International Journal of Data and Network Science 7 (2023) 1595–1602

Contents lists available at GrowingScience

International Journal of Data and Network Science

homepage: www.GrowingScience.com/ijds

Non-homogeneous continuous time Markov chain model for information dissemination 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*

CHRONICLE

Article history:
Received: July 5, 2023
Received in revised format: July 25, 2023
Accepted: August 7, 2023
Available online: August 7, 2023

Keywords:
Information dissemination
Twitter
Non-homogeneous continuous time Markov chain
Sigmoid function
Maximum likelihood estimation

ABSTRACT

Nonhomogeneous Continuous-Time Markov Chain (NH-CTMC) is a stochastic process that can be used to model problems where the future state depends only on the current state and is independent of the past. The transition intensity in NH-CTMC is not constant but is a function of time. In this paper, NH-CTMC is employed to model information dissemination on Twitter, where transitions occur only from followee groups to follower groups. Information is considered spread on Twitter when followers retweet posts from their followees. The tweet-retweet process on Twitter satisfies the Markov property, as a retweet from a follower depends only on the tweet posted just before by the corresponding followee. The probability of a tweet spreading is determined by the transition intensity, assumed to be a Sigmoid function whose parameters are estimated using Maximum Likelihood Estimation (MLE). This method is applied to Twitter data from Indonesia related to discussions on Covid-19 vaccination. The results indicate that information about Covid-19 vaccination on Twitter spreads rapidly from followees to followers in the first 20 hours, and then slows down after 40 hours. The NH-CTMC model outperforms the Homogeneous Continuous-Time Markov Chain (H-CTMC) approach, where the transition intensity (tweet spreading intensity) is assumed to be constant.

1. Introduction

A Markov chain is defined as a stochastic process with the property of conditional distribution of future events given the current and past events, relying solely on the present state and being independent of the past (memoryless) (Bremaud, 2020). This property is known as the Markov property. Markov chains have been widely used by researchers to predict a state influenced only by the preceding state. The Continuous-Time Markov Chain (CTMC) model can be used to observe transitions that occur in very small time intervals (Begun et al., 2013). CTMC itself is classified into Homogeneous CTMC (H-CTMC) and Nonhomogeneous CTMC (NH-CTMC). In practice, H-CTMC is more commonly used in modeling due to the assumption of homogeneity, where the transition rates are assumed constant, making it easier to solve (Ocaña-Riola, 2005). Generally, H-CTMC is used in medical problems, such as the clinical course of chronic diseases like cancer analysis (Kalbfleisch and Lawless, 1985; Ocaña-Rilola, 2002), and disease progression in poultry (Trajstman, 2002). Though limited, H-CTMC is also applied to certain issues in social media information dissemination (Li et al., 2014; Zhu et al., 2014; Firdaniza et al., 2022b). However, not all cases are suitable for modeling with H-CTMC, as in reality, the transition intensities may depend on time.

The information dissemination on Twitter can be modeled using a Markov chain because it satisfies the Markov property, where a user's tweet dissemination depends only on the tweets they received previously and is independent of the tweet history.

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

ISSN 2561-8156 (Online) - ISSN 2561-8148 (Print)

doi: 10.5267/j.ijdns.2023.8.004

If user A follows user B, then A is considered a follower of B, and B is referred to as A's followee (Firdaus et al., 2018). Information is said to be disseminated if a tweet posted by a followee is retweeted by their follower. The use of Markov models in information spread phenomena on Twitter is still relatively uncommon (Firdaniza et al., 2022a). In Firdaniza et al. (2022a), it is mentioned that studies on information dissemination or information diffusion on Twitter have been conducted using various methods such as the Quantum q-attention model (Shuai et al., 2012), Poisson regression (Kwon et al., 2017), Time decay features cascade model (Agarwal and Mehta, 2020), Susceptible-Infected-Recovered (SIR) model (Zheng et al., 2018), and Susceptible-Exposed-Infected (SEI) model (Kumar et al., 2020), as well as Bayesian networks (Varshney et al., 2017). Only one article, Li et al. (2014), discusses information dissemination on Twitter using continuous-time Markov chain. Furthermore, Firdaniza et al. (2022b) used H-CTMC to model information diffusion among Twitter users, where the transition intensity of information dissemination on Twitter users, representing the state of CTMC, is assumed constant. However, this assumption does not truly reflect reality, as the rate of information spread depends on time, making the NH-CTMC model more appropriate. Nonetheless, unlike H-CTMC, the NH-CTMC model is quite complex and challenging. Determining transition intensities is not as straightforward as in the H-CTMC model.

Several researchers have employed different approaches to estimate transition intensities. For instance, Hsieh et al. (2002) assumed the transition intensity to be a piecewise exponential function, while Ocaña-Riola (2005) assumed it to be a piecewise constant function. However, these assumptions have a drawback in the interval partitioning, leading to drastic changes in the intensity matrix. Subsequently, Hubbard et al. (2008) utilized time transformation to make the process homogeneous in the new time domain. The weakness of this model lies in the non-homogeneity, as the Markov chain's non-homogeneity will be the same for every state. Titman (2011) then assumed the transition intensity function to be a Bspline function and estimated the parameters using maximum likelihood estimation (MLE). All these NH-CTMC models have been previously used by researchers in medical issues, where the Markov chain states represent patients' treatment statuses, such as assessing the progression of chronic diseases (Hsieh et al., 2002), biomedical data analysis (Ocaña-Riola, 2005), delirium in cancer patients (Hubbard et al., 2008), spiking neuron response (Tapson, 2009), post-heart transplant patient disease progression (Titman 2011), and kidney disease development (Begun et al., 2013). However, so far, no NH-CTMC research has been found regarding information dissemination on Twitter (Firdaniza, 2022a).

This paper discusses the NH-CTMC model for information dissemination on Twitter data. The contribution of this paper lies in determining the transition intensity matrix as a function of time and its application to the problem of information dissemination on Twitter. Here, the transition intensity is assumed to be a family of Sigmoid functions, and the parameters are estimated using Maximum Likelihood Estimation (MLE). Furthermore, this NH-CTMC model is applied to analyze the dissemination of Covid-19 vaccination information on Twitter in Indonesia. In this study, it is assumed that the two states in NH-CTMC are represented by the followee and follower groups of Twitter users. Lastly, a comparison of the performance of the NH-CTMC and H-CTMC models in information dissemination on Twitter is also presented in the discussion.

2. Problem Description

The tweet-retweet process that occurs on Twitter heavily depends on the tweets posted by Twitter users. If a Twitter user, the followee, posts a tweet and it is subsequently retweeted by their follower, then it is considered that the information has been disseminated. The probability of information spreading from one user to another on Twitter can be determined through the intensity of information dissemination. Firdaniza et al. (2022b) have discussed a model for information diffusion on Twitter using H-CTMC, where the intensity of information dissemination is assumed to be constant. However, this assumption does not align with reality, as the intensity of information dissemination is not constant but varies over time, necessitating the use of the NH-CTMC model in this phenomenon.

The research questions in this study are as follows:
1. How to estimate the transition rate matrix and transition probability matrix that describe the dissemination of information from one user to another on Twitter?
2. How does the NH-CTMC model compare to the H-CTMC model for the problem of information dissemination on Twitter?

3. Literature Review

Referring to Bremaud (2020), a stochastic process $\{X(t), t \geq 0\}$ is said to be a CTMC) if it satisfies the Markov property, which means that for any $s, t \geq 0$ and non-negative integer $i, j, x(\gamma), 0 \leq \gamma < s$, the following holds,

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

In NH-CTMC, it holds that $P\{X(s+t) = j | X(s) = i\} = p_{ij}(s,t)$, meaning that the transition probability from state i to state j at time t will depend on the starting point s. Determining the transition probability matrix heavily relies on the transition intensity matrix. The transition intensity is not constant but varies with time (as a function of t). In this case, the transition probability matrix needs to be estimated for each time t.

For small time h, the transition intensity satisfies (Ocaña-Riola, 2005),

$$q_{ij}(t) = \lim_{h \to 0} \frac{p_{ij}(t, t+h)}{h} \quad (2)$$

and

$$v_i(t) = \lim_{h \to 0} \frac{1 - p_{ii}(t, t+h)}{h}. \quad (3)$$

The transition intensity matrix $\mathbf{Q}(t)$ with entries $q_{ij}(t)$ of size $(K \times K)$ can be written in the form of

$$\mathbf{Q}(t) = \begin{bmatrix} q_{11}(t) & q_{12}(t) & \dots & q_{1K}(t) \\ q_{21}(t) & q_{22}(t) & \dots & q_{2K}(t) \\ \vdots & \vdots & \vdots & \vdots \\ q_{K1}(t) & q_{K2}(t) & \dots & q_{KK}(t) \end{bmatrix} \quad (4)$$

under the following conditions:

(i) The off-diagonal entries are non-negative; $q_{ij}(t) \geq 0$ untuk $i \neq j$.

(ii) The sum of each row's entries is zero; $\sum_{j=1}^{K} q_{ij}(t) = 0$.

As a result, the entries for the diagonal of the transition intensity matrix are $-\sum_{i,j=1}^{K} q_{ij}(t), i \neq j$ and $v_i(t) = -q_{ii}(t) = \sum_{i,j=1}^{K} q_{ij}(t), i \neq j$.

The transition probability of NH-CTMC is the solution of the Forward Kolmogorov equation (Titman, 2011),

$$\frac{d\mathbf{P}(t_0,t)}{dt} = \mathbf{P}(t_0, t)\mathbf{Q}(t) \text{ dan } \mathbf{P}(t_0, t_0) = \mathbf{I} \quad (5)$$

where $\mathbf{P}(t_0, t)$ is the transition probability matrix with entries $p_{ij}(t_0, t)$. Equation (5) must be satisfied for every different value of t. This transition probability matrix $\mathbf{P}(t_0, t) = [p_{ij}(t_0, t)]$ can be determined by first estimating the transition intensity matrix $\mathbf{Q}(t) = [q_{ij}(t)]$.

4. Results and Discussion

4.1 Transition intensity in NH-CTMC

In this study, the entries of the transition intensity matrix $q_{ij}(t)$ are assumed to be a family of Sigmoid Functions. Based on the phenomenon of information dissemination on Twitter, the transition rate or rate of information dissemination on Twitter will decrease over time, so the function used as the assumption of transition intensity in this study is

$$q_{ij}(t) = \frac{e^{\alpha_{ij}}}{e^{\beta_{ij}} + e^{\gamma_{ij} t}} \quad (6)$$

with $\alpha_{ij}, \beta_{ij}, \gamma_{ij}$ are the parameters of the function estimated using MLE.

Suppose there are N individuals observed at time t_{ij}, $i = 1,2, \dots, N; j = 1,2, \dots, J_i$. $X_i(t_{ij})$ is the state of individual i at time $t_{ij.}$ $\boldsymbol{\theta} = (q_{12}(t), q_{23}(t), \dots, q_{kk-1}(t))$ is the transition intensity vector.

The Likelihood function for N individuals is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N} \prod_{j=1}^{J_i} p_{X_i(t_{ij-1}), X_i(t_{ij})}(t_{ij-1}, t_{ij}; \boldsymbol{\theta}) \quad (7)$$

where

$$p_{X_i(t_{ij-1}), X_i(t_{ij})}(t_{ij-1}, t_{ij}; \boldsymbol{\theta}) = P\{X_i(t_{ij}) = x_i | X_i(t_{ij-1}) = x_{ij-1}; \boldsymbol{\theta}\},$$

In this case, $p_{X_i(t_{ij-1}),X_i(t_{ij})}(t_{ij-1},t_{ij};\boldsymbol{\theta}) = p_{ij}(t_0,t;\boldsymbol{\theta})$ is the entry of the transition probability matrix $\mathbf{P}(t_0,t;\boldsymbol{\theta})$ that satisfies the initial value problem,

$$\frac{d\mathbf{P}(t_0,t;\boldsymbol{\theta})}{dt} = \mathbf{P}(t_0,t;\boldsymbol{\theta})\mathbf{Q}(t;\boldsymbol{\theta}) \text{ dan } \mathbf{P}(t_0,t_0;\boldsymbol{\theta}) = \mathbf{I}. \tag{8}$$

The solution of $\mathbf{P}(t_0,t;\boldsymbol{\theta})$ is determined numerically, by substituting the intensity function used. To obtain an estimate of $\boldsymbol{\theta}$, the first and second partial derivatives of the Log-likelihood function of Eq. (7) are used,

$$\ell = ln\, L(\boldsymbol{\theta}) = \sum_{i=1}^{N}\sum_{j=1}^{J_i} \ln\left(p_{X_i(t_{ij-1}),X_i(t_{ij})}(t_{ij-1},t_{ij};\boldsymbol{\theta})\right) \tag{9}$$

Then $\boldsymbol{\theta}$ is determined iteratively with the Fisher scoring algorithm. The estimated parameter $\widehat{\boldsymbol{\theta}}$ at the $m+1$ iteration is derived from the second-degree Taylor series of the Log-Likelihood function $\ell(\boldsymbol{\theta})$ around $\widehat{\boldsymbol{\theta}}^{(1)}$ with the following steps:

1. Enter the initial estimate $\widehat{\boldsymbol{\theta}}^{(1)}$.
2. Iteration starts from $m = 1$.
3. Determine the first partial derivatives of all estimated parameters

$$X(\widehat{\boldsymbol{\theta}}^{(m)})_{z\times 1} = \begin{bmatrix} \frac{\partial \ell}{\partial \theta_1{}^{(m)}} \\ \frac{\partial \ell}{\partial \theta_2{}^{(m)}} \\ \vdots \\ \frac{\partial \ell}{\partial \theta_z{}^{(m)}} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N}\sum_{j=1}^{J_i} \frac{\frac{\partial}{\partial\theta_1} p_{X_i(t_{ij-1}),X_i(t_{ij})}(t_{ij-1},t_{ij};\boldsymbol{\theta})}{p_{X_i(t_{ij-1}),X_i(t_{ij})}(t_{ij-1},t_{ij};\boldsymbol{\theta})} \\ \sum_{i=1}^{N}\sum_{j=1}^{J_i} \frac{\frac{\partial}{\partial\theta_2} p_{X_i(t_{ij-1}),X_i(t_{ij})}(t_{ij-1},t_{ij};\boldsymbol{\theta})}{p_{X_i(t_{ij-1}),X_i(t_{ij})}(t_{ij-1},t_{ij};\boldsymbol{\theta})} \\ \vdots \\ \sum_{i=1}^{N}\sum_{j=1}^{J_i} \frac{\frac{\partial}{\partial\theta_z} p_{X_i(t_{ij-1}),X_i(t_{ij})}(t_{ij-1},t_{ij};\boldsymbol{\theta})}{p_{X_i(t_{ij-1}),X_i(t_{ij})}(t_{ij-1},t_{ij};\boldsymbol{\theta})} \end{bmatrix}$$

where z denotes the number of parameters being estimated.

4. Determines the inverse of the matrix $\Im(\widehat{\boldsymbol{\theta}}^{(m)}) = \left[E\left(-\frac{\partial^2 \ell}{\partial\theta_r\partial\theta_s}\right)\right]$ by

$$E\left(-\frac{\partial^2 \ell}{\partial\theta_r\partial\theta_s}\right) = \sum_{i=1}^{N}\sum_{j=1}^{J_i} \frac{\frac{\partial}{\partial\theta_r} p_{X_i(t_{ij-1}),\ X_i(t_{ij})}(t_{ij-1},t_{ij};\boldsymbol{\theta}) \frac{\partial}{\partial\theta_s} p_{X_i(t_{ij-1}),X_i(t_{ij})}(t_{ij-1},t_{ij};\boldsymbol{\theta})}{p_{X_i(t_{ij-1}),X_i(t_{ij})}(t_{ij-1},t_{ij};\boldsymbol{\theta})}.$$

5. Determines the estimate of $\widehat{\boldsymbol{\theta}}$ at the $m+1$ by iteration

$$\widehat{\boldsymbol{\theta}}^{(m+1)} = \widehat{\boldsymbol{\theta}}^{(m)} + X(\widehat{\boldsymbol{\theta}}^{(m)})\Im(\widehat{\boldsymbol{\theta}}^{(m)})^{-1}$$

6. Iteration continues until $\left|\widehat{\boldsymbol{\theta}}^{(m+1)} - \widehat{\boldsymbol{\theta}}^{(m)}\right| < \varepsilon$, with ε a certain small number.

From this process, the parameter $\boldsymbol{\theta}$ of the transition intensity matrix $\mathbf{Q}(t)$ is obtained and then the transition probability matrix $\mathbf{P}(t_0,t)$)) is determined numerically using the Finite Difference method. For efficient calculation of $\mathbf{P}(t_j,t)$ with arbitraryt_j, it is calculated using Chapman Kolmogorov, $\mathbf{P}(t_0,t) = \mathbf{P}(t_0,t_j)\mathbf{P}(t_j,t)$ such that

$$\mathbf{P}(t_j,t) = \mathbf{P}(t_0,t_j)^{-1}\mathbf{P}(t_0,t). \tag{10}$$

4.2 Data Description

The data used in this study is tweet data from Indonesian Twitter users who talk about the topic of Covid19 vaccination taken from https://netlytic.org/ (Firdaniza et al., 2022b). The number of tweets that occur between Twitter users is large, so in NH-CTMC it is not possible for all users to be declared as Markov chain states. This is because every move that occurs is a function of time, as a result there are a lot of parameters involved and it is very difficult to solve.

In this study, it is assumed that the transfer occurs only from the followee group to the follower group, meaning that in this NH-CTMC model of information dissemination on Twitter there are two states, namely the follower group and the followee group. Data on information transfer (tweet-retweet) from the followee group to the follower group of 20 Twitter users is seen based on a 20×20 matrix of time between tweets by taking entries that are above the main diagonal (see data in Firdaniza et al., 2022b). The retweet data of these 20 Twitter users are expressed in 145 IDs, indicating the number of transitions that occur is 145 (N = 145). The retweet times of the 20 followers are shown in Table 1.

Table 1
Retweet time of 20 followers (in hours)

	A	B	C	D	E	F	G	H	I	J	K	L	M	O	P	Q	R	S	T
A		0.00	0.00	1.40	10.95	0.08	11.94	0.61	0.47	0.63	2.84	0.00	12.48	0.00	30.29	0.34	16.24	57.31	2.50
B			0.00	1.37	11.18	0.32	12.17	0.85	0.27	0.48	3.07	0.00	13.02	0.00	30.51	0.58	16.48	57.85	2.72
C				1.48	11.26	1.13	12.25	0.88	0.14	0.67	3.14	0.00	13.18	0.00	30.59	0.65	15.53	58.01	2.18
D					10.71	4.98	11.70	1.84	0.86	4.52	4.06	0.00	14.31	0.00	30.04	1.78	17.47	59.14	1.40
E						14.31	0.99	9.01	8.03	13.85	11.23	0.00	4.40	0.00	19.33	8.95	24.64	49.23	10.74
F							12.29	0.56	0.77	5.87	2.75	0.00	12.79	0.00	30.64	1.69	16.16	57.61	2.45
G								8.02	7.04	12.86	10.24	0.00	3.41	0.00	36.86	7.96	23.65	48.24	9.75
H									0.92	5.46	4.13	0.00	14.37	0.00	30.32	1.84	17.53	59.20	1.89
I										5.88	3.20	0.00	13.45	0.00	30.58	0.92	16.61	58.28	2.77
J											3.17	0.00	13.09	0.00	30.58	0.68	16.58	57.92	2.54
K												0.00	22.24	0.00	37.17	17.65	13.41	67.07	18.37
L													13.23	0.00	28.58	0.70	16.76	58.06	2.86
M														0.00	47.71	18.88	34.57	44.83	20.67
N															30.17	0.40	16.30	57.72	2.55
O																19.48	35.17	42.72	21.27
P																	16.92	57.36	2.16
Q																		89.24	33.67
R																			26.53
S																			2.34
T																			0

4.3 Information dissemination on Twitter data with NH-CTMC

In this study there are 145 IDs with two states (followee group and follower group), so the Log-likelihood function of the information dissemination model on Twitter is in the form of

$$\ell = \prod_{i=1}^{145} \prod_{j=1}^{J_i} ln\left(p_{X_i\left(t_{ij-1}\right),X_i\left(t_{ij}\right)}\left(t_{ij-1}, t_{ij}; \boldsymbol{\theta}\right)\right). \tag{11}$$

The transition intensity is assumed to be a family of Sigmoid Functions, i.e.

$$\mathbf{Q}(t) = \begin{pmatrix} -q_{12}(t) & q_{12}(t) \\ 0 & 0 \end{pmatrix} \tag{12}$$

With

$$q_{12}(t) = \frac{e^{\alpha}}{e^{\beta}+e^{\gamma t}} \quad . \tag{13}$$

Based on Eq. (8) and with the transition intensity matrix as in equation (12), it can be written

$$\begin{pmatrix} \frac{d}{dt}\left(p_{11}(t_0,t;\boldsymbol{\theta})\right) & \frac{d}{dt}\left(p_{12}(t_0,t;\boldsymbol{\theta})\right) \\ \frac{d}{dt}\left(p_{21}(t_0,t;\boldsymbol{\theta})\right) & \frac{d}{dt}\left(p_{22}(t_0,t;\boldsymbol{\theta})\right) \end{pmatrix} = \begin{pmatrix} p_{11}(t_0,t;\boldsymbol{\theta}) & p_{12}(t_0,t;\boldsymbol{\theta}) \\ p_{21}(t_0,t;\boldsymbol{\theta}) & p_{22}(t_0,t;\boldsymbol{\theta}) \end{pmatrix} \begin{pmatrix} -q_{12}(t;\boldsymbol{\theta}) & q_{12}(t;\boldsymbol{\theta}) \\ 0 & 0 \end{pmatrix} \tag{14}$$

and $\begin{pmatrix} p_{11}(t_0,t_0;\boldsymbol{\theta}) & p_{12}(t_0,t_0;\boldsymbol{\theta}) \\ p_{21}(t_0,t_0;\boldsymbol{\theta}) & p_{22}(t_0,t_0;\boldsymbol{\theta}) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Substituting Eq. (13) into Eq. (14) yields a system of differential equations

$$\begin{aligned} \frac{d}{dt}\left(p_{11}(t_0,t;\alpha,\beta,\gamma)\right) &= -p_{11}(t_0,t;\alpha,\beta,\gamma)\ \frac{e^{\alpha}}{e^{\beta}+e^{\gamma t}}\ , \\ \frac{d}{dt}\left(p_{12}(t_0,t;\alpha,\beta,\gamma)\right) &= p_{11}(t_0,t;\alpha,\beta,\gamma)\ \frac{e^{\alpha}}{e^{\beta}+e^{\gamma t}}\ , \\ \frac{d}{dt}\left(p_{21}(t_0,t;\alpha,\beta,\gamma)\right) &= -p_{21}(t_0,t;\alpha,\beta,\gamma)\ \frac{e^{\alpha}}{e^{\beta}+e^{\gamma t}}\ , \\ \frac{d}{dt}\left(p_{22}(t_0,t;\alpha,\beta,\gamma)\right) &= p_{21}(t_0,t;\alpha,\beta,\gamma)\ \frac{e^{\alpha}}{e^{\beta}+e^{\gamma t}}\ . \end{aligned} \tag{15}$$

Next, the first partial derivative of the system of Eq. (15) is determined with respect to the three parameters α, β, γ. The intensity function in Eq. (13) and its derivatives are written as function definitions in R software. The resulting $\left(p_{ij}(t_0,t;\alpha,\beta,\gamma)\right)$ is then substituted into the Log-likelihood Eq. (11). Then the parameters α, β, γ are estimated by Fisher scoring algorithm to get $\mathbf{Q}(t)$. Furthermore, the transition probability of the process being in a certain state is calculated using the Finite Difference method,

$$\frac{d\mathbf{P}(t_0,t_j)}{dt} = \frac{h}{2}\left[\mathbf{P}(t_0,t_{j+1}) - \mathbf{P}(t_0,t_{j-1})\right]. \tag{16}$$

The calculation of estimated parameters and transition probability was carried out with the help of R software, and obtained the value $\alpha = -2{,}09203$; $\beta = -1245{,}927$; $\gamma = 0{,}020737$, so that the transition intensity matrix of NH-CTMC was produced in the form of

$$\mathbf{Q}(t) = \begin{pmatrix} -\frac{e^{-2{,}09203}}{e^{-1245{,}927}+e^{0{,}020737\,t}} & \frac{e^{-2{,}09203}}{e^{-1245{,}927}+e^{0{,}020737\,t}} \\ 0 & 0 \end{pmatrix}. \tag{17}$$

and the transition intensity from state 1 to state 2 are shown in Fig. 1.

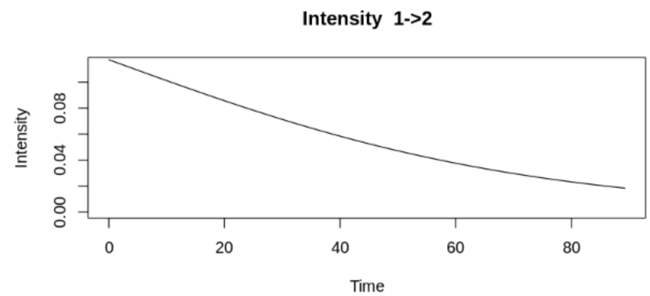

Fig. 1. Transition intensity for function $q_{12}(t) = \frac{e^{\alpha}}{e^{\beta}+e^{\gamma t}}$

It can be seen from Fig. 1 that the intensity of information dissemination decreases gradually over time.

Furthermore, the transition probability which is the description of information dissemination on Twitter with NH-CTMC is shown in Fig. 2.

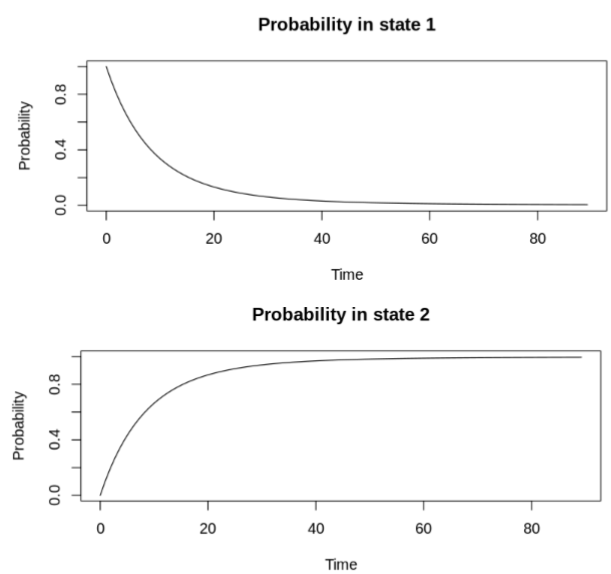

Fig. 2. Transition probability by using the function $q_{12}(t) = \frac{e^{\alpha}}{e^{\beta}+e^{\gamma t}}$

Based on Fig. 2, it can be seen that information on Twitter about Covid-19 vaccination spread rapidly from followee to follower in the first 20 hours, then began to slow down after 40 hours.

4.4 Comparison of H-CTMC and NH-CTMC Models for Information Dissemination on Twitter Data

Comparison of the H-CTMC and NH-CTMC Models for Information dissemination on Twitter data is done through the Likelihood Ratio Test (LRT). Referring to Hubbard et al. (2008), $LRT = -2(\ell_H(\theta) - \ell_{NH}(\theta)) \sim \chi^2_{(k)}$. If $LRT > \chi^2_{(k)} tabel$, then it is said that NH-CTMC provides a significant improvement in fit.

Based on the NH-CTMC transition intensity matrix in the form of the Sigmoid Function family in equation (17), numerically with the help of R software, $-2\ell_{NH} = 674{,}57$ is generated. Furthermore, the same is done for the H-CTMC Model, estimating the transition intensity using MLE with data from 20 Twitter users (145 IDs), resulting in

$$\mathbf{Q} = \begin{pmatrix} -0{,}09586 & 0{,}09586 \\ 0 & 0 \end{pmatrix}. \tag{18}$$

Based on the H-CTMC transition intensity matrix in Eq. (18), it is obtained that $-2\ell_H = 683{,}051$, so it is obtained that

$$LRT = -2(\ell_H(\theta) - \ell_{NH}(\theta)) = 683{,}051 - 674{,}57 = 8{,}481.$$

If 5% significance level and 3 independent degrees are used, the value of $\chi^2_{(3)}$ table = 7,815, and this result shows that $LRT > \chi^2_{(3)}$, meaning that the NH-CTMC model with intensity function

$$q_{12}(t) = \frac{e^{-2{,}09203}}{e^{-1245{,}927} + e^{0{,}020737\,t}}$$

provides a significant improvement in fit or it is said that the NH-CTMC Model is better than the H-CTMC Model in the problem of information dissemination on Twitter data.

5. Conclusion

In this paper, Information dissemination with NH-CTMC Model on Twitter data has been solved by using the assumption of transition intensity in the form of Sigmoid Function family. Sigmoid function parameters are estimated using Maximum Likelihood Estimation (MLE). The Sigmoid pattern can capture the pattern of information dissemination on Twitter. Information spreads quickly from followee to follower on the first day, then slows down after the second day and so on. Comparison of the NH-CTMC Model with the H-CTMC Model for information dissemination problems on Twitter data results in that the NH-CTMC Model is better than the H-CTMC Model based on the Log-Likelihood Ratio Test.

Acknowledgement

This research is funded by the Universitas Padjadjaran Doctoral Dissertation Research Grant (RDDU), with contract number: 1595/UN6.3.1/PT.00/2021 We are grateful for the discussion on social media analytics study with the collaborators in the RISE_SMA H2020 Project.

References

Agarwal, S., & Mehta, S. (2020). Effective influence estimation in twitter using temporal, profile, structural and interaction characteristics. *Information Processing and Management*, *57*(6). https://doi.org/10.1016/j.ipm.2020.102321

Begun, A., Icks, A., Waldeyer, R., Landwehr, S., Koch, M., & Giani, G. (2013). Identification of a multistate continuous-time nonhomogeneous markov chain model for patients with decreased renal function. *Medical Decision Making*, *33*(2), 298–306. https://doi.org/10.1177/0272989X12466731

Bremaud, P. (2020) *Probability Theory and Stochastic Processes*, Switzerland, Springer.

Firdaniza, F., Ruchjana, B. N., Chaerani, D., & Radianti, J. (2022a). Information Diffusion Model in Twitter: A Systematic Literature Review. *Information*, *13*(1). https://doi.org/10.3390/info13010013

Firdaniza, Ruchjana, B. N., Chaerani, D., & Radianti, J. (2022b). Information diffusion model with homogeneous continuous time Markov chain on Indonesian Twitter users. *International Journal of Data and Network Science*, *6*(3), 659–668. https://doi.org/10.5267/j.ijdns.2022.4.006

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. https://doi.org/10.1016/j.osnem.2018.04.001

Hsieh, H. J., Chen, T. H. H., & Chang, S. H. (2002). Assessing chronic disease progression using non-homogeneous exponential regression Markov models: An illustration using a selective breast cancer screening in Taiwan. *Statistics in Medicine*, *21*(22), 3369–3382. https://doi.org/10.1002/sim.1277

Hubbard, R. A., Inoue, L. Y. T., & Fann, J. R. (2008). Modeling nonhomogeneous Markov processes via time transformation. *Biometrics*, *64*(3), 843-850.

Kalbfleisch, J. D., & Lawless, J. F. (1985). The analysis of panel data under a Markov assumption. *Journal of the American*

Statistical Association, *80*(392), 863–871. https://doi.org/10.1080/01621459.1985.10478195

Kumar, Sanjay, 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. https://doi.org/10.1016/j.procs.2020.04.073

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, pp. 896–909). https://doi.org/10.1108/IMDS-07-2016-0290

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*. https://www.sciencedirect.com/science/article/pii/S0957417414001146

Ocaña-Rilola, R. (2002). Two methods to estimate homogenous Markov processes. *Journal of Modern Applied Statistical Methods*, *1*(1), 131–138. https://doi.org/10.22237/jmasm/1020255480

Ocaña-Riola, R. (2005). Non-homogeneous Markov processes for biomedical data analysis. *Biometrical Journal*, *47*(3), 369–376. https://doi.org/10.1002/bimj.200310114

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. https://www.igi-global.com/article/content/75772

Tapson, J., Jin, C., van Schaik, A. and Etienne-Cummings, R. (2009). A first-order nonhomogeneous markov model for the response of spiking neurons stimulated by small phase-continuous signals. *Neural Computation*, *21*(6), pp. 1554–1588. doi:10.1162/neco.2009.06-07-548.

Titman, A. C. (2011). Flexible nonhomogeneous Markov models for panel observed data. *Biometrics*. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1541-0420.2010.01550.x

Trajstman, A. C. (2002). A Markov Chain Model for Newcastle Disease and its Relevance to the Intracerebral Pathogenicity Index. *Biometrical Journal*, *44*(1), 43. https://doi.org/10.1002/1521-4036(200201)44:1<43::aid-bimj43>3.3.co;2-v

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. https://doi.org/10.1016/j.knosys.2017.07.003

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). https://doi.org/10.1103/PhysRevE.97.062306

Zhu, T., Wang, B., Wu, B., & Zhu, C. (2014). Maximizing the spread of influence ranking in social networks. *Information Sciences*, *278*, 535–544. https://doi.org/10.1016/j.ins.2014.03.070