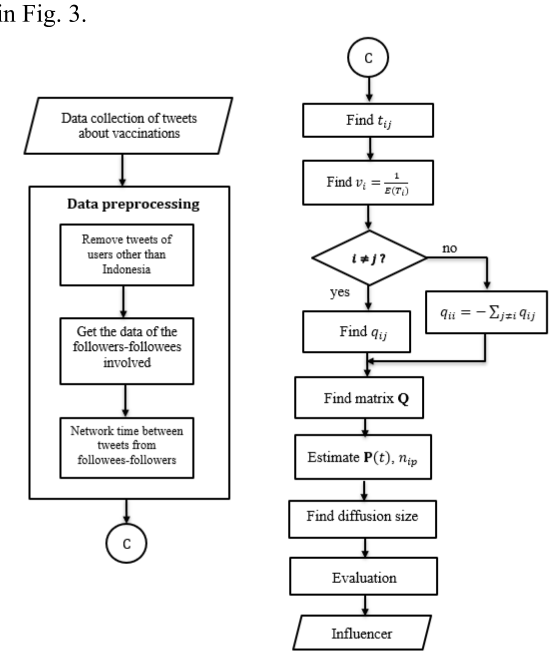


Fig 3. Flowchart of Information Diffusion Model with CTMC

4. Results and Discussion

Information Diffusion Model of Indonesian Twitter data with homogeneous CTMC requires that information is said to diffuse or spread only if Twitter users are bound in a followers-followees network and are involved in discussing a topic. Users involved in tweeting a topic describe the state of the CTMC. Then the time lapse between user i uploading a tweet and then user j who is a follower of i uploads a tweet with the same topic is denoted by t_{ij}. To answer the first research question (RQ1), it is necessary to define the transition rate matrix (in this case it is assumed to be constant) that satisfies Eq. (5). Estimation of transition probability that describes the information dissemination model is used in Eq. (15). After obtaining the transition probability matrix, then to answer the second research question (RQ2), first predict the number of other users affected by the user i following the logarithmic function. To be more precise, in Section 4.1 Dataset Description is discussed, and in Section 4.2. the result of CTMT Information Diffusion Model with selected data set is presented.

4.1 Dataset Description

In this study, we processed data limited to 5,000 Indonesian Twitter users who were involved in the "vaksinasicovid19" tweet with the most followers. The time matrix is calculated between followees' and followers' tweets (obtained a time matrix with a size of 5,000 × 5,000). An example of the time data between consecutive tweets from 10 users with the highest number of followers can be seen in Table 1. The rows in Table 1 represent followees, while the columns represent followers. The number 0 in the i-th row of the j-th column indicates that user j is not a follower of i or user j has not posted a tweet about covid19 vaccination or $i = j$. Users involved in this tweet follower and followee activity are A = jokowi, B = detikcom, C = kompascom, D = ganjarpranowo, E = KPK_RI, F = CNNIndonesia, G = aniesbaswedan, H = KompasTV, I = budimandjatmiko. J = tvOneNews. The time between tweets is expressed in hours.

Table 1
Time between successive tweets about covid19 vaccination between 10 followers-followees (in hours)

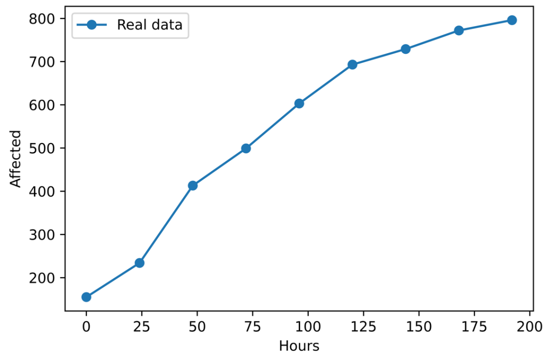
User	A	B	C	D	E	F	G	H	I	J
A	0.00	0.00	0.00	1.40	10.95	0.08	11.94	0.61	0.47	0.63
B	55.79	0.00	0.00	1.37	11.18	0.32	12.17	0.85	0.27	0.48
C	55.91	0.00	0.00	1.48	11.26	1.13	12.25	0.88	0.14	0.67
D	54.42	0.00	0.00	0.00	10.71	4.98	11.70	1.84	0.86	4.52
E	56.07	0.00	0.00	9.33	0.00	14.31	0.99	9.01	8.03	13.85
F	55.85	0.00	0.00	1.42	11.31	0.00	12.29	0.56	0.77	5.87
G	55.08	0.00	0.00	8.34	17.53	13.32	0.00	8.02	7.04	12.86
H	55.37	0.00	0.00	0.94	10.99	5.92	11.98	0.00	0.92	5.46
I	55.79	0.00	0.00	1.17	11.25	6.34	12.24	0.98	0.00	5.88
J	55.82	0.00	0.00	1.40	11.25	1.01	12.24	0.75	0.44	0.00

Table 1 shows the value of t_{ij} which is the time between user i's tweet and user j's tweet as follower i. For example, t_{BA} = 55.79 means that the time difference between follower A tweeting B after tweeting B is 55.79 hours. $t_{\text{AB}} = 0$ means maybe B isn't A's follower or maybe B isn't tweeting after A posted the tweet.

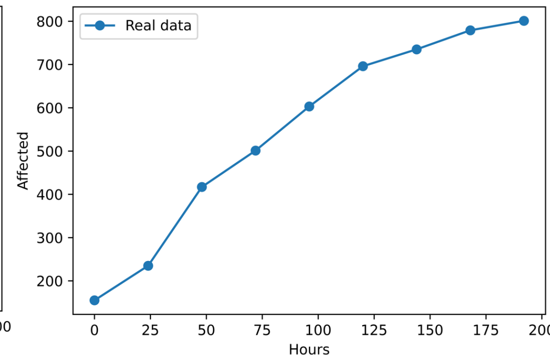
4.2 Information Diffusion Model and Influencer Ranking

From the time matrix data between followees' and followers' tweets, then the average time between tweets of each user is determined. Furthermore, the transition rate matrix or the information dissemination rate $\mathbf{Q}$ is calculated according to Eq. (5) with each entry using Eq. (8). In this case, the transition rate matrix $\mathbf{Q}$ of size 5,000 × 5,000 is obtained. Next, answer RQ1: *How to estimate a transition probability matrix that describes the spread of information from one user to another on Twitter?* In this case, Eq. (15) is used to estimate the transition probability matrix ($\mathbf{P}(t)$). The obtained transition probability matrix $\mathbf{P}(t)$ is 5,000 × 5,000, so it cannot be displayed in this paper. This transition probability matrix describes the probability of information spreading from one user to another.

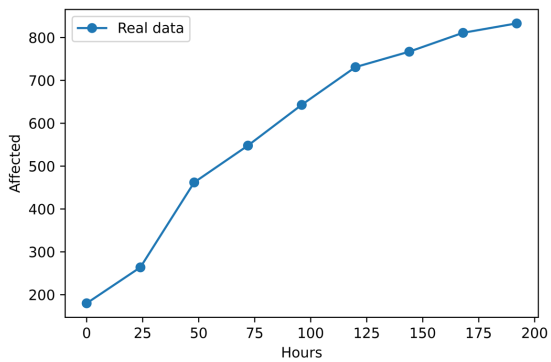
After the transition probability matrix $\mathbf{P}(t)$ is obtained, then answer RQ2: *How to determine the top influencers dynamically over times?*. In this case the top influencer is determined by the value of the diffusion size based on Eqs. (16-17). To obtain the predicted diffusion size, we used 80% of the time of this data as training data. The number of users affected by user i is predicted by describing this data. For example, Fig. 4. shows a plot of data on the number of other users affected by four users selected randomly.



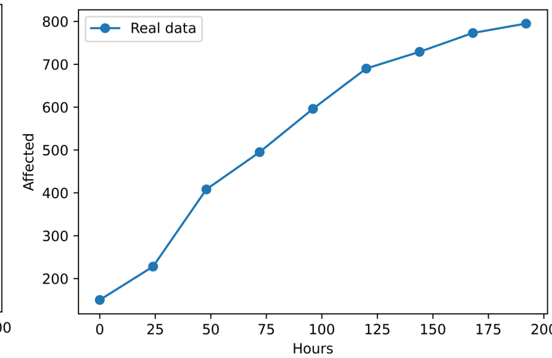
(a) 39th account



(b) 1462th account



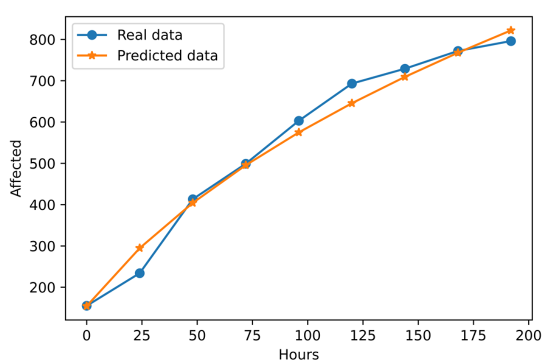
(b) 1778th account



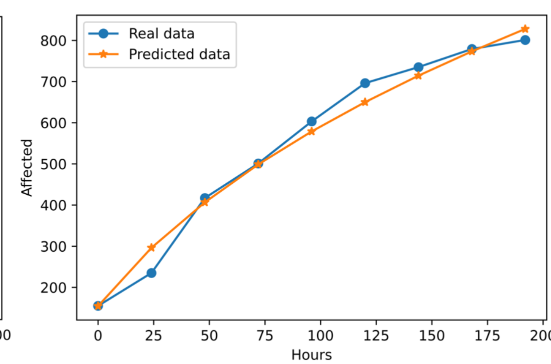
(d) 4345th account

Fig. 4. Number of affected users from randomly selected users

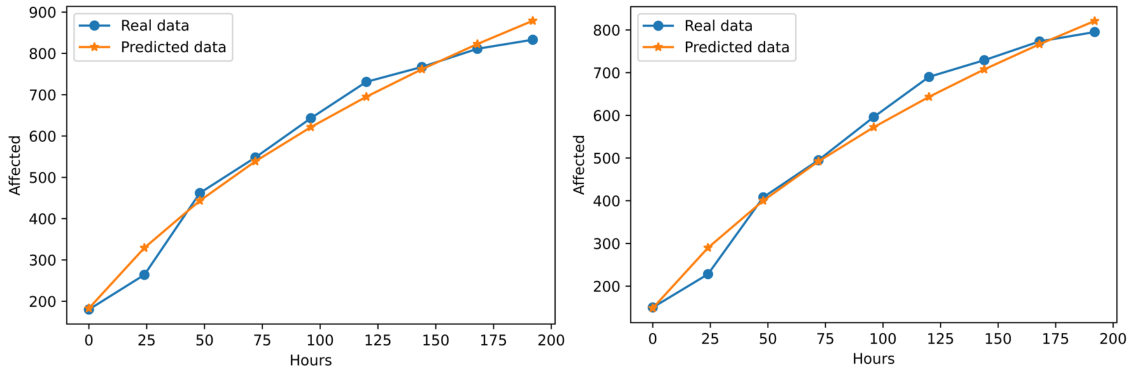
Fig. 4. (a) shows the number of other users affected by the 39th account in the training period, (b) the number of other users affected by the 1462nd account in the training period, (c) the number of other users affected by the 1778th account in the training time span, and (d) the number of other users affected by the 4345th account in the training time span. From Fig. 4. we can see that the number of affected users is a function of time which is similar to a logarithmic function. This training data is used to estimate the values of a and b in Eq. (17). Furthermore, this logarithm function is used to obtain the value of n_{ip} as the predicted value of n_i over time. A visualization of the comparison of real data with the predicted number of affected users from four randomly selected users is shown in Fig. 5.



(a) 39th account



(b) 1462th account



(c) 1778th account (d) 4345th account

Fig. 5. Comparison between real and predicted number of affected data from four randomly selected users

Fig. 5. shows a comparison between the real data and the estimated number of other users affected during the training time by (a) the 39th account, (b) the 1462nd account, (c) the 1778th account and (d) the 4345th account. Referring to Lawrence et al. (2009), the model's performance in predicting the number of affected users generated by this logarithmic function is said to be accurate with MAPE 14.8%.

We ended the process by calculating the diffusion size for several different times, namely day 1 (t=24), day 3 (t=72), day 5 (t=120) and day 8 (t=192). Table 2 shows the top 20 users by number of followers and ranking of influencers by spread size with CTMC for each time.

Table 2
Influencer ranking by diffusion size with CTMC

#	most followers	Influencer	Influencer	Influencer (t=120)	Influencer (t=192)
1	jokowi	Kemdikbud_RI	belindch	uradn	lensaRTV
2	detikcom	MasTBP_	womencoalition	R_Uddarojat	R_Uddarojat
3	kompascom	BiasaAgung	agama_nusantara	tercelup	suarasidoarjofm
4	ganjarpranowo	mummtaz_qadhifa	andienas	agama_nusantara	InfoSeJogja
5	KPK_RI	gunoto7	akuluthfi	SaintChyril	_wafflepancake
6	CNNIndonesia	mprgoid	SaintChyril	Adiitoo	NUCreativeMedia
7	aniesbaswedan	andreasharsono	Bayu_99091	_wafflepancake	radarbogorID
8	KompasTV	mangkuthinks	Ignastherry	IanSPCC	NrimoBerkah
9	budimandjatmiko	evndari	nickoakbar	lensaRTV	CikManisDonut
10	tvOneNews	sendysuwanto	ExcelJosaphine	Sakamilenialjtg	Valkyreism
11	501Awani	bulanrudrigo	WahyuDhyatmika	belindch	ictss_iium
12	kumparan	elHurryKoRn_2	vivi3284	NrimoBerkah	BlackcoTanya
13	Metro_TV	FirzaHusain	shandya	Ignastherry	kalilawar_16
14	tempodotco	kutang_bolong	BaleBengong	radarbogorID	april_hamsa
15	DivHumas_Polri	mahesatiwi	tercelup	goldeneagle2707	Sudutkota5
16	KemenkesRI	Dhinijogja	suradiparlan5	vivi3284	Adiitoo
17	DGHisham	Lionelbarcalope	Scortariuz	NUCreativeMedia	capkucing
18	TirtoID	kitkitttin	kemendagri	kamalacempaka	yudiseti
19	liputan6dotcom	4rios16	germantantra	Bayu_99091	FajarRamdhn972
20	republikaonline	Morodi5	Gelangalit4	Gelangalit4	JakartaGreater

The second column in Table 2 states the ranking of users based on the number of followers. Furthermore, the third until the sixth columns show the order of influencers for different times. It can be seen that the user with the account name "Jokowi" is the user with the highest number of followers but does not guarantee that he will become an influencer. On the first day it is predicted that "Kemdikbud_RI" will be the top influencer, while on the 5th day "uradn" will be the top influencer, and on the 8th day it will be the top influencer " lensaRTV ".

In this paper, the Homogeneous CTMC is adopted. This means that the transition rate or information dissemination time in Eq. (5) is assumed to be constant. However, the phenomenon of information dissemination is actually not constant, but is a function of time (nonhomogeneous). This motivate us to conduct further research, namely the diffusion of information on Twitter will be modeled with a non-homogeneous CTMC.

5. Conclusion

In this paper, information diffusion or information dissemination is described by a homogeneous CTMC model. In this case the transition rate of information dissemination is assumed to be constant. The probability of information dissemination is predicted with a transition opportunity matrix. Through this CTMC model, influencer rankings have been determined dynamically over time with predictions of affected users following a logarithmic function. Diffusion size for different times give different influencer ratings. From the influencer rankings obtained, the most influential users are not the users with the most followers.

Acknowledgement

We thank the Rector of the Universitas Padjadjaran, Indonesia, who have provided funding via the Universitas Padjadjaran Doctoral Dissertation Research Grant (RDDU), with a contract number: 1427/UN6.3.1/LT/2020). We are grateful for the discussion on social media analytics study with the collaborators in the RISE-SMA H2020 Project.

References

Alp, Z. Z., & Öğüdücü, Ş. G. (2019). Influence Factorization for identifying authorities in Twitter. *Knowledge-Based Systems*, *163*, 944–954.

Bhowmick, A. (2019). Temporal Sequence of Retweets Help to Detect Influential Nodes in Social Networks. *IEEE Transactions on Computational Social Systems*, *6*(3), 441–455.

Fernando, S., Amador Díaz López, J., Şerban, O., Gómez-Romero, J., Molina-Solana, M., & Guo, Y. (2019). Towards a large-scale twitter observatory for political events. *Future Generation Computer Systems*, *110*(3).

Firdaniza, F., Ruchjana, B. N., Chaerani, D., & Radianti, J. (2022). Information Diffusion Model in Twitter : A Systematic Literature Review. *Information*,*13*(1).

Firdaniza, Ruchjana, B. N., & Chaerani, D. (2021). Information diffusion model using continuous time Markov chain on social media. *Journal of Physics: Conference Series*, *1722*(1).

Firdaus, S. N., Ding, C., & Sadeghian, A. (2018). Retweet: A popular information diffusion mechanism – A survey paper. *Online Social Networks and Media*, *6*, 26–40.

Hamzehei, A., Jiang, S., Koutra, D., Wong, R., & Chen, F. (2017). Topic-based social influence measurement for social networks. *Australasian Journal of Information Systems*, *21*.

Han, J., Kamber, M., and Pei,J., (2012) '*Data Mining: Concept and Techniques*'. Elsevier,Inc.

Jiang, C., Chen, Y., & Liu, K. J. R. (2014). Evolutionary dynamics of information diffusion over social networks. *IEEE Transactions on Signal Processing*, *62*(17), 4573–4586.

Kawamoto, T. (2013). A stochastic model of tweet diffusion on the Twitter network. *Physica A: Statistical Mechanics and Its Applications*, *392*(16), 3470–3475. https://doi.org/10.1016/j.physa.2013.03.048

Kim, Y., & Seo, J. (2020). Detection of Rapidly Spreading Hashtags via Social Networks. *IEEE Access*, *8*, 39847–39860.

Kumar, S., Saini, M., Goel, M., & Aggarwal, N. (2020). Modeling Information Diffusion in Online Social Networks Using SEI Epidemic Model. *Procedia Computer Science*, *171*(2019), 672–678.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? *Proceeding of the 19th International Conference on World wide web, 591-600.*

Kwon, J. (2017). Effects of source influence and peer referrals on information diffusion in Twitter. In *Industrial Management and Data Systems* (Vol. 117, Issue 5, 896–909).

Lawrence, K.D., Klimberg, R.K. and Lawrence, S.M. (2009) *Fundamentals of Forecasting Using Excel*, *Australian Health Service Alliance*.

Li, J., Peng, W., Li, T., Sun, T., Li, Q., & Xu, J. (2014). Social network user influence sense-making and dynamics prediction. *Expert Systems with Applications, 41*, 5115-5124.

Mittal, D., Suthar, P., Patil, M., Pranaya, P. G. S., Rana, D. P., & Tidke, B. (2020). Social Network Influencer Rank Recommender Using Diverse Features from Topical Graph. *Procedia Computer Science*, *167*(2019), 1861–1871.

Oo, M. M., & Lwin, M. T. (2020). Detecting Influential Users in a Trending Topic Community Using Link Analysis Approach. *International Journal of Intelligent Engineering and Systems*, *13*(6), 178–188.

Peng, S., Zhou, Y., Cao, L., Yu, S., Niu, J., & Jia, W. (2018). Influence analysis in social networks: A survey. *Journal of Network and Computer Applications*, *106*, 17–32.

Qasem, Z., Jansen, M., Hecking, T., & Hoppe, H. U. (2017). Using attractiveness model for actors ranking in social media networks. *Computational Social Networks*, *4*(1).

Riquelme, F., & González-Cantergiani, P. (2016). Measuring user influence on Twitter: A survey. *Information Processing & Management,* 1-27.

Ross, S. M. (2010). *Introduction to Probability Models* (10th ed.). Elsevier Inc.

Shuai, X., Ding, Y., Busemeyer, J., Chen, S., Sun, Y., & Tang, J. (2012). Modeling indirect influence on Twitter. *International Journal on Semantic Web and Information Systems*, *8*(4), 20–36.

Simmie, D., Vigliotti, M. G., & Hankin, C. (2014). Ranking twitter influence by combining network centrality and influence observables in an evolutionary model. *Journal of Complex Networks*, *2*(4), 495-517.

Song, X., Chi, Y., Hino, K., & Tseng, B. L. (2007). Information flow modeling based on diffusion rate for prediction and ranking. *16th International World Wide Web Conference, WWW2007*, 191–200.

Tidke, B., Mehta, R., & Dhanani, J. (2020). Consensus-based aggregation for identification and ranking of top-k influential nodes. *Neural Computing and Applications*, *32*(14), 10275–10301.

Varshney, D., Kumar, S., & Gupta, V. (2017). Predicting information diffusion probabilities in social networks: A Bayesian networks based approach. *Knowledge-Based Systems*, *133*, 66–76.

Yoo, E., Rand, W., Eftekhar, M., & Rabinovich, E. (2016). Evaluating information diffusion speed and its determinants in social media networks during humanitarian crises. *Journal of Operations Management, 45*, 123-133.

Zhang, L., Luo, M., & Boncella, R. J. (2020). Product information diffusion in a social network. *Electronic Commerce Research*, *20*(1), 3–19.

Zheng, Z., Yang, H., Fu, Y., Fu, D., Podobnik, B., & Stanley, H. E. (2018). Factors influencing message dissemination through social media. *Physical Review E*, 97(6), 1–6.

Zhu, T., Wang, B., Wu, B., & Zhu, C. (2014). Maximizing the spread of influence ranking in social networks. *Information Sciences*, 278, 535–544.

https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/acces accessed on December 31, 2021