Research Article

The Spread of Information in Virtual Communities

Zhen Zhang ¹, Jin Du,¹ Qingchun Meng,² Xiaoxia Rong ³, and Xiaodan Fan ¹

¹Department of Statistics, The Chinese University of Hong Kong, Hong Kong SAR, China

²School of Management, Shandong University, Jinan, China

³School of Mathematics, Shandong University, Jinan, China

Correspondence should be addressed to Xiaoxia Rong; rongxiaoxia@sdu.edu.cn and Xiaodan Fan; xfan@cuhk.edu.hk

Received 7 October 2020; Revised 6 November 2020; Accepted 10 November 2020; Published 26 November 2020

Academic Editor: Liang Wang

Copyright © 2020 Zhen Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the growth of online commerce, companies have created virtual communities (VCs) where users can create posts and reply to posts about the company's products. VCs can be represented as networks, with users as nodes and relationships between users as edges. Information propagates through edges. In VC studies, it is important to know how the number of topics concerning the product grows over time and what network features make a user more influential than others in the information-spreading process. The existing literature has not provided a quantitative method with which to determine key points during the topic emergence process. Also, few researchers have considered the link between multilayer physical features and the nodes' spreading influence. In this paper, we present two new ideas to enrich network theory as applied to VCs: a novel application of an adjusted coefficient of determination to topic growth and an adjustment to the Jaccard coefficient to measure the connection between two users. A two-layer network model was first used to study the spread of topics through a VC. A random forest method was then applied to rank various factors that might determine an individual user's importance in topic spreading through a VC. Our research provides insightful ways for enterprises to mine information from VCs.

1. Introduction

Virtual communities (VCs) provide an interactive experience that, if positive, may instil customer loyalty [1]. They enable consumers to learn the functions of products and follow up conveniently with buying online, as well as provide a channel for receiving customer feedback, which plays an important role in product innovation [2, 3]. Mining information provided by consumers in VCs enables companies to adjust the next generation of their products to improve customer satisfaction [4].

Complex network theory has been a major tool in the study of the physical structure and dynamic processes of social, biological, and technological networks [5]. In analyses of information spreading, users are represented as nodes [6]; these nodes may reside in multiple possible states, depending on whether they have learned information and whether they can transmit it to a neighbour [7–9]. Among real-world VCs, social networks such as Weibo, WeChat,

Twitter, and Facebook have different physical structures, leading to different patterns of topic transmission.

Consumer VCs are different from traditional social networks in that they centre around products. They provide a real-time look into customers' experiences with a product from the date of release. Users in VCs may post their feedback for other users to view, which, in the best-case scenario, may encourage loyalty among existing consumers while encouraging new consumers to buy in. Thus, understanding consumer VCs can provide insight into public opinion trends, helping the company maintain existing markets or develop new markets [10].

In consumer VCs, information is transmitted via posts, which are about particular topics [11]. Networks change over time; thus, time dimensions, i.e., temporal networks, have been incorporated into network analyses [12]. For consumer VCs, the identification of key time points during topic emergence can provide an angle for addressing the consulting service appropriately. Additionally, there are

multiple ways for users in VCs to interact with one another, of which “replying” and “searching for interest” are common ways to get information. Mining the influential users and specifying the important network features are also critical for improving the efficiency of multilayer network information transformation [13].

Thus, this paper tries to answer the following questions:

- (i) Following the introduction of a new product, how does the number of topics concerning the product grow over time?
- (ii) How do topics spread through a VC? What features characterise the influence of individual users during the information-spreading process?

Here, we focused specifically on the Huawei P10/P10 Plus for our case study. Two new concepts are introduced. First, the coefficient of determination was applied to the growth of topics after a new product was introduced to yield the node sequential emergence coefficient of determination (NSECD), which was used to identify the moment when most of the growth had finished. Second, the Jaccard coefficient was adjusted to gain a new measure of similarity between two users in a VC. Subsequently, a two-layered network model, representing two ways of spreading information, was introduced to study the spread of a topic through a VC and identify the most influential users. The random forest method was then applied to rank the importance of various factors affecting a user’s impact on spreading topics through a VC. Finally, some suggestions are provided for future research.

The remainder of this paper is organised as follows. Section 2 provides a literature review on existing methods. Section 3 describes the Huawei P10/P10 Plus dataset and its preprocessing. Section 4 introduces the NSECD statistic for studying the growth of new topics after the introduction of a new product. Section 5 proposes the adjusted Jaccard coefficient and the two-layered network model. Later in the section, simulations carried out to identify key users in the network are described, with the random forest method used to find features important to users’ information-spreading performance. Section 6 recaps suggestions for enterprise management of VCs and proposes directions for further research.

2. Literature Review

Online social networks exert a major influence on life today [14]. Sange et al. found that online social networks provide a platform for spreading both objective facts and fake news [15]. Park et al. noted the rapid growth of mobile devices such as smartphones and examined rapid information propagation in mobile social networks [16]. Up to now, the spreading of information has been investigated intensively in interdisciplinary fields [17].

Research into information propagation in VCs has two main foci: influence factor analysis and propagation path analysis. Influence factor analysis attempts to identify which factors make a node influential in a network; such factors may include gender, age, beliefs, etc. On the other hand, propagation path analysis studies the way in which

information is transmitted through the network by, for instance, assigning weights to edges and setting transmission probabilities based on these weights.

We first provide examples of influence factor analysis. Li et al. developed a multinomial naive Bayes classifier, categorised microblog posts based on content, and found that the information type has a significant influence on propagation patterns in terms of scale and topological features [18]. Zeng and Zhu proposed an emotional model for information propagation based on the emotional states of network users [19]. Hsu developed an integrated conceptual model and explored the effects of brand-evangelism-related behavioural decisions of enterprises on VC members [20]. Wan et al. used a least squares support vector machine to study consumer electronics supply chains [21]. However, all of the above studies were mainly analyses of the influencing factors but did not consider the differences among VCs.

With regard to propagation form, Huo and Cheng established a modified ignorant-wiseman-spreader-stifler model to analyse the spread of rumours through a network [22]. Xu et al. proposed a new iterative algorithm called SpectralRank, which assumes that a node’s propagation capability is proportional to the number of neighbouring nodes after adding a ground node to the network [23]. Shao et al. introduced the NL centrality algorithm to identify influential nodes in a network; the algorithm considers both the semilocal structure of a node and its topological position. [24]. Wang et al. proposed a method based on an integral k-shell to identify the influential nodes in a command-and-control network [25]. Escalante and Odehnal proposed a deterministic SIRS-type model for rumour propagation and applied it in simulations with two types of rumours: an original rumour, followed by a second counteracting rumour based on a complex network [26]. Li et al. presented the Potential Concentration Label method to help locate multiple sources of contagion under a susceptible-infected-recovered model [27]. Zhang et al. introduced a susceptible-infected-true-removed model of rumour spreading to account for members in a network who know or can discern the truth [28]. Xiong et al. introduced the location concept to a local social network model [29] and further extended it to the recommendation system via information spreading in a local-based social network [30, 31]. In their recent research, Xiong et al. combined the location and temporal effects of a social network, proposing constructive advice on dynamic management [32]. Zhang et al. studied networks that can be subdivided into smaller groups called communities and proposed a node-ranking algorithm called AI Rank using two factors: attractive power (which measures the number of followers a node has compared to its neighbours) and initiating power (which accounts for the communities that a node’s neighbours belong to) [33].

Although these studies considered node importance ranking together with the topology of the network, they generally treated networks as single-layered. This does not account for the fact that there may be more than one way for information to propagate through a network. Hence, a multilayer network model may be more appropriate. We propose a two-layered model in this paper.

3. Data Preprocessing and Description

The Pollen Club is the official VC for Huawei's products, including smartphones, laptops, and other electronic devices. Each user is assigned a unique identifier with which they can express their opinions about products freely, look through other users' posts, and reply to posts [34].

For our initial dataset, we selected 2000 web pages about the Huawei P10/P10 Plus, containing 2,392,035 posts about the product. After removing duplicate and spam posts, we retained 57,560 original posts and 826,328 reply posts by 129,362 users as our dataset. The data were acquired directly from club.huawei.com to avoid interview effects [35].

Next, core topics were extracted from the posts. In a previous study [36], 100 topics about phones were selected initially (see Table 1). After sorting out those with higher frequency, the remaining topics were grouped into three categories: system, software, and hardware, according to their features, as shown in Table 2.

The remainder of this paper focuses on 61 of these topics. These topics did not all appear on the first day of our dataset, but emerged over the course of the study as users bought and used the product and formulated questions about the product. In the next section, we will discuss the emergence of topics.

4. Dynamic Analysis of Topic Emergence

According to the classic product lifecycle, a new product goes through three stages: emergence, growth, and maturation. Emergence refers to the period before the product's launch [37]. Growth refers to a period of high consumer activity after the product launch, when consumers who had been eagerly awaiting the launch are ready to buy the product. Maturity refers to a period of low consumer activity afterwards, when consumers may continue to buy the product but do so at a slower pace because enthusiastic consumers have already done so.

For VCs, the transition point from growth to maturation is of interest to the company, because it signals the point after which fewer resources should be needed to monitor and respond to VC posts concerning the product. The purpose of this section is to introduce a statistic that will identify such a transition point.

For this purpose, we used the growth in the cumulative number of topics from the 61 topics up to a given date to measure the VC's interest in the product. We could, for instance, select the date at which the cumulative number of topics reached 90% or 95% of 61. However, any such choice would involve some uncertainty regarding which threshold to use. Instead, we propose a transition point that avoids such an arbitrary threshold choice.

Based on the adjusted coefficient of determination in statistics [38], we define the NSECD as follows:

$$r_t = 1 - \frac{T}{T-t+1} \left(1 - \frac{n_t}{n+1} \right) (1 \leq t \leq T), \quad (1)$$

where r_t is the NSECD value on the t^{th} day, n_t is the cumulative number of new topics on the t^{th} day, n is the total number of topics, and T is the total number of days.

As t increases, $T/(T-t+1)$ increases while $1 - (n_t/n+1)$ decreases. We expect that in a typical dataset, strong growth in new topics early in the study period will lead r_t to increase initially while faster growth in the factor $T/(T-t+1)$ together with a saturation of new topics will lead to r_t decreasing later on. Our key moment will be when this function reaches its maximum; i.e.,

$$t^* = \arg \max_t (r_t). \quad (2)$$

Figure 1 shows the NSECD on each day in our dataset. In this dataset, it was calculated that $t^* = 33$. Accordingly, $r_{t^*} = 0.8457$.

To summarise the lifecycle of topic growth in our case study, the period from the start of the study to the product launch date (Day 9) was considered the "emergence" stage. Twelve new topics were added during this stage, reflecting early interest in the Huawei P10/P10 Plus. The period from the product launch date to our key moment (Day 33) was the "growth" stage. The period from our key moment to the end of the study was considered the "maturity" stage.

In Figure 2, 12 new topics were added during the "emergence" stage, all on Day 9, the final day of the stage. This reflected early interest in the Huawei P10/P10 Plus.

The "growth" stage continued the trend of new topics from the last day of the emergence stage. New topics appeared rapidly early in this stage, as users shared their opinions towards the product from different aspects, but the rate at which new topics appeared slowed towards the end. By the end of this stage, 58 topics had emerged, 95.08% of the total number of topics considered.

The "maturity" stage witnessed even slower growth in new topics compared to the earlier stages, as most topics had already appeared earlier.

5. Network Modelling

In this section, we first introduce a two-layer network model and then perform information-propagation simulations to determine which users are the most effective at spreading information in a VC.

5.1. Structure of the Network. Here, we establish a two-layered network model for VCs. Each user is represented by a node that occurs on both layers. The two layers correspond to two ways in which a user in a VC may interact with another user: by replying to their posts or by searching for their posts.

The first layer is the "flow of information by replies (FIR)" network (denoted as $\text{FIR}(V_1, E_1, W_1)$). Given two users U_i and U_j , let w_{ij}^1 be the number of times U_j replied to a post by U_i within the dataset. If $w_{ij}^1 > 0$, then an arrow a_{ij}^1 from i to j is drawn. The FIR network consists of V_1 , the set of all users, W_1 , the set of all w_{ij}^1 weights, and E_1 , the set of all a_{ij}^1 arrows.

The second layer is called the "flow of information by interest (FII)" network (denoted as $\text{FII}(V_2, E_2, W_2)$). It is inspired by the idea that two users are likely to search for each other's posts, only when they share the same interests.

TABLE 1: Topics and frequency.

Serial number	Topics	Frequency	Serial number	Topics	Frequency	Serial number	Topics	Frequency	Serial number	Topics	Frequency
1	System	6569	26	Game	2810	51	Withdraw money	22	76	Motherboard	183
2	Life	1408	27	Light	692	52	Defrayment	1218	77	Touch	97
3	Taking pictures	3162	28	Flash back	695	53	King Glory	1471	78	Translation	42
4	Hibernation	204	29	Record	216	54	Beta	425	79	Clipboard	3
5	Power consumption	2905	30	Noise	123	55	Jingdong	601	80	Power saving mode	318
6	Data	1780	31	Font	499	56	Vmall	468	81	Wall paper	469
7	Wifi	1992	32	Loudspeaker	130	57	Theme	850	82	3D paper	15
8	Upgrade	4133	33	Program	1012	58	Antifingerprint oleophobic coating	622	83	Firmware bag	54
9	Charging	2459	34	Code	1007	59	Ring	472	84	Desktop	1066
10	Power off	934	35	Cloud service	191	60	Battery	2094	85	Radio	75
11	Unlock	1795	36	Lock screen	1968	61	Navigation key	236	86	NFC	622
12	WeChat	5569	37	Memory	1946	62	System B172	228	87	Developer options	129
13	Video	2574	38	Pattern	2684	63	System B213	231	88	Automatic rotation	56
14	Clarity	83	39	Backups	421	64	System B167	269	89	Vague	975
15	Telephone	2806	40	Resolution	398	65	Home key	495	90	Root	432
16	Net	1946	41	Touch off	2	66	Android8.0	95	91	Log version	6
17	Pixel	349	42	Pick-up hand	1719	67	Calendar	117	92	Kugou	54
18	Fingerprint	3241	43	Update	6422	68	Voice assistant	454	93	Compatibility	59
19	Payment	219	44	Camera	1778	69	Type	217	94	GPS	330
20	Block	172	45	Lightness	781	70	Infrared remote control	46	95	Weibo	709
21	Bank card	139	46	Player	137	71	Gesture operation	17	96	QQ	2253
22	Transportation card	345	47	Music	841	72	Application treasure	34	97	Position	803
23	Consumption	790	48	Screen capture	618	73	Split screen	216	98	Internet speed	443
24	Message	1075	49	Color	774	74	Heat	2700	99	4G	2277
25	Black screen	591	50	Location	600	75	Chip	279	100	Dominant frequency	33

TABLE 2: Classification of topics.

Type	Topics
System	System, upgrade, lock screen, unlock, black screen, font, resolution, update, lightness, color, screen capture, location, telephone, net, mode, vague, data, Wifi, power off, beta, theme, ring, voice assistant, heat, wall paper, desktop, NFC, root, 4G, internet speed, GPS, position
Software	WeChat, fingerprint, transposition card, consumption, message, game, flash back, program, code, backups, music, defrayment, video, King Glory, Jingdong, Vmall, Weibo, QQ
Hardware	Life, taking pictures, power consumption, charging, memory, pick-up hand, camera, light, antifingerprint oleophobic coating, battery, home key

FIGURE 1: Graph of the node sequential emergence coefficient of determination.

FIGURE 2: The topic lifecycle of Huawei P10/P10 Plus.

Given two users U_i and U_j , we can construct a measure w_{ij}^2 representing the commonality of interests and draw an edge a_{ij}^2 between them when w_{ij}^2 is above some preset threshold ε . The FII network consists of V_2 , the set of all users, W_2 , the set of all w_{ij}^2 weights, and E_2 , the set of all a_{ij}^1 arrows. It remains to define each w_{ij}^2 weight.

Let N be the total number of topics and let the topics have a fixed order. Then, let $\vec{F}^i = (F_1^i, \dots, F_k^i, \dots, F_N^i)^T$ be the vector comprising the numbers of the main U_i posts concerning each of the topics. Because some users may focus on only a portion of the topics, most entries in $\vec{F}^i$ are zero. As such, a Jaccard coefficient is taken into consideration. Let $R_i = \{k: F_k^i > 0\}$, i.e., the set of all topics for which U_i has at least one main post. By definition, the Jaccard coefficient of $\vec{F}^i$ and $\vec{F}^j$ can be calculated as follows:

$$\rho_{ij} = \frac{\#(R_i \cap R_j)}{\#(R_i \cup R_j)}, \quad (3)$$

where $\#(R_i \cap R_j)$ and $\#(R_i \cup R_j)$ denote the number of elements in the intersection and union of topics mentioned by users U_i and U_j , respectively.

The disadvantage of the Jaccard coefficient is that it does not distinguish a pair of casual users who post about only one topic, which happens to be the same, and a pair of enthusiastic users who post about many shared topics. For example, consider the following two situations:

Situation 1: users U_i and U_j post only on “system” and “updates.”

Situation 2: users U_i and U_j post only on “system.”

Based on equation (3), the Jaccard coefficient assigns a weight of 1 to both situations. However, this may not be appropriate, because the users in Situation 1 may be more active in the VC and, hence, more likely to exchange information by searching. Thus, we propose the following adjusted Jaccard coefficient:

$$\rho_{ij}^a = \frac{\#(R_i \cap R_j)}{\#(R_i \cup R_j)} \frac{\#(R_i \cup R_j)}{N} = \frac{(\#(R_i \cap R_j))^2}{\#(R_i \cup R_j) \times N}. \quad (4)$$

In our case study, $N = 61$. Based on equation (3), $\rho_{ij}^a = 0.0328$ in Situation 1 and $\rho_{ij}^a = 0.0164$ in Situation 2. Like the Jaccard coefficient, the adjusted Jaccard coefficient is always between 0 and 1, and is 0 when both users share no common topics. However, unlike the classical Jaccard coefficient, it is 1 only when both users share all topics.

In Figure 3, we plot the distribution of values of ρ_{ij}^a over all pairs of users in our case study. Figures 3(a) and 3(b) show that the overwhelming majority of pairs of users were associated with a small value of ρ_{ij}^a . A small ρ_{ij}^a means that there is little possibility for information spreading. In our

FIGURE 3: Distribution of ρ_{ij}^a : (a) distribution of ρ_{ij}^a ; (b) distribution of ρ_{ij}^a ($\rho_{ij}^a < 0.2$).

analysis, we used a threshold value of $\varepsilon = 0.2$. That is, an edge a_{ij}^2 was drawn between User U_i and User U_j only when $\rho_{ij}^a > 0.2$. This made the FII network a sparse network. Notably, the difference in the range longitudinal axis between Figures 3(a) and 3(b) is caused by the difference in the bar intervals.

Gephi software was used to draw illustrations of both layers for our case study. The results are shown in Figure 4.

To illustrate how our two-layer network model works, consider the following simple example with seven users, as shown in Figure 5.

Here, $a-g$ represent the seven users. Users may send or receive information from the FIR or FII networks simultaneously. For example, User a can receive information from User b or c in the FIR network, or from User b or d in the FII network. User a can send information to User b or d in the FII network. Information can spread through both channels simultaneously. The larger the weight of an edge, the more likely information is to flow through it at any given step in both layers.

5.2. Information Propagation and User Importance. Next, the two-layered network model will be expanded with information transmission mechanisms, to simulate the flow of information through the VC. Suppose we are interested in the propagation of a specific piece of information through the network. The information-propagation model will be based on a simple two-state framework; that is, at any given time, a node is in one of two possible states:

- (1) A susceptible state, representing the states of users who do not receive the information
- (2) An infected state, indicating the states of users who receive the information

Time is treated as discrete. At each time point t , infected nodes have a probability of transmitting the information to

their susceptible out-neighbours. The spread of the information is thus a random process. The mathematical model of this process is described below.

For convenience, susceptible and infected states are denoted by α and β , respectively. We denote the FIR layer and the FII layer as $l = 1$ and $l = 2$, respectively. The key notations are listed as follows:

- (i) $H_i^{(t)}$ is the state of User i at time t
- (ii) $r_i^{1,t}$ and $r_i^{2,t}$ are the indicator variables for User i being in states α or β , respectively, at time t
- (iii) $N^1(i)$ is the set of out-neighbours of User i in the FIR network
- (iv) $N^2(i)$ is the set of neighbours of User i in the FII network
- (v) $N(i) = N^1(i) \cup N^2(i)$ is the group of target users that User i may spread information to
- (vi) $I(\cdot)$ is the indicator function. It takes the value of 1 when the assertion inside is true and 0 when the assertion inside is false

The key mechanisms for this model are as follows:

- (1) For User i ($i = 1, 2, 3, \dots, n$), each day $t - 1$ except for the last step and each layer l $p_i^{l,t-1}$ is a draw from $U(0, 1)$. If User i is infected at $t - 1$, the value of this draw will decide which of the i 's out-neighbours on the layer l become infected at time t . More probable infections are always prioritised over less probable infections. A low value of $p_i^{l,t-1}$ means that the node will be easily infected, whereas a high value means that it will be difficult to infect.
- (2) For each User i , each layer l , and each in-neighbor j of i , p_{ji}^l will measure the infectiousness of the $j \rightarrow i$ channel. Large values indicate that j easily infects i , whereas small values indicate that j has difficulty

FIGURE 4: The two-layered network model applied to Huawei P10/P10 Plus: (a) FIR network; (b) FII network.

FIGURE 5: An example of FII and FIR networks.

infecting i . The values will depend only on the network structure from the previous section.

Note that out-neighbours and in-neighbours of node i are different only when the layer's edges are directed, i.e., the FIR network in this case. If the layer's edges are not directed, out-neighbours and in-neighbours are the same and are simply called neighbours.

At the starting time $t = 1$, a single node i_0 is infected while all others are susceptible; i.e., $H_{i_0}^1 = \beta$ while $H_i^1 = \alpha$ for

$i \neq i_0$. Inductively, assume we know H_i^{t-1} for all i . Then, the node states of the next day are given as

$$H_i^t = \begin{cases} \alpha, & \text{if } r_i^{1,t} = 1, \\ \beta, & \text{if } r_i^{2,t} = 1, \end{cases} \quad (5)$$

where $r_i^{1,t}$ and $r_i^{2,t}$ can be determined by

$$\begin{cases} r_i^{1,t} = 1 - r_i^{2,t}, \\ r_i^{2,t} = I(H_i^{t-1} = \alpha) I \left(\left[\sum_{l=1}^2 \sum_{\substack{j: i \in N^l(j), \\ H_j^{t-1} = \alpha}} I(p_{ji}^l > p_j^{l,t-1}) \right] > 0 \right) + I(H_i^{t-1} = \beta), \end{cases} \quad (6)$$

and where values of p_{ji}^l are calculated as follows:

$$\begin{cases} p_{ji}^1 = \frac{w_{ji}^1}{\sum_{k \in N^1(j)} w_{jk}^1}, \\ p_{ji}^2 = \rho_{ji}^a. \end{cases} \quad (7)$$

This means that infected nodes stay infected, whereas noninfected nodes become infected if, on some layers, one of its infected in-neighbours has a strong enough connection to overcome the given node's threshold at that time and layer. The information-spreading process begins with a single node i_0 infected, while all other nodes are susceptible. Any node can be used as the starting node. The process ends when either no infected node has any noninfected out-neighbours, or t has reached some specified time limit. In our simulations, we used a time limit of $t = 10$. Note that because the spreading process depends on random draws ($p_{ji}^{l,t-1}$), the process itself will be random.

A standard way to measure node i_0 's importance in a network is by the extent of the infection that begins at i_0 [39]. To this end, we define the information-spreading rate of a node i_0 as $C(i_0) = n_{i_0}/n$, where n_{i_0} is the number of infected nodes at the end of the spreading process and n is the total number of nodes. Because the spread of information depends on random variables ($p_{ji}^{l,t-1}$), $C(i)$ will itself be a random variable. Its expectation is too complicated to calculate explicitly; however, it can be estimated by repeated sampling as follows:

$$E(C(i)) \approx \frac{1}{N_0} \sum_{k=1}^{N_0} C_k(i), \quad (8)$$

where N_0 is the number of trials used, i.e., the number of simulations with i as the starting node, and $C_k(i)$ is the value obtained in the k^{th} trial. Due to the large size of our dataset and the resulting high computation time, N_0 was set as 30. Users were then ranked according to their mean information-spreading rate across these 30 trials.

The simulation procedure implemented in MATLAB (MathWorks, Natick, MA, USA) is as follows. Let i_0 be the node where the information starts. We track the spread of information throughout the network through the set of infected nodes, denoted as I , and the set of noninfected nodes, denoted as S :

Step 1: begin with the network nodes $V = \{1, 2, \dots, n\}$ and edges $E = \{a_{ij}^1, a_{ij}^2\}$ with weights $W = \{w_{ij}^1, w_{ij}^2\}$ and starting node i_0 .

Step 2: if i_0 has no out-neighbours, halt and output $t = 1$ and $C(i_0) = 1/n$. Otherwise, initialise $t = 1$, $I = \{i_0\}$, and $S = \{i \leq n: i \neq i_0\}$.

Step 3: while $t \leq 10$,

Step 3a: let $\partial I = \{i \in I: N(i) \cap S \neq \emptyset\}$. If this is empty, exit the while loop, otherwise continue.

Step 3b: for i in ∂I , (1) draw $p_i^{1,t}$ and $p_i^{2,t}$ from $U(0, 1)$, and (2) for each $j \in N(i) \cap S$, calculate p_{ij}^l as above. If $p_{ij}^l > p_i^{l,t}$, insert j into I and remove j from S .

Step 3c: update $C(i_0) = \#(I)/n$ and $t = t + 1$.

Step 4: output $C(i_0)$.

An example of a two-layered network and an information-spreading history with node a as the starting node is provided in Figure 6.

Notice that node a cannot spread information to node b in the FIR layer but may do so via the FII layer. From there, the information can spread from node b to c via the FIR layer. For a different spreading path, notice that a cannot spread to d in the FIR layer but may do so through the FII layer. From there, the information can spread to e through the FIR layer.

The algorithm was run on this simple example once for each of the nodes as the starting node. Table 3 shows the average spreading rate with examples in Figure 5 across 30 trials.

In their corresponding simulations Nodes a , b , c , and d had the highest mean spreading rates, with a possibility of spreading information to all nodes, except for the isolated node g . Of these four, node b performed best in terms of average spreading rate. Also, Nodes c and d were more efficient by 1% compared to node a , although node a had more links than those two. This suggests that c and d have special positions in this network, which matches the graph.

The algorithm was then applied to the real Pollen Club dataset described in Section 3. The results for the top 20 of 129,362 users, as ranked by mean spreading rate across 30 trials in MATLAB, are summarised in Table 4. The standard deviations were not too large for only 30 trials and could likely be shrunk further by running more trials as needed.

In Table 4, IUG and OUG stand for "intermediary user group" and "ordinary user group," respectively [36]. The OUG refers to customers who bought Huawei products and registered for the Pollen Club. The IUG refers to customers who received official training from Huawei and are willing to answer questions from other customers.

Specially, one of the IUG members is on the top of the list, indicating that the Pollen Club is organised by the customers themselves, saving the company and the effort of doing so. The number next to each OUG member is the user level. There are 12 levels in total, with higher levels indicating greater user experience. Users can advance their levels by joining activities in the Pollen Club. As can be seen, except for User 4948, the OUG members in our top 20 generally have high user levels. This shows that our method of ranking a user's spreading rate aligns well with Huawei's own method of ranking users.

To prove our model's effectiveness, comparisons were conducted using the probability 1 transformation model shown in Appendix. The results using our model were more consistent with trends observed in the real data.

5.3. Feature Selection of the Spreading Process. In this section, we investigate the relationships between spreading rate and network features that can be computed simply, without the

FIGURE 6: One possible information-spreading history in a 7-node network: (a) $t=1$, (b) $t=2$, (c) $t=3$, (d) $t=4$, (e) $t=5$, and (f) $t=6$.

TABLE 3: Mean and standard deviations of spreading rate of each node in Figure 5.

Source node	Mean spreading rate	Standard deviations of spreading rate
a	0.8048	0.1093
b	0.8333	0.0541
c	0.8143	0.0764
d	0.8095	0.0781
e	0.2762	0.0362
f	0.2810	0.0261
g	0.1429	0.0000

need for repeated simulations. Twenty-two network features [40] (denoted as x_1 through x_{22} , respectively) of the two-layer model for a node i are considered. Ten are for the FII

layer, and 12 are associated with the FIR layer. The 22 features are listed in Table 5.

Each feature x_1 through x_{22} is normalised to have a mean of 0 and a variance of 1. We then ask which of these features can predict the spreading rate. This provides insight into which features may cause a node to spread information more efficiently through the networks. We ran a random forest algorithm [41] by Breiman with the code referenced therein using scikit-learn [42]. Random forest is an ensemble method used to model regression in a nonlinear way.

Recall that the random forest algorithm consists of randomly selecting a subset of the users and a subset of the network features, forming a tree by selecting at each node the network feature and a boundary value for that feature. This splits the users into two branches at each such node, and continuing until in all branches, the number of users is

TABLE 4: Pollen Club Top 20 P10/P10 Plus users.

Rank	User ID	Group (level)	$C(i_0)$ (mean (std))
1	6050	IUG	0.0823 (0.0156)
2	65412	OUG (7)	0.0804 (0.0180)
3	78838	IUG	0.0800 (0.0157)
4	73370	IUG	0.0798 (0.0137)
5	62489	OUG (5)	0.0772 (0.0150)
6	9014	IUG	0.0762 (0.0206)
7	78835	OUG (6)	0.0761 (0.0205)
8	59364	OUG (10)	0.0757 (0.0217)
9	1008	OUG (5)	0.0753 (0.0167)
10	126016	OUG (6)	0.0743 (0.0191)
11	50128	OUG (9)	0.0742 (0.0199)
12	9381	OUG (7)	0.0729 (0.0204)
13	71961	OUG (6)	0.0725 (0.0201)
14	4948	OUG (2)	0.0716 (0.0150)
15	93168	OUG (7)	0.0710 (0.0210)
16	89378	OUG (7)	0.0708 (0.0216)
17	91874	OUG (7)	0.0704 (0.0154)
18	79875	OUG (6)	0.0700 (0.0188)
19	85851	OUG (7)	0.0697 (0.0299)
20	84435	OUG (5)	0.0695 (0.0141)

TABLE 5: Meaning of features.

Network layer	Feature name	Symbol	Meaning
FII	Degree	x_1	The number of neighbours of node i
	Weighted degree	x_2	The sum of the weights of all edges containing node i
	Eccentricity	x_3	The maximum distance from node i to any other node
	Closeness centrality	x_4	The reciprocal of the sum of the shortest distances from node i to all other nodes
	Harmonic closeness centrality	x_5	The number of pairs of distinct nodes j, k both different from node i for which the shortest path between j and k passes through node i
	Betweenness centrality	x_6	The number of pairs of distinct nodes j, k both different from node i for which the shortest path between j and k passes through node i
	Hub score	x_7	Let A be the adjacency matrix of the network. The hub score of node i is the i th component of the eigenvector corresponding to the maximum eigenvalue of $A\tau A$
	Authority score	x_8	It equals to the weighted sum of its neighbours' hub score. That is, $x_8(i) = \sum w_{ij}x_7(j)j \in N(i)$
	Local clustering coefficient	x_9	It is given by the proportion of links between the vertices within node i 's neighbourhood divided by the number of links that could possibly exist between them
	Eigen centrality1	x_{10}	The same as hub score except for we use the matrix adjacency matrix a instead of $A\tau A$ here
FRI	Indegree	x_{11}	The number of number of edges ending at node i
	Outdegree	x_{12}	The number of number of edges starting at node i
	Weighted indegree	x_{13}	The sum of the weight of all edges ending at node i
	Weighted outdegree	x_{14}	The number of number of edges ending at node i
	Eccentricity	x_{15}	Same as x_3
	Closeness centrality	x_{16}	Same as x_4
	Harmonic closeness centrality	x_{17}	Same as x_5
	Betweenness centrality	x_{18}	Same as x_6
	Hub score	x_{19}	Same as x_7
	Authority score	x_{20}	Same as x_8
	Page rank coefficient	x_{21}	An algorithm introduced by Google to rank web pages. It does so by estimating the probability that a person randomly clicking on links will arrive at the given page
	Local clustering coefficient	x_{22}	It is given by the proportion of links between the vertices within node's neighbourhood divided by the number of links that could possibly exist between them

TABLE 6: R^2 from random forest regression.

Round	Training dataset	Testing dataset
1	0.9703	0.7818
2	0.9706	0.7755
3	0.9704	0.7758
4	0.9697	0.7970
5	0.9699	0.7913

TABLE 7: Rank of the indexes for features in 10 rounds.

Rank	Round 1	Round 2	Round 3	Round 4	Round 5	Round 6	Round 7	Round 8	Round 9	Round 10
1	16	16	16	16	16	16	16	16	16	16
2	17	17	17	17	17	17	17	17	17	17
3	15	15	15	15	15	15	15	15	15	15
4	19	19	19	19	19	19	19	19	19	19
5	12	22	12	22	22	22	22	22	22	22
6	22	12	22	12	12	12	12	12	12	12
7	14	14	14	14	14	14	14	14	14	14
8	18	18	18	18	18	18	18	18	18	18
9	20	20	20	20	20	20	20	20	20	20
10	13	13	13	13	10	10	13	13	13	13
11	11	11	11	11	5	7	11	11	11	10
12	10	21	21	10	13	8	21	21	21	11
13	7	7	10	21	7	11	10	10	10	7
14	21	5	8	7	11	5	9	7	7	1
15	8	10	5	8	8	13	7	5	5	9
16	1	4	7	5	21	4	5	8	8	2
17	2	8	4	4	4	21	4	4	4	21
18	9	3	1	2	1	2	8	2	1	8
19	5	1	2	3	2	1	2	1	2	5
20	4	2	9	9	6	9	3	6	9	3
21	3	6	3	1	3	3	1	9	3	4
22	6	9	6	6	9	6	6	3	6	6
23	16	16	16	16	16	16	16	16	16	16
24	17	17	17	17	17	17	17	17	17	17
25	15	15	15	15	15	15	15	15	15	15

at most some threshold value. Given a new vector of values for the features, each tree is used to predict a value for the spreading rate. The forest as a whole then makes a prediction by averaging the values of each of the individual trees' predictions. The key parameters in this process are follows:

- (i) n_{tree} : the number of trees used. Increasing this value should decrease variance without leading to overfitting [43]. As suggested by reference [44], the number of trees was set to 500.
- (ii) m' : the number of features selected in each tree. We followed the suggestion in [44] to take $m' = \log_2 m \approx 4$, where m is the total number of features.
- (iii) n' : the number of users to sample in each tree. We used all the users, i.e., $n' = n = 129362$.
- (iv) $leaf'$: the maximum leaf size, which controls when the construction of each tree halts. We used $leaf' = 1$, i.e., we continued splitting until there was only one user in each branch.

Recall that after running the random forest algorithm on a training set, it outputs a regression function. Given a new point $x = (x_1, \dots, x_{22})$, the regression function outputs a predicted adjusted spreading rate $f(x)$, which can then be compared with the spreading rates from the simulations. We performed five-fold cross-validation on the Pollen Club dataset together with the spreading rates. Table 6 shows the resulting R^2 values when these regression functions were tested against the training set and the testing set.

The random forest algorithm can also be used to rank the importance of network features. Each tree ranks its m' selected features according to the decrease in variance at the corresponding node. That is, the variance in spreading rates at the parent node is compared to the sum of variances at the two child nodes; the greater the decrease, the more important that feature according to that tree. During the same process, the decreasing variance can be calculated for each tree. Finally, the decreasing variance for each of the m features is summed as the corresponding feature importance.

FIGURE 7: The importance ranking of features.

FIGURE 8: Probability 1 information-spreading history in a 7-node network: (a) $t=1$, (b) $t=2$, (c) $t=3$, and (d) $t=4$.

To check the stability, we again divided the dataset into 10 pieces, to check whether each piece yielded a similar ranking of the network features. The result is shown in

Table 7. We used Kendall's W-test [45] to assess the degree of agreement in the rankings across the ten runs and obtained a W value of 0.7570. According to the χ^2 test, this corresponds

TABLE 8: Top 20 P10/P10 Plus users.

Rank	User ID	Group (level)	$C(i_0)$
1	119794	OUG(2)	0.904757
2	119744	OUG(1)	0.904749
2	119781	OUG(1)	0.904749
2	119803	OUG(1)	0.904749
5	119741	OUG(2)	0.904741
5	119802	OUG(1)	0.904741
5	119823	OUG(1)	0.904741
8	119772	OUG(4)	0.904733
8	119826	OUG(1)	0.904733
10	119771	OUG(3)	0.904725
10	50128	OUG(2)	0.904725
12	9381	OUG(3)	0.904701
13	71961	OUG(5)	0.904693
13	4948	OUG(3)	0.904693
15	93168	OUG(1)	0.904686
16	89378	OUG(6)	0.904670
16	91874	OUG(3)	0.904670
16	79875	OUG(2)	0.904670
16	85851	OUG(1)	0.904670
20	84435	OUG(1)	0.904662

to a p value of 3.8064×10^{-11} , indicating very high confidence in a genuine agreement in the rankings.

Lastly, we obtained a single overall ranking based on the whole dataset. The results are shown in Figure 7. The top nine features all belong to the FIR layer. The most important features are closeness centrality (x_{16}) and harmonic closeness centrality (x_{17}) in the FIR layer. In the FII layer, the most important feature is eigencentrality; it indicates that the user in the central position in the FIR network can affect the information spreading. Also, the quality of the neighbours in the FII network plays an important role in transformation.

6. Conclusions

In this paper, we proposed the NSECD to identify the key moment in topic growth in a VC after a new product is introduced. A two-layer model was developed for assessing information propagation in a VC, where information can flow among users either by replies to posts (the FIR layer) or searching for topics of common interest (the FII layer). We applied this model to our case study, which focused on the P10/P10 Plus device in Huawei's Pollen Club, to identify which users were most effective at spreading information through the network. Lastly, we compared these results with commonly used network features using a random forest algorithm and found that spreading effectiveness correlated best with closeness centrality and harmonic closeness centrality in the FIR layer and eigencentrality in the FII layer.

We have two suggestions for how our model may be improved in future research. First, the infectiousness formula in the FIR layer can be modified to consider not just the quantity of postreplies but also their quality. For instance, natural language analysis [46] could be used to score the quality of posts. Second, the network model can be extended to have more than two layers. For example, many VCs enable

users to follow other users; thus, a third layer could be used to capture these follower relations.

In conclusion, this research introduces new concepts for network theory and provides suggestions for how companies could manage their VCs.

Appendix

Comparisons with a probability 1 transformation model: the probability 1 transformation model is defined as follows. The information-propagation model will also be based on a simple two-state framework; that is, at any given time, a node is in one of two possible states. In contrast to the model described in Section 5.2, if node i 's state is infected at time t , the states of neighbours (*i.e.* $N(i) = \cup N^l(i)$) in all layers of node i will turn into infected states.¹ This means that once infected, it will transform the information to all neighbours with probability 1 (Figure 8).

To illustrate this clearly, the same example used in Figure 5 in Section 5.2 is demonstrated. If node a is chosen as the initial source of the information, the information will spread to all other nodes in both layers, except for the isolated node, in four steps. Because the probability 1 model does not involve uncertainty, it can be implemented with one time step. To coordinate with our model as described in the main text, the maximum spreading time was set as 10. Where the meaning of the headline of the table is the same as that of Table 4. The results shown in Table 8 suggest that the lower-grade OUG members are more influential, which contradicts the meaning of the grade. This is due to the probability 1 model, which considers only the degree of the nodes and ignores the weight preference and uncertainty. Additionally, our model was more effective in predicting how information spreads among the nodes, compared with the probability 1 model.

Data Availability

All data, the models used during the study that appear in the submitted article, and the original data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study was supported by the National Natural Science Foundation of China (no. 71974115), the Innovation Method Fund of China (2018IM020200), and a theme-based research grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (project no. T12-710/16-R).

References

- [1] D. Ariely, *Predictably Irrational: The Hidden Forces that Shape Our Decisions*, HarperCollins, New York, NY, USA, 2008.
- [2] R. J. Brodie, A. Ilic, B. Juric, and L. Hollebeek, "Consumer engagement in a virtual brand community: an exploratory analysis," *Journal of Business Research*, vol. 66, no. 1, pp. 105–114, 2013.
- [3] S.-C. Chu and Y. Kim, "Determinants of consumer engagement in electronic word-of-mouth (eWOM) in social networking sites," *International Journal of Advertising*, vol. 30, no. 1, pp. 47–75, 2011.
- [4] M. E. Zaglia, "Brand communities embedded in social networks," *Journal of Business Research*, vol. 66, no. 2, pp. 216–223, 2013.
- [5] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca, "Network analysis in the social sciences," *Science*, vol. 323, no. 5916, pp. 892–895, 2009.
- [6] S. G. B. Roberts, R. I. M. Dunbar, T. V. Pollet, and T. Kuppens, "Exploring variation in active network size: constraints and ego characteristics," *Social Networks*, vol. 31, no. 2, pp. 138–146, 2009.
- [7] S. Boccaletti, V. Latora, and Y. Moreno, "Complex networks: structure and dynamics," *Physics Reports*, vol. 424, no. 4–5, pp. 175–308, 2006.
- [8] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167–256, 2003.
- [9] M. E. J. Newman, *Networks: An Introduction*, Oxford University Press, Oxford, UK, 2010.
- [10] X.-W. Wang, Y.-M. Cao, and C. Park, "The relationships among community experience, community commitment, brand attitude, and purchase intention in social media," *International Journal of Information Management*, vol. 49, pp. 475–488, 2019.
- [11] D. Baum, M. Spann, J. Fuller, and C. Thürridl, "The impact of social media campaigns on the success of new product introductions," *Journal of Retailing and Consumer Services*, vol. 50, pp. 289–297, 2019.
- [12] P. Holme and J. Saramäki, "Temporal networks," *Physics Reports*, vol. 519, no. 3, pp. 97–125, 2012.
- [13] M. Kivelä, A. Arenas, M. Barthélemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, "Multilayer networks," *Journal of Complex Networks*, vol. 2, no. 3, pp. 203–271, 2014.
- [14] C.-W. Tsai and S.-J. Liu, "SEIM: search economics for influence maximization in online social networks," *Future Generation Computer Systems*, vol. 93, pp. 1055–1064, 2019.
- [15] C. Sang, T. Li, S. Tian, Y. Xiao, and G. Xu, "SFTRD: a novel information propagation model in heterogeneous networks: modeling and restraining strategy," *Physica A: Statistical Mechanics and Its Applications*, vol. 524, pp. 475–490, 2019.
- [16] J. Park, J. Ryu, and S.-B. Yang, "ActiveDBC: learning knowledge-based information propagation in mobile social networks," *Wireless Networks*, vol. 25, no. 4, pp. 1519–1531, 2019.
- [17] P. Lu, L. Deng, and H. Liao, "Conditional effects of individual judgment heterogeneity in information dissemination," *Physica A: Statistical Mechanics and Its Applications*, vol. 523, pp. 335–344, 2019.
- [18] L. Li, Q. Zhang, J. Tian, and H. Wang, "Characterizing information propagation patterns in emergencies: a case study with Yiliang earthquake," *International Journal of Information Management*, vol. 38, no. 1, pp. 34–41, 2018.
- [19] R. Zeng and D. Zhu, "A model and simulation of the emotional contagion of netizens in the process of rumor refutation," *Scientific Reports*, vol. 9, no. 1, pp. 1–15, 2019.
- [20] L. C. Hsu, "Investigating the brand evangelism effect of community fans on social networking sites," *Online Information Review*, vol. 43, no. 5, 2019.
- [21] X. L. Wan, Z. Zhang, and X. X. Rong, "Exploring an interactive value-adding data-driven model of consumer electronics supply chain based on least squares support vector machine," *Scientific Programming*, vol. 2016, Article ID 3717650, , 2016.
- [22] L. A. Huo and Y. Y. Cheng, "Dynamical analysis of a IWSR rumor spreading model with considering the self-growth mechanism and indiscernible degree," *Physica A: Statistical Mechanics and Its Applications*, vol. 536, Article ID 120940, 2019.
- [23] S. Xu, P. Wang, and C. X. Zhang, "Spectral learning algorithm reveals propagation capability of complex networks," *IEEE Transactions on Cybernetics*, vol. 49, no. 12, pp. 4253–4261, 2018.
- [24] Z. Shao, S. Liu, and Y. Zhao, "Identifying influential nodes in complex networks based on Neighbours and edges," *Peer-to-Peer Networking and Applications*, vol. 12, no. 6, pp. 1528–1537, 2019.
- [25] Y. Wang, B. Chen, and W. Li, "Influential node identification in command and control networks based on integral k-shell," *Wireless Communications and Mobile Computing*, vol. 2019, Article ID 6528431, , 2019.
- [26] R. Escalante and M. Odehna, "A deterministic mathematical model for the spread of two rumors," *Afrika Matematika*, vol. 31, no. 2, pp. 315–331, 2020.
- [27] X. Li, Y. Liu, and C. Zhao, "Locating multiple sources of contagion in complex networks under the SIR model," *Applied Sciences*, vol. 9, no. 20, p. 4472, 2019.
- [28] J. P. Zhang, H. M. Guo, and W. J. Jing, "Dynamic analysis of rumor propagation model based on true information spreader," *Acta Physica Sinica*, vol. 68, no. 15, Article ID 150501, 2019.
- [29] X. Xiong, S. Qiao, and Y. Li, "Affective impression: sentiment-awareness POI suggestion via embedding in heterogeneous LBSNs," *IEEE Transactions on Affective Computing*, 2019.

- [30] X. Xiong, S. Qiao, and N. Han, "Where to go: an effective point-of-interest recommendation framework for heterogeneous social networks," *Neurocomputing*, vol. 373, pp. 56–69, 2020.
- [31] X. Xiong, S. Qiao, Y. Li, N. Han, G. Yuan, and Y. Zhang, "A point-of-interest suggestion algorithm in multi-source geo-social networks," *Engineering Applications of Artificial Intelligence*, vol. 88, Article ID 103374, 2020.
- [32] X. Xiong, F. Xiong, and J. Zhao, "Dynamic discovery of favorite locations in spatio-temporal social networks," *Information Processing & Management*, vol. 57, no. 6, Article ID 102337, 2020.
- [33] W. Zhang, J. Yang, X.-Y. Ding, X.-M. Zou, H.-Y. Han, and Q.-C. Zhao, "Groups make nodes powerful: identifying influential nodes in social networks based on social conformity theory and community features," *Expert Systems with Applications*, vol. 125, pp. 249–258, 2019.
- [34] Huawei Pollen Club, "Huawei pollen club," 2020, <https://club.huawei.com/forum.html>.
- [35] P. V. Marsden, "Interviewer effects in measuring network size using a single name generator," *Social Networks*, vol. 25, no. 3, pp. 1–16, 2003.
- [36] Q. C. Meng, Z. Zhang, and X. L. Wan, "Properties exploring and information mining in consumer community network: a case of Huawei pollen club," *Complexity*, vol. 2018, Article ID 9470580, , 2018.
- [37] R. Vernon, "International investment and international trade in the product cycle," *International Economic Policies and Their Theoretical Foundations*, Academic Press, vol. 80, no. 2, pp. 415–435, 1992.
- [38] X. Q. He and W. Q. Liu, *Applied Regression Analysis*, China Renmin University Press, Beijing, China, 2nd edition, 2014.
- [39] X. L. Wan, Z. Zhang, and C. Zhang, "Stock market temporal complex networks construction, robustness analysis, and systematic risk identification: a case of CSI 300 index," *Complexity*, vol. 2020, Article ID 7195494, , 2020.
- [40] G. R. Chen, X. F. Wang, and L. I. Xiang, *Introduction to Complex Networks: Models Structures and Dynamics*, Higher Educational Press, Beijing China, 2012.
- [41] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [42] F. Pedregosa, G. Varoquaux, and A. Gramfort, "Scikit-learn: machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [43] G. Biau and E. Scornet, "A random forest guided tour," *Test*, vol. 25, no. 2, pp. 197–227, 2016.
- [44] L. Raynal, J.-M. Marin, P. Pudlo, M. Ribatet, C. P. Robert, and A. Estoup, "ABC random forests for Bayesian parameter inference," *Bioinformatics*, vol. 35, no. 10, pp. 1720–1728, 2019.
- [45] A. P. Field, *Kendall's Coefficient of Concordance*, Wiley StatsRef: Statistics Reference Online, Hoboken, NJ, USA, 2014.
- [46] X. Xiong, Y. Li, R. Zhang, Z. Bu, G. Li, and S. Ju, "DGI: recognition of textual entailment via dynamic gate matching," *Knowledge-Based Systems*, vol. 194, no. 22, Article ID 105544, 2020.