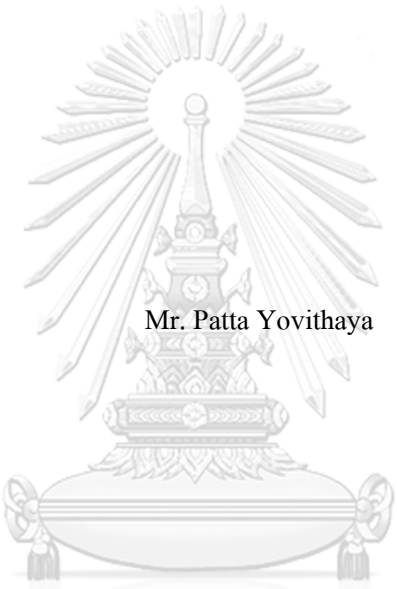
Using graph evolutionary to retrieve related tweets

Mr. Patta Yovithaya



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

A Thesis Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science in Computer Science

Department of Computer Engineering

FACULTY OF ENGINEERING

Chulalongkorn University

Academic Year 2022

Copyright of Chulalongkorn University

การใช้วิวัฒนาการของกราฟสำหรับสืบค้นข้อมูลทวิตเตอร์

นายภัททะ ยอวิทยา

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาศาสตร์คอมพิวเตอร์ ภาควิชาวิศวกรรมคอมพิวเตอร์
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2565
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Thesis Title	Using graph evolutionary to retrieve related tweets
By	Mr. Patta Yovithaya
Field of Study	Computer Science
Thesis Advisor	Assistant Professor SUKREE SINTHUPINYO, Ph.D.

Accepted by the FACULTY OF ENGINEERING, Chulalongkorn University in Partial Fulfillment of the Requirement for the Master of Science

..... Dean of the FACULTY OF ENGINEERING

(Professor SUPOT TEACHAVORASINSKUN)

THESIS COMMITTEE

..... Chairman

(Assistant Professor NATTEE NIPARNAN, Ph.D.)

..... Thesis Advisor

(Assistant Professor SUKREE SINTHUPINYO, Ph.D.)

..... Examiner

(Associate Professor NUTTAPONG CHENTANEZ, Ph.D.)

..... External Examiner

(Assistant Professor Dr. Denduang Pradubsuwan)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

ภัททะ ยอวิทยา : การใช้วิวัฒนาการของกราฟสำหรับสืบค้นข้อมูลทวิตเตอร์. (Using graph evolutionary to retrieve related tweets) อ.ที่ปรึกษาหลัก : ผศ. ดร.สุกรี สินธุภิญโญ

เนื่องด้วยความนิยมและจำนวนผู้ใช้งานแต่ละวันทำให้โซเชียลมีเดียจึงมีอิทธิพลอย่างมากในช่วงทศวรรษที่ผ่านมา ด้วยลักษณะการใช้ข้อความโต้ตอบและบันทึกข้อความสั้น ๆ ผู้ใช้งานส่วนใหญ่จึงใช้ทวิตเตอร์เพื่ออัปเดตข่าวสารหรือเหตุการณ์ต่าง ๆ ที่เกิดขึ้นล่าสุด จากข้อความจำนวนมากที่ถูกส่งบนทวิตเตอร์ทำให้วิวัฒนาการของเหตุการณ์ใดเหตุการณ์หนึ่งสามารถพัฒนาไปสู่เหตุการณ์ที่เกี่ยวข้องโดยที่เป็นหัวข้อเดียวกัน โดยงานวิจัยนี้ได้นำเสนอวิธีการใหม่ในการสืบค้นข้อมูลทวิตเตอร์ที่เกี่ยวข้องกับคำที่ใช้ค้นหา นอกจากข้อมูลทวิตเตอร์ที่คล้ายคลึงกันจะถูกสืบค้นแล้ว ข้อมูลทวิตเตอร์อื่น ๆ ที่เกี่ยวข้องจะถูกสืบค้นขึ้นมาด้วย งานวิจัยนี้ใช้วิธีการ Simplification Summarization เพื่อลดทอนข้อมูลที่มีความสำคัญน้อยและคงไว้เพียงข้อมูลที่มีความสำคัญมากในกราฟเครือข่าย โดยใช้มาตรวัดค่ากลางและค่าสัมประสิทธิ์จากทฤษฎีการวิเคราะห์กราฟเครือข่าย ในขณะที่การใช้วิวัฒนาการของกราฟสามารถช่วยแสดงการแพร่กระจายของความสัมพันธ์ของข้อมูลที่เกี่ยวข้อง การศึกษานี้นำข้อมูลจากทวิตเตอร์ของสำนักข่าวไทย เพื่อใช้วัดผลการสืบค้นข้อมูลที่เกี่ยวข้องและพบว่าการใช้วิวัฒนาการของกราฟสามารถสืบค้นข้อมูลที่เกี่ยวข้องได้มากขึ้น เมื่อเทียบกับการใช้คะแนนความคล้ายคลึงเพื่อสืบค้นข้อมูล

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สาขาวิชา	วิทยาศาสตร์คอมพิวเตอร์	ลายมือชื่อนิสิต .....
ปีการศึกษา	2565	ลายมือชื่อ อ.ที่ปรึกษาหลัก .....

6370220721 : MAJOR COMPUTER SCIENCE

KEYWORD: graph evolutionary, graph summarization, information retrieval

Patta Yovithaya : Using graph evolutionary to retrieve related tweets . Advisor: Asst. Prof. SUKREE SINTHUPINYO, Ph.D.

Due to its popularity and daily active users, social media has become powerful and influential in the last decade. With the nature of a micro-blogging platform, instant messages and the latest short posts are sent throughout the network on Twitter. Therefore, most users utilize Twitter to update breaking news or the latest events. Since a huge volume of tweet messages have been published on Twitter, event evolution has also rapidly developed into related events within similar topics. In this study, we present a novel method to retrieve tweets that relate to a given query term. Not only perfectly matched tweets, but more related tweets will be retrieved. The collected tweet data are processed and constructed as an original network. With the benefits of social network analysis, a simplification-based summarization approach is applied to ignore information that has less importance while preserving significant information in the network based on centrality measurement and clustering coefficient. Using the evolutionary of graph-based representation extends the relationship diffusion to assist related information retrieval. Experiments were performed using Thai news datasets and the framework performance was evaluated by precision, recall, and f-score. The experimental results show that our framework outperformed the baseline methods which derived a similarity score based on the word-embedded vector to find relevant documents.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Field of Study:	Computer Science	Student's Signature .....
Academic Year:	2022	Advisor's Signature .....

ACKNOWLEDGEMENTS

I would like to thank my supervisor Asst. Prof. Sukree Sinthupinyo for his consistent support, guidance, and overall insights throughout the research. Furthermore, I would like to thank all thesis commitees for providing comments and recommendations. I am also thankful to the faculty of Computer Engineering and all staff members for their considerate support. Finally, I would like to thank my family and friends for all unconditional support.

Patta Yovithaya

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

TABLE OF CONTENTS

Page

..... iii

ABSTRACT (THAI)..... iii

..... iv

ABSTRACT (ENGLISH)..... iv

ACKNOWLEDGEMENTS.....v

TABLE OF CONTENTS..... vi

LIST OF FIGURES 8

LIST OF TABLES.....9

1. Introduction.....10

1.1 Objectives.....11

1.2 Scope.....11

2. Related works.....12

2.1 Twitter API.....12

2.2 Text Summarization 12

2.3 Graph-based Summarization.....13

2.4 Tweet Summarization 14

2.5 Event Evolution.....15

2.6 Social Network Analysis.....15

3. Proposed methodology.....17

3.1 Dataset.....21

3.2 Construct a simplified network 22

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

3.3 Building graph evolutionary properties.....25

4. Evaluation 33

5. Experimental results.....35

5.1 Performance experiment 35

5.2 Our improvement over the ordinary approaches.....37

6. Conclusion.....41

REFERENCES 43

VITA 48

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

LIST OF FIGURES

Page

Figure 1: The system architecture 17

Figure 2: Network graph 22

Figure 3: Top 30 nodes of clustering coefficient 22

Figure 4: Top 30 nodes of degree centrality 23

Figure 5: Top 30 nodes of betweenness centrality 23

Figure 6: Top 30 nodes of closeness centrality 23

Figure 7: A network graph with selected nodes 24

Figure 8: A simplified network graph 24

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

LIST OF TABLES

Page

Table 1: Total number of collected tweet messages 21

Table 2: Added and overlapped nodes/edges.....32

Table 3: Experimental configurations.....35

Table 4: Overall performance comparison.....37

Table 5: Predicted positive class performance.....40

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

1. Introduction

With the number of active users over 400 million, Twitter becomes one of the most popular social media platforms in the world [1]. Users are allowed to share their contents on the social network in diverse forms such as text, image, and video. Since a huge number of users are actively generating several tweets every day, tweets are reproduced in many different topics. Twitter is a microblogging platform that can create either short blogs or instant messages by posting the current status or latest news. More than a half of Twitter's users in United States typically get news on Twitter which is the biggest portion among the social network platforms [2]. Due to the popularity in news, several popular news channels are listed on the chart of the most Twitter followers in 2021 [3]. According to a findings [4], 70% new users exploit Twitter to keep live events up in real time. With a nature of using Twitter, users could easily explore through hot trending topics and instantly receive quick updates what is happening around the world. A tweet is not only containing a creator message, but it also includes thought and sentiment from different users related to its tweet. Therefore, there are a diverse of vocabulary that have been using on Twitter every day. A typical representation that best fits on both words and relations is a network graph which basically consists of vertices and edges.

A graph or network is a representation composed of a collection of both vertices and edges. An edge is a linkage between two vertices. A graph typically visualizes the connection that interacts among vertices. With the time involved, vertices/edges can be added, overlapped, and deleted which is called a dynamic graph [5]. Since there are many graph-based summarization approaches that have been proposed, we utilize a simplification-based summarization technique that helps filtering important nodes or edges from an original network to better understand the

large social network. Furthermore, we exploit centrality measures which are one of social network analysis theories to gain a deeper understanding of network. To visualize the evolutionary of vertices in the network, we illustrate vertices/edges that have added and overlapped in a consecutive day.

1.1 Objectives

This research aims to exploit the graph evolutionary for extracting news that correlate with the others. Furthermore, we create a framework of using graph-based evolutionary to conclude key important news daily topics that can retrieve more related tweets compared to word-embedded similarity approaches. The time-evolved analysis can assist the interpretation of how connections evolve from one to another and develop the expansion of related news when the time changes.

1.2 Scope

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

This research collects tweet messages from news agency accounts including MatichonOnline and MorningNewsTV3. The experiment covers ten news topics to extract related news and evaluate the performance using precision, recall, and f-score.

2. Related works

This chapter covers the detail of related works that will be used in this study. We describe the way to collect Twitter data, a literature review of summarization approaches, and event evolution. Plus, a theory of Social Network Analysis is also explained in the following section.

2.1 Twitter API

The application process interface (API) provided by Twitter, also known as Twitter API [6], can be used to programmatically create communication on Twitter. Twitter API provides various programmatic endpoints to retrieve Twitter data with respect to the level of access permission. In this study, we take an applicable endpoint of retrieving multiple tweets by a single user. This endpoint will response the most recent tweets from a specific user ID.

2.2 Text Summarization

Text summarization is a very popular technique for compacting the full-length original text into a concise summary while maintaining key important information. There are commonly two approaches in text summarization: extractive-based summarization and abstraction-based summarization [7]. In extractive-based summarization approach, the significant subsets of the original text will be pulled out and integrating them to make a summary. On the other hand, an abstraction-based summarization is a technique of reproducing key crucial information using advanced techniques into a new shorter way. There are a variety of methods proposed in graph-based text summarization. TextRank [8] is one of the most prominent algorithms where sentences represent the vertices of a graph and edges determine a similarity between sentences. Yongkiatpanich and Wichadakul [9] have been studied

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

using PageRank and a combination of ontology and word embedded. Joshi et al. [10] proposed a semantic graph-based text summarization approach for documents. They also used graph theory measures for choosing the summary sentences based on a top of their semantic scores. Natesh et al. [11] studied a text summarization in graph-based approach where the graph was built on co-occurrence base. That method shows a good performance on news articles, Wikipedia searches, and technical documents. Recently, several studies have been proposed in abstraction-based summarization [12-14] and extractive-based summarization approaches [15-18].

2.3 Graph-based Summarization

A graph-based summarization can be categorized in a variety of methods. In order to completely illustrate the approaches, Liu et al. [19] summed up the graph-based summarization methods in four basis techniques relying on different objectives. Aggregation based [20-24], bit compression based [25], simplification based [26, 27], and influence based [28] are a suite of core techniques that have been used in many studies. Aggregation based is one of the most prominent techniques that grouped nodes or edges into super nodes/edges using its optimization function which can be divided into node-grouping and edge-grouping methods. Node-grouping methods help grouping all nodes within its cluster into a single node called a supernode. Whereas edge-grouping methods are an approach to minimize number of edges into compressor or virtual nodes. One of the edge-grouping methods is the graph dedensification which enhances query processing by compressing dense high-degree vertices in a graph [24]. Recently [23], a distributed aggregation-based graph summarization for large-scale approximate attributed graph queries was proposed to semantically produce a graph summary for improving computational costs and maintaining query accuracy. A bit compression in graph-

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

based summarization focuses on reducing number of bits in the original graph, where reporting core structural pattern along with some original components in the modeled graph. In simplification-based method, the main idea is to pull out crucial nodes and edges to simplify the original graph where disconnecting less important nodes or edges from the graph. Zeqian et al. [26] proposed a simplification summarization technique called “OntoVis” which utilizes the ontology information to semantically prune the large networks. OntoVis also supports both structural abstraction and semantic abstraction. In structural abstraction, the essential structure of the entire network is required to be preserved while pruning the network. Semantic abstraction allows users to construct a derived graph from the original graph by including only nodes whose types are selected in the ontology graph. Li and Lin [27] proposed an unsupervised algorithm for egocentric information abstraction in heterogeneous social networks where obtains similarly a resulting graph as OntoVis. Various recent studies [29-31] introduced how to simplify a network that includes discovering important vertices/edges with different metrics and pruning non-essential components of the network. Influence-based approach is an understanding of high-level description of the influence propagation in large-scale networks. This approach typically utilizes an optimization process to formulate graph summarization. Mehmood et al. [28] researched on the social network summarization with the use of social influence analysis and information propagation.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

2.4 Tweet Summarization

Tweet summarization has been widely studied in several areas of interest with different techniques. Naik et al. [32] introduced Particle Swarm Optimization Algorithm which enhances the accuracy for clustering tweets for a summarization. With a benefit of using pre-trained language models shows an outstanding

performance on various fields of study, Dusart et al. [33] proposed a neural network based model for summarizing huge tweet streams and dynamic output summary based on the input size automatically. Chakraborty et al .[34] authored the summarization of tweets related to news article approach that can capture the diversity of tweet opinions.

2.5 Event Evolution

Several studies that researched the event evolution of social media platform. With a rapid change of events and opinions exchange, event evolution approaches emerged for detecting occurring events and discovering the evolution of events. Panagiotou et al. [35] summarized definitions, challenges, and trends in event detection frameworks that have been researched in the last few years. Cordeiro and Gama introduced a research survey of event detection techniques [36] that can generally be classified into two major methods. The first method is the Online New Event detection (NED) which points out to discovering events in real-time documents. The second technique, Retrospective Event Detection (RED), is a retrieval document process of event- relevant documents from the collection of historical documents. A discovery of event evolution from news stories was studied by Dou et al. [37]. The graph-based event evolution visualizes the core structure of events that can help the information extraction task and information retrieval.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.6 Social Network Analysis

Social network analysis has emerged in the past recent years and widely studied in diverse area of interest. To systematically perform analysis on social network, the social network analysis metrics assist in many different aspects:

discovering the pattern of connection, producing an underlying to differentiate networks, monitoring changes in a network over time, and analyzing locations in a network [38]. Centrality measurement is one of social network analysis metrics that can determine how significant a vertex locates within the network. Vertex-specific network metrics provide several quantitative measures: degree centrality, betweenness centrality, closeness centrality, and clustering coefficient. Degree centrality (DC) measures a number of edges connected to a node. The degree centrality of vertex i is given in (1) where A_{ij} is the adjacency matrix of the graph and n is a number of vertices [39].

$$D(i) = \frac{1}{n-1}\sum_{j=1}^{n} A_{ij} \qquad (1)$$

Betweenness centrality (BC) is a measure that the extent of a node lies on the shortest path to other nodes. An equation is given in (2) where V is a set of nodes and $\sigma(s,t)$ is the number of shortest (s,t)-paths. Also, $\sigma(s,t|v)$ is the number of shortest (s,t)-paths passing through some node v other than s,t. If $s = t$, let $\sigma(s,t) = 1$, and if $v \in \{s,t\}$, let $\sigma(s,t|v) = 0$ [40].

Chulalongkorn University

$$c_B(v) = \sum_{s,t \in V} \frac{\sigma(s,t|v)}{\sigma(s,t)} \qquad (2)$$

Closeness centrality (CC) is a fraction of the sum of shortest path lengths to all other nodes given in (3). The notion of closeness centrality points up a node that can possibly reach any other nodes in the graph either a very far off or a nearby with small hops [41].

$$C(u) = \frac{1}{\sum_y d(u,y)} \qquad (3)$$

Clustering coefficient (CF) measures the density of connections in the network. The equation is given in (4) where $T(u)$ is the number of triangles through node u and $\deg(u)$ is the degree of u [42].

$$C_u = \frac{2T(u)}{\deg(u)(\deg(u)-1)} \quad (4)$$

3. Proposed methodology

In this chapter, we are going to explain the proposed methodology of using graph evolutionary to retrieve related tweets. The overall system architecture, as illustrated in Figure 1, shows how our framework processes a given query to find associated terms using graph evolutionary and retrieve related tweets.

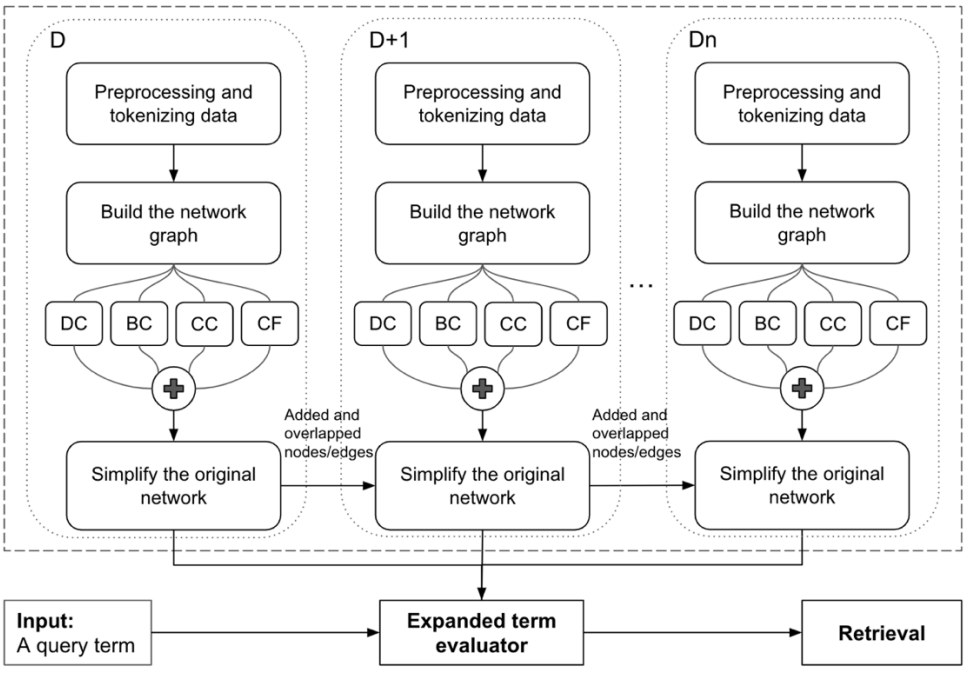


Figure 1: The system architecture

The data acquisition in this study is required to use the application process interface provided by Twitter. We firstly retrieve tweet contents by utilizing the use of Twitter's API and save the collection of data to proceed further steps. Since Twitter and any social media platforms allow users to generate posts in several forms, the pre-processing step is mandatory to deal with the retrieved data. As this study focuses on text only, the cleansing process takes place to help removing irrelevant information on the data e.g., special characters, emoji, hashtags. After cleansing process completed, we apply a tokenization on each tweet to split data into words. Each word represents an individual vertex in the network. However, the nature of tweets is relatively small. The relations become very crucial to capture the information association between words in the tweets. Hence, the edges are built the relationship on a pair of consecutive words in the same tweet message. With these steps, we can construct a network graph based on the daily collected tweets.

In a collection of daily tweets, the network graph is relatively dynamic relying on a trending topic that twitter users are currently interested in. We consider the approach that can extract the core components in the primitive network graph into a simplified graph. It can assists filtering the important information which literally contribute to the network using simplification-based summarization technique. Hence, social network theory is significantly important for measuring nodes that have an impact and popularity in the network. We take the advantages of the DC, BC, and CC measurements that can show an excellent popularity and reputation in the network [43]. Furthermore, clustering coefficient is an essential measurement that can measure how connected neighbor's vertices tend to be a cluster.

Since the original network is measured using centrality and clustering metrics, high values give different benefits relying on a particular indicator. Since these

measurements help to filter out nodes from any other nodes in the graph, we integrate all important nodes into a set of selected vertices by selecting top N nodes from DC, BC, CC, and CF. As a result of combining nodes, the simplified-graph process starts to distill the connected components out of the original graph based on selected vertices. First, we find neighbors of the vertex in the network and all possible edges among neighbors of the vertex that will be removed using the Cartesian product. After a set of new possible edges is created, each of them will be added to the network and both nodes and edges in between will also be removed. At the end, the simplified graph shows the core connected components for summarizing tweets on daily basis.

Considering a time-evolved analysis in a certain period, the graph evolutionary assists the interpretation of expanded connections when the time changes. Since a simplified graph is representative of daily tweet messages, we compose an evolution of graph-based representation into two approaches; added and overlapped nodes/edges. The nodes/edges, which join on the next day (d_{i+1}), but do not appear in d_i is called “added nodes/edges”. However, the “overlapped nodes/edges” are the nodes/edges that exist in both d_i and d_{i+1}.

Chulalongkorn University

The details to simplify a large network and create graph evolutionary properties are described in Algorithm I. Starting with data acquisition, data preprocessing, and tokenization. Building the network graph and centrality measurements are processed in step 4. A framework then applies a graph-based simplification process and creates graph evolutional properties respectively.

ALGORITHM I: Simplify network graph and construct graph evolutionary

1. Collect tweets data using Twitter API

2. Preprocess collected data (remove special characters, emoji, hashtags)

3. Tokenize each tweet into words

4. Construct a graph $G_i = (V, E)$ where a set of nodes (V) represent the tokenized words and a set of edges (E) represent adjacent word in the same tweet
   - 4.1 Measure a degree centrality, betweenness centrality, closeness centrality, and clustering coefficient
   - 4.2 Sort all measurements in step 4.1 in descending order

5. Select unique nodes based on conditions:
   - 5.1 Clustering coefficient top N nodes and > 0
   - 5.2 Degree centrality top N nodes
   - 5.3 Betweenness centrality top N nodes
   - 5.4 Closeness centrality top N nodes

6. Simplify the graph based on selected nodes in step 5

```
foreach d_{i=1} in D = {d_{1,2,..,n}}:
    foreach v in V_d:
        if v not in selected nodes:
          src ← neighbors of v
          dst ← neighbors of v
          new_edges ← cartesian_product(src, dst)
          foreach s, t in new_edges:
            if s ≠ t :
              G_d ← add_edges_from(s, t)
              G_d ← remove_node(v )
```

7. Create graph evolutionary properties

```
foreach d_{i=2} in D = {d_{1,2,..,n}}:
    G_d ← Simplified graph of d_i
    G_{d_{i-1}} ← Simplified graph of d_{i-1}
    added_nodes ← (G_d nodes) – (G_{d_{i-1}} nodes)
    overlapped_nodes ← (G_d nodes) ∩ (G_{d_{i-1}} nodes)
    added_edges ← (G_d edges) – (G_{d_{i-1}} edges)
    overlapped_edges ← (G_d edges) ∩ (G_{d_{i-1}} edges)
```

8. End of Algorithm

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

3.1 Dataset

In the experiment, we pulled out tweet messages from Thai news agencies through Twitter API including two accounts: MorningNewsTV3 and MatichonOnline. We selected these news agencies to collect tweets because of the popularity. Both are the most renowned news industries in Thailand which have over 3.1 and 1.1 million followers respectively. The number of collected data from both accounts in June-July 2022 is displayed in Table 1.

	MatichonOnline	MorningNewsTV3
1-7 June 2022	612	464
8-14 June 2022	704	551
15-21 June 2022	1,015	632
22-30 June 2022	1,489	692

Table 1: Total number of collected tweet messages

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

However, a combination of two famous agencies is used in the experiment. After the data has been collected, the preparation process was needed to be done by cleansing and transforming tweet messages to appropriate format. Data cleansing is a very important process which helps removing unused and irrelevant message. With the help of well-cleansing step, the tokenization operates splitting tweet messages in a form of words correctly. The tweets data of public accounts will be used in the preliminary experiment that is retrieved in June 2022.

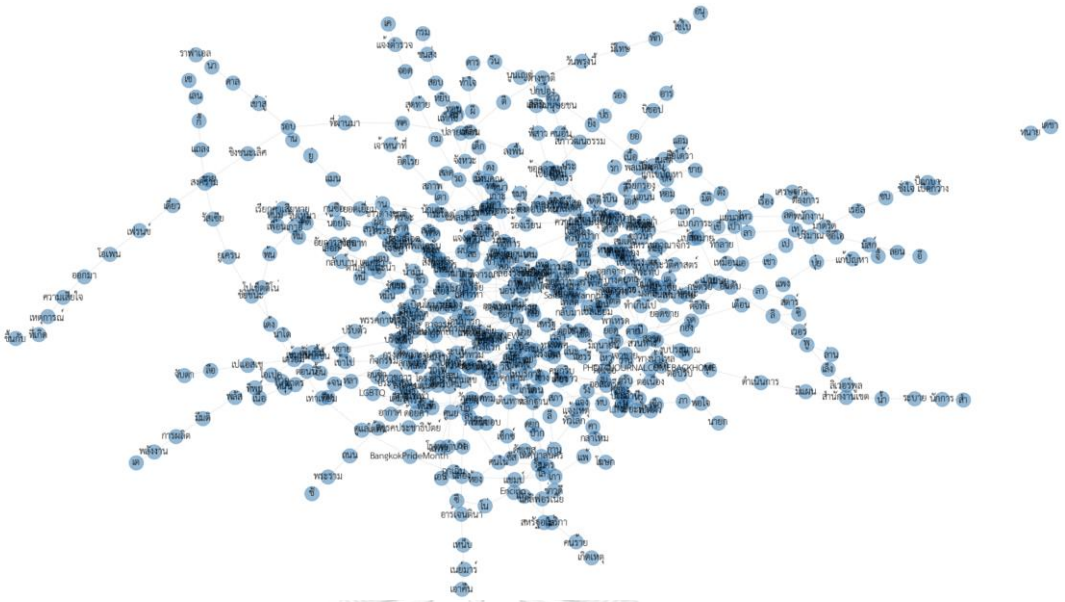


Figure 2: Network graph

3.2 Construct a simplified network

We constructed a network graph in which vertices represent words and edges represent a relationship of consecutive words using bigram. After an original graph is generated in Figure 2, four measures will be calculated from a graph including clustering coefficient, degree, betweenness, and closeness centrality. These measures are sorted in descending order and top n nodes will be chosen for producing a simplified graph. In the preliminary experiment, a top of 30 nodes is used for filtering the original network as shown in Figure 3-5.

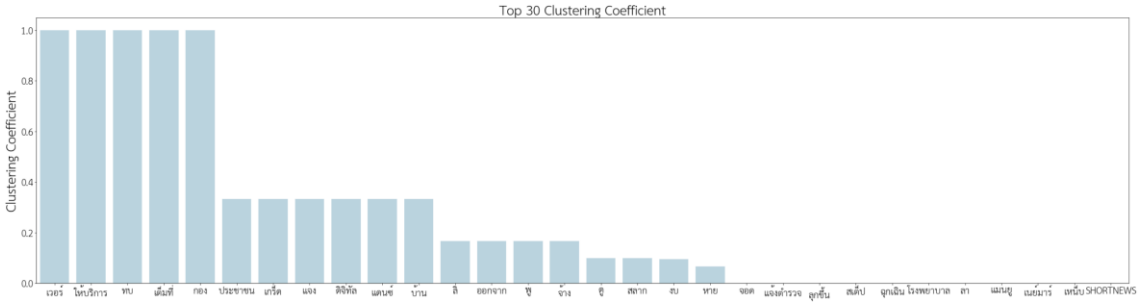


Figure 3: Top 30 nodes of clustering coefficient

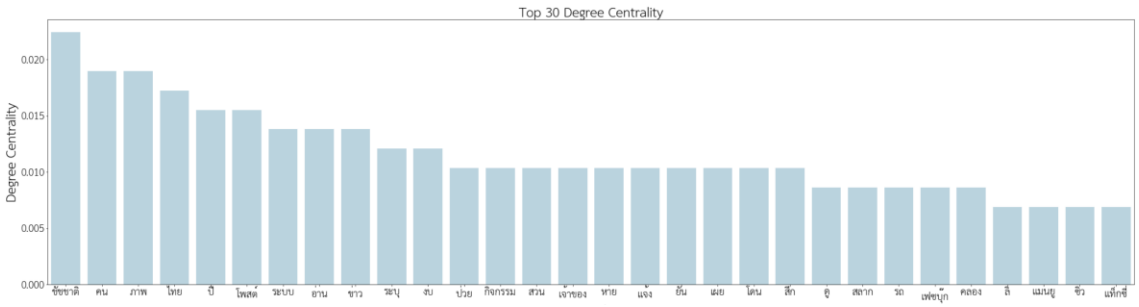


Figure 4: Top 30 nodes of degree centrality

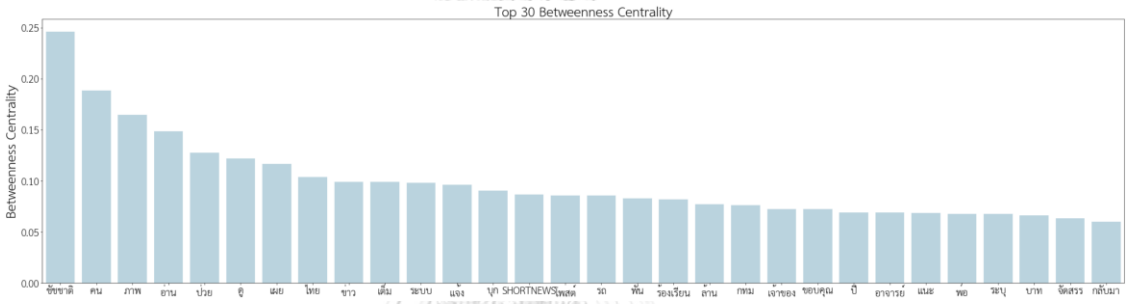


Figure 5: Top 30 nodes of betweenness centrality

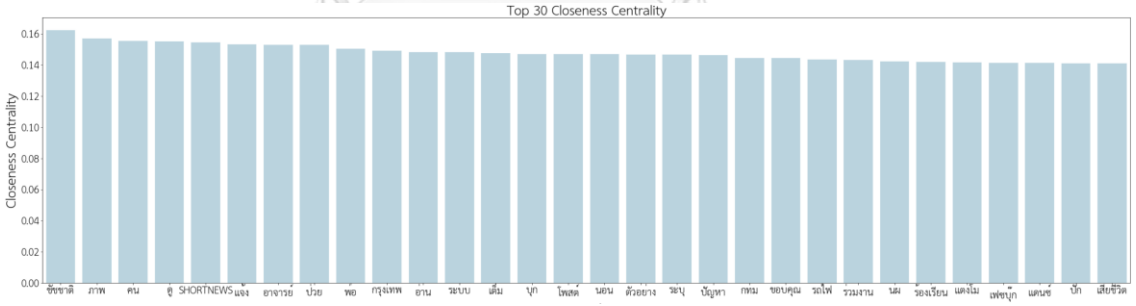


Figure 6: Top 30 nodes of closeness centrality

Filtering top 30 nodes from CF, DC, BC, and CC will give us a group of important nodes that play a significant component on its network. Figure 7 displays nodes that have been selected will be rendered in blue, whereas the light blue color represents nodes that will be removed in a further step.

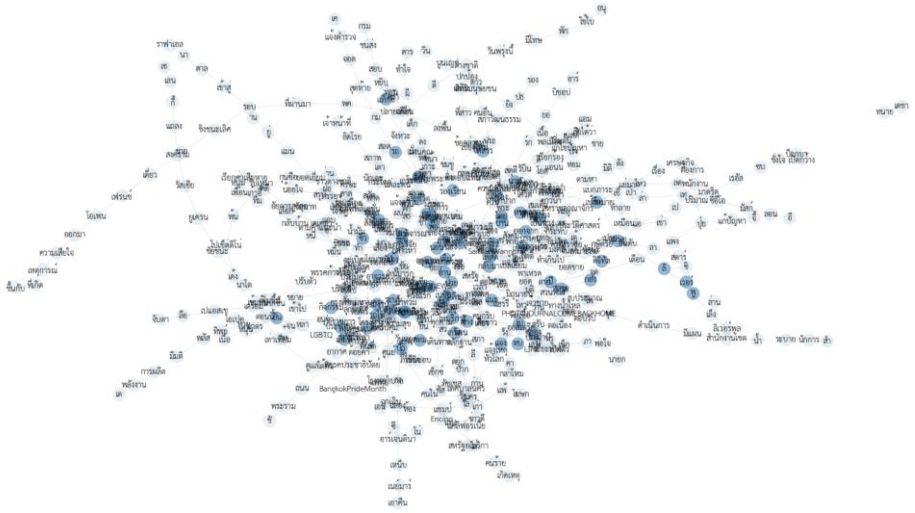


Figure 7: A network graph with selected nodes

As shown in Figure 8, a simplified graph can be modeled by a graph-based simplification process. This process deletes both nodes and edges that do not correspond to prior selected nodes. Finally, a simplified graph will be generated based on the top 30 clustering coefficient, degree centrality, betweenness centrality, and closeness centrality.

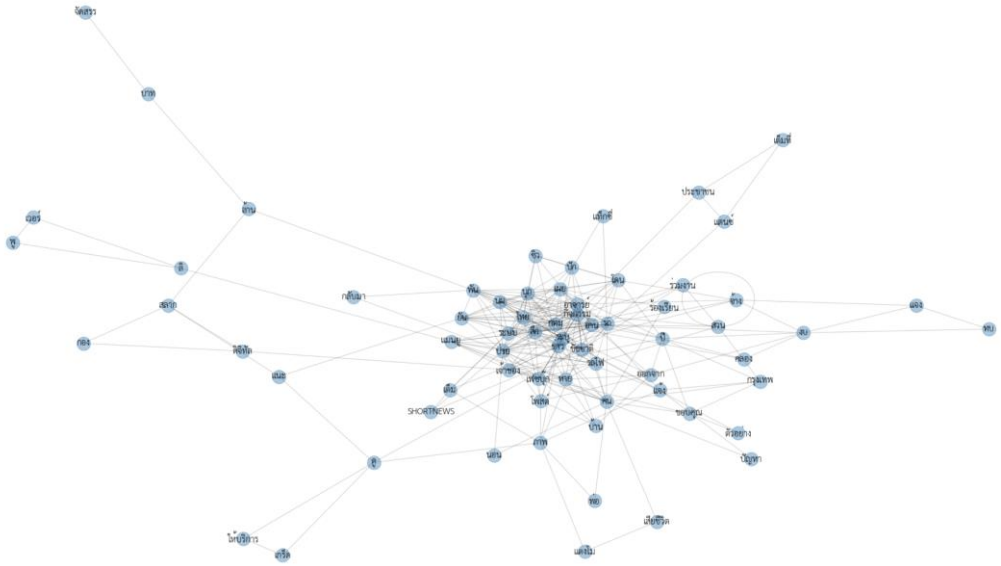


Figure 8: A simplified network graph

3.3 Building graph evolutionary properties

Below is a result table of graph-based evolutionary properties which are produced using simplified graphs on 4-5 June 2022. The graph-based evolutionary properties include a set of added and overlapped nodes/edges that can display how nodes dynamically change over time as shown in Table 2. Furthermore, it gives us two major crucial aspects. Firstly, nodes represent a set of words that many twitter users are talking at that time that means node is the important key headline. Node expansion therefore indicates the change of core events and activities. Secondly, the edge changes show a conversion of connection between nodes. This can significantly assist to detect the event drifts and see the node evolves from one to another.

4 June 2022 - 5 June 2022	Added edges	(เสียชีวิต, เรื่อง), (เสียชีวิต, แตงโม), (เสียชีวิต, ปี), (เสียชีวิต, คลอง), (เสียชีวิต, คน), (เสียชีวิต, นู), (เสียชีวิต, จำนวน), (เสียชีวิต, ไทย), (เสียชีวิต, จีน), (เสียชีวิต, เจอ), (เสียชีวิต, ตาย), (เสียชีวิต, มิถุนายน), (เสียชีวิต, ชัด), (เสียชีวิต, ระบุ), (เสียชีวิต, อ่าน), (เสียชีวิต, โลก), (เสียชีวิต, หญิง), (เสียชีวิต, เตรียม), (เสียชีวิต, สัปดาห์), (เสียชีวิต, ชาย), (เสียชีวิต, วอลเลย์บอล), (เสียชีวิต, GT), (เสียชีวิต, ชัชชาติ), (เสียชีวิต, อ่าน), (เสียชีวิต, ติด), (เสียชีวิต, ใบ), (เสียชีวิต, รัฐบาล), (เสียชีวิต, ประชาชน), (เสียชีวิต, เผย), (เสียชีวิต, ขวิด), (เสียชีวิต, พล), (เสียชีวิต, โพสต์), (เสียชีวิต, เครื่อง), (เสียชีวิต, แม่น้ำ), (เสียชีวิต, ศึก), (เสียชีวิต, สลาก), (เสียชีวิต, ทีม), (เสียชีวิต, ล้าน), (เสียชีวิต, ชิว), (เสียชีวิต, แชง), (เสียชีวิต, สำเร็จ), (เสียชีวิต, ชาติ), (เสียชีวิต, สร้าง), (เสียชีวิต, คว้าแชมป์), (เสียชีวิต, เซ็ต), (เสียชีวิต, เนชั่น), (เสียชีวิต, ตัว), (เสียชีวิต, ขยะ), (เสียชีวิต, คดี), (เสียชีวิต, เอาชนะ), (เสียชีวิต, มาโน่), (เสียชีวิต, เสนอ), (ระบุ, พล), (ระบุ, คลอง), (ระบุ, ตัว), (ระบุ, ส่งผล), (ระบุ, คน), (ระบุ, เรื่อง), (ระบุ, นู), (ระบุ, จำนวน), (ระบุ, จีน), (ระบุ, เจอ), (ระบุ, โลก), (ระบุ, หญิง), (ระบุ, เตรียม), (ระบุ, สัปดาห์), (ระบุ, ชาย), (ระบุ, ชัด), (ระบุ, วอลเลย์บอล), (ระบุ, GT), (ระบุ, ปี), (ระบุ, อ่าน), (ระบุ, ติด), (ระบุ, ใบ), (ระบุ, มิถุนายน), (ระบุ, แตงโม), (ระบุ, รัฐบาล), (ระบุ, ประชาชน), (ระบุ, ขวิด), (ระบุ, เครื่อง), (ระบุ, แม่น้ำ), (ระบุ, ศึก), (ระบุ, สลาก), (ระบุ, ทีม), (ระบุ, ล้าน), (ระบุ, ชิว), (ระบุ, แชง), (ระบุ, สำเร็จ), (ระบุ, ตาย), (ระบุ, ชาติ), (ระบุ, สร้าง), (ระบุ, คว้าแชมป์), (ระบุ, เซ็ต), (ระบุ, เนชั่น), (ระบุ, ขยะ), (ระบุ, คดี), (ระบุ, เอาชนะ), (ระบุ, มาโน่), (ระบุ, เสนอ), (ชัด, เปิดหัว), (ชัด, ชาย), (ชัด, ชาติ), (ชัด, เนชั่น), (ชัด, เผย), (ชัด, วอลเลย์บอล), (ชัด, GT), (ชัด, ชัชชาติ), (ชัด, ปี), (ชัด, ชัย), (ชัด, อ่าน), (ชัด, ติด), (ชัด, ใบ), (ชัด, มิถุนายน), (ชัด, เจอ), (ชัด, แตงโม), (ชัด, ไทย), (ชัด, เรื่อง), (ชัด, รัฐบาล), (ชัด, ประชาชน), (ชัด, เซ็ต), (ชัด, ขวิด), (ชัด, จำนวน), (ชัด, พล), (ชัด, ตาย), (ชัด, โพสต์), (ชัด, เครื่อง), (ชัด, โลก), (ชัด,

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

		หญิง), (ชัด, เตรียม), (ชัด, สัปดาห์), (ชัด, คน), (ชัด, คลอง), (ชัด, แม่น้ำ), (ชัด, ศึก), (ชัด, สลาก), (ชัด, ทีม), (ชัด, ล้าน), (ชัด, อ่าน), (ชัด, นู), (ชัด, จีน), (ชัด, ชิว), (ชัด, แชง), (ชัด, สำเร็จ), (ชัด, สร้าง), (ชัด, คว้าแชมป์), (ชัด, ตัว), (ชัด, ขยะ), (ชัด, คดี), (ชัด, เอาชนะ), (ชัด, มาโน่), (ชัด, เสนอ), (อ่าน, ทีม), (อ่าน, เผย), (อ่าน, ติด), (อ่าน, ใบ), (อ่าน, มิถุนายน), (อ่าน, ปี), (อ่าน, เจอ), (อ่าน, แตงโม), (อ่าน, สลาก), (อ่าน, ล้าน), (อ่าน, ไทย), (อ่าน, ชาย), (อ่าน, วอลเลย์บอล), (อ่าน, GT), (อ่าน, ชัชชาติ), (อ่าน, เรื่อง), (อ่าน, รัฐบาล), (อ่าน, ประชาชน), (อ่าน, ขวิด), (อ่าน, จำนวน), (อ่าน, พล), (อ่าน, โพสต์), (อ่าน, เครื่อง), (อ่าน, โลก), (อ่าน, หญิง), (อ่าน, เตรียม), (อ่าน, สัปดาห์), (อ่าน, คน), (อ่าน, คลอง), (อ่าน, แม่น้ำ), (อ่าน, ศึก), (อ่าน, อ่าน), (อ่าน, ชิว), (อ่าน, นู), (อ่าน, จีน), (อ่าน, แชง), (อ่าน, สำเร็จ), (อ่าน, ตาย), (อ่าน, ชาติ), (อ่าน, สร้าง), (อ่าน, คว้าแชมป์), (อ่าน, เซ็ต), (อ่าน, เนชั่น), (อ่าน, ตัว), (อ่าน, ขยะ), (อ่าน, คดี), (อ่าน, เอาชนะ), (อ่าน, มาโน่), (อ่าน, เสนอ), (สร้าง, ไทย), (สร้าง, นักบินอวกาศ), (สร้าง, เอาชนะ), (สร้าง, คว้าแชมป์), (สร้าง, วอลเลย์บอล), (สร้าง, ตาย), (สร้าง, แม่น้ำ), (สร้าง, โลก), (สร้าง, ชาย), (สร้าง, เซ็ต), (สร้าง, ทีม), (สร้าง, อ่าน), (สร้าง, GT), (สร้าง, ชัชชาติ), (สร้าง, ปี), (สร้าง, ติด), (สร้าง, ใบ), (สร้าง, มิถุนายน), (สร้าง, เจอ), (สร้าง, แตงโม), (สร้าง, เรื่อง), (สร้าง, รัฐบาล), (สร้าง, ประชาชน), (สร้าง, เผย), (สร้าง, ขวิด), (สร้าง, จำนวน), (สร้าง, พล), (สร้าง, โพสต์), (สร้าง, เครื่อง), (สร้าง, หญิง), (สร้าง, เตรียม), (สร้าง, สัปดาห์), (สร้าง, คน), (สร้าง, คลอง), (สร้าง, ศึก), (สร้าง, สลาก), (สร้าง, ล้าน), (สร้าง, นู), (สร้าง, จีน), (สร้าง, ชิว), (สร้าง, แชง), (สร้าง, สำเร็จ), (สร้าง, ชาติ), (สร้าง, เนชั่น), (สร้าง, ตัว), (สร้าง, ขยะ), (สร้าง, คดี), (สร้าง, มาโน่), (สร้าง, เสนอ), (ทีม, ชาติ), (ทีม, ขึ้นไป), (ทีม, จีน), (ทีม, นู), (ทีม, โลก), (ทีม, มาโน่), (ทีม, เอาชนะ), (ทีม, เผย), (ทีม, นักบินอวกาศ), (ทีม, สลาก), (ทีม, ล้าน), (ทีม, ปี), (ทีม, อ่าน), (ทีม, ติด), (ทีม, ใบ), (ทีม, มิถุนายน), (ทีม, เจอ), (ทีม, แตงโม), (ทีม, ไทย), (ทีม, ชาย), (ทีม, วอลเลย์บอล), (ทีม, GT), (ทีม, ชัชชาติ), (ทีม, เรื่อง), (ทีม, รัฐบาล), (ทีม, ประชาชน), (ทีม, ขวิด), (ทีม, จำนวน), (ทีม, พล), (ทีม, โพสต์), (ทีม, เครื่อง), (ทีม, หญิง), (ทีม, เตรียม), (ทีม, สัปดาห์), (ทีม, คน), (ทีม, คลอง), (ทีม, แม่น้ำ), (ทีม, ศึก), (ทีม, ชิว), (ทีม, แชง), (ทีม, สำเร็จ), (ทีม, ตาย), (ทีม, คว้าแชมป์), (ทีม, เซ็ต), (ทีม, เนชั่น), (ทีม, ตัว), (ทีม, ขยะ), (ทีม, คดี), (ทีม, เสนอ), (ชิว, ชัย), (ชิว, นู), (ชิว, จีน), (ชิว, โลก), (ชิว, สัปดาห์), (ชิว, เตรียม), (ชิว, โพสต์), (ชิว, สลาก), (ชิว, คลอง), (ชิว, หญิง), (ชิว, แชง), (ชิว, วอลเลย์บอล), (ชิว, GT), (ชิว, ชัชชาติ), (ชิว, ปี), (ชิว, ชาย), (ชิว, ติด), (ชิว, ใบ), (ชิว, มิถุนายน), (ชิว, เจอ), (ชิว, แตงโม), (ชิว, เรื่อง), (ชิว, รัฐบาล), (ชิว, ประชาชน), (ชิว, เผย), (ชิว, ขวิด), (ชิว, จำนวน), (ชิว, พล), (ชิว, เครื่อง), (ชิว, คน), (ชิว, แม่น้ำ), (ชิว, ศึก), (ชิว, ล้าน), (ชิว, สำเร็จ), (ชิว, ตาย), (ชิว, ชาติ), (ชิว, คว้าแชมป์), (ชิว, เซ็ต), (ชิว, เนชั่น), (ชิว, ตัว), (ชิว, ขยะ), (ชิว, คดี), (ชิว, เอาชนะ), (ชิว, มาโน่), (ชิว, เสนอ), (ชัย, เตรียม), (ชัย, โพสต์), (ชัย, เนชั่น), (ผ่อนคลาย, คน), (ผ่อนคลาย, ศบค), (คน, ชาย), (คน, นักบินอวกาศ), (คน, จำนวน), (คน, ขึ้นไป), (คน, ศบค), (คน, คว้าแชมป์), (คน, เรื่อง), (คน, นู), (คน, เครื่อง), (คน, คลอง), (คน, ไทย), (คน, จีน), (คน, เจอ), (คน, อ่าน), (คน, ปี),

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

		(คน, ติด), (คน, ใบ), (คน, มิถุนายน), (คน, แตงโม), (คน, วอลเลย์บอล), (คน, GT), (คน, ชัชชาติ), (คน, รัฐบาล), (คน, ประชาชน), (คน, เผย), (คน, ขวิด), (คน, พล), (คน, โพสต์), (คน, โลก), (คน, หญิง), (คน, เตรียม), (คน, สัปดาห์), (คน, แม่น้ำ), (คน, ศึก), (คน, สลาก), (คน, ล้าน), (คน, แซง), (คน, สำเร็จ), (คน, ตาย), (คน, ชาติ), (คน, เซ็ต), (คน, เนชั่น), (คน, ตัว), (คน, ขยะ), (คน, คดี), (คน, เอาชนะ), (คน, มาโน่), (คน, เสนอ), (เรื่อง, เผย), (เรื่อง, โพสต์), (เรื่อง, รัฐบาล), (เรื่อง, ประชาชน), (เรื่อง, นู), (เรื่อง, มาโน่), (เรื่อง, สัปดาห์), (เรื่อง, เครื่อง), (เรื่อง, จำนวน), (เรื่อง, คลอง), (เรื่อง, ไทย), (เรื่อง, จีน), (เรื่อง, ติด), (เรื่อง, ใบ), (เรื่อง, มิถุนายน), (เรื่อง, ปี), (เรื่อง, เจอ), (เรื่อง, แตงโม), (เรื่อง, ชาย), (เรื่อง, วอลเลย์บอล), (เรื่อง, GT), (เรื่อง, ชัชชาติ), (เรื่อง, ขวิด), (เรื่อง, พล), (เรื่อง, โลก), (เรื่อง, หญิง), (เรื่อง, เตรียม), (เรื่อง, แม่น้ำ), (เรื่อง, ศึก), (เรื่อง, สลาก), (เรื่อง, ล้าน), (เรื่อง, อ่าน), (เรื่อง, แซง), (เรื่อง, สำเร็จ), (เรื่อง, ตาย), (เรื่อง, ชาติ), (เรื่อง, คว้าแชมป์), (เรื่อง, เซ็ต), (เรื่อง, เนชั่น), (เรื่อง, ตัว), (เรื่อง, ขยะ), (เรื่อง, คดี), (เรื่อง, เอาชนะ), (เรื่อง, เสนอ), (นู, จำนวน), (นู, คลอง), (นู, ไทย), (นู, จีน), (นู, อ่าน), (นู, เจอ), (นู, โลก), (นู, สัปดาห์), (นู, เตรียม), (นู, โพสต์), (นู, สลาก), (นู, หญิง), (นู, แซง), (นู, ชาย), (นู, วอลเลย์บอล), (นู, GT), (นู, ชัชชาติ), (นู, ปี), (นู, ติด), (นู, ใบ), (นู, มิถุนายน), (นู, แตงโม), (นู, รัฐบาล), (นู, ประชาชน), (นู, เผย), (นู, ขวิด), (นู, พล), (นู, เครื่อง), (นู, แม่น้ำ), (นู, ศึก), (นู, ล้าน), (นู, สำเร็จ), (นู, ตาย), (นู, ชาติ), (นู, คว้าแชมป์), (นู, เซ็ต), (นู, เนชั่น), (นู, ตัว), (นู, ขยะ), (นู, คดี), (นู, เอาชนะ), (นู, มาโน่), (นู, เสนอ), (คลอง, คลอง), (คลอง, รัฐบาล), (คลอง, แม่น้ำ), (คลอง, สลาก), (คลอง, โพสต์), (คลอง, จีน), (คลอง, เตรียม), (คลอง, ไทย), (คลอง, สัปดาห์), (คลอง, โลก), (คลอง, ปี), (คลอง, สำเร็จ), (คลอง, จำนวน), (คลอง, หญิง), (คลอง, แซง), (คลอง, เจอ), (คลอง, ใบ), (คลอง, มิถุนายน), (คลอง, ติด), (คลอง, แตงโม), (คลอง, ชาย), (คลอง, วอลเลย์บอล), (คลอง, GT), (คลอง, ชัชชาติ), (คลอง, ประชาชน), (คลอง, เผย), (คลอง, ขวิด), (คลอง, พล), (คลอง, เครื่อง), (คลอง, ศึก), (คลอง, ล้าน), (คลอง, อ่าน), (คลอง, ตาย), (คลอง, ชาติ), (คลอง, คว้าแชมป์), (คลอง, เซ็ต), (คลอง, เนชั่น), (คลอง, ตัว), (คลอง, ขยะ), (คลอง, คดี), (คลอง, เอาชนะ), (คลอง, มาโน่), (คลอง, เสนอ), (ไทย, เอาชนะ), (ไทย, ล้ม), (ไทย, วอลเลย์บอล), (ไทย, ขยับ), (ไทย, ส่งผล), (ไทย, จีน), (ไทย, หญิง), (ไทย, สัปดาห์), (ไทย, เจอ), (ไทย, โลก), (ไทย, เตรียม), (ไทย, สลาก), (ไทย, จำนวน), (ไทย, แตงโม), (ไทย, ปี), (ไทย, ติด), (ไทย, ใบ), (ไทย, มิถุนายน), (ไทย, ชาย), (ไทย, GT), (ไทย, ชัชชาติ), (ไทย, รัฐบาล), (ไทย, ประชาชน), (ไทย, ขวิด), (ไทย, พล), (ไทย, แซง), (ไทย, เครื่อง), (ไทย, แม่น้ำ), (ไทย, ศึก), (ไทย, ล้าน), (ไทย, สำเร็จ), (ไทย, ตาย), (ไทย, ชาติ), (ไทย, คว้าแชมป์), (ไทย, เซ็ต), (ไทย, เนชั่น), (ไทย, ตัว), (ไทย, ขยะ), (ไทย, คดี), (ไทย, มาโน่), (ไทย, เสนอ), (เอาชนะ, สำเร็จ), (เอาชนะ, เซ็ต), (เอาชนะ, จีน), (เอาชนะ, แซง), (เอาชนะ, โอเพน), (เอาชนะ, นักตบ), (เอาชนะ, โลก), (เอาชนะ, ติด), (เอาชนะ, ใบ), (เอาชนะ, มิถุนายน), (เอาชนะ, ปี), (เอาชนะ, เจอ), (เอาชนะ, แตงโม), (เอาชนะ, ชาย), (เอาชนะ, วอลเลย์บอล), (เอาชนะ, GT), (เอาชนะ, ชัชชาติ), (เอาชนะ, รัฐบาล), (เอาชนะ, ประชาชน), (เอาชนะ, เผย), (เอาชนะ, ขวิด), (เอาชนะ, จำนวน), (เอาชนะ, พล), (เอาชนะ,

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

		โพสต์), (เอาชนะ, เครื่อง), (เอาชนะ, หญิง), (เอาชนะ, เตรียม), (เอาชนะ, สัปดาห์), (เอาชนะ, แม่น้ำ), (เอาชนะ, ศึก), (เอาชนะ, สลาก), (เอาชนะ, ล้าน), (เอาชนะ, อ่าน), (เอาชนะ, ตาย), (เอาชนะ, ชาติ), (เอาชนะ, คว้าแชมป์), (เอาชนะ, เนชั่น), (เอาชนะ, ตัว), (เอาชนะ, ขยะ), (เอาชนะ, คดี), (เอาชนะ, มาโน่), (เอาชนะ, เสนอ), (เผย, ติด), (เผย, ตั้งเป้า), (เผย, เจอ), (เผย, ออกความเห็น), (เผย, เตรียม), (เผย, ชัชชาติ), (เผย, เป็นมา), (เผย, รัฐบาล), (เผย, สลาก), (เผย, ประชาชน), (เผย, ล้าน), (เผย, ตัว), (เผย, ขวิด), (เผย, จำนวน), (เผย, วอลเลย์บอล), (เผย, GT), (เผย, ปี), (เผย, ชาย), (เผย, พล), (เผย, ใบ), (เผย, มิถุนายน), (เผย, แตงโม), (เผย, โพสต์), (เผย, เครื่อง), (เผย, โลก), (เผย, หญิง), (เผย, สัปดาห์), (เผย, แม่น้ำ), (เผย, ศึก), (เผย, อ่าน), (เผย, จีน), (เผย, แชง), (เผย, สำเร็จ), (เผย, ตาย), (เผย, ชาติ), (เผย, คว้าแชมป์), (เผย, เซ็ต), (เผย, เนชั่น), (เผย, ขยะ), (เผย, คดี), (เผย, ส่งผล), (เผย, มาโน่), (เผย, เสนอ), (ชาย, แม่น้ำ), (ชาย, โลก), (ชาย, วอลเลย์บอล), (ชาย, GT), (ชาย, ชัชชาติ), (ชาย, ปี), (ชาย, ติด), (ชาย, ใบ), (ชาย, มิถุนายน), (ชาย, เจอ), (ชาย, แตงโม), (ชาย, รัฐบาล), (ชาย, ประชาชน), (ชาย, ขวิด), (ชาย, จำนวน), (ชาย, พล), (ชาย, โพสต์), (ชาย, เครื่อง), (ชาย, หญิง), (ชาย, เตรียม), (ชาย, สัปดาห์), (ชาย, ศึก), (ชาย, สลาก), (ชาย, ล้าน), (ชาย, อ่าน), (ชาย, จีน), (ชาย, แชง), (ชาย, สำเร็จ), (ชาย, ตาย), (ชาย, ชาติ), (ชาย, คว้าแชมป์), (ชาย, เซ็ต), (ชาย, เนชั่น), (ชาย, ตัว), (ชาย, ขยะ), (ชาย, คดี), (ชาย, มาโน่), (ชาย, เสนอ), (สลาก, ใบ), (สลาก, โพสต์), (สลาก, จีน), (สลาก, เตรียม), (สลาก, สัปดาห์), (สลาก, โลก), (สลาก, หญิง), (สลาก, ศึก), (สลาก, แชง), (สลาก, ปี), (สลาก, ติด), (สลาก, มิถุนายน), (สลาก, เจอ), (สลาก, แตงโม), (สลาก, วอลเลย์บอล), (สลาก, GT), (สลาก, ชัชชาติ), (สลาก, รัฐบาล), (สลาก, ประชาชน), (สลาก, ขวิด), (สลาก, จำนวน), (สลาก, พล), (สลาก, เครื่อง), (สลาก, แม่น้ำ), (สลาก, อ่าน), (สลาก, สำเร็จ), (สลาก, ตาย), (สลาก, ชาติ), (สลาก, คว้าแชมป์), (สลาก, เซ็ต), (สลาก, เนชั่น), (สลาก, ตัว), (สลาก, ขยะ), (สลาก, คดี), (สลาก, มาโน่), (สลาก, เสนอ), (ตาย, มิถุนายน), (ตาย, วอลเลย์บอล), (ตาย, คว้าแชมป์), (ตาย, เซ็ต), (ตาย, เตรียม), (ตาย, จำนวน), (ตาย, จีน), (ตาย, เจอ), (ตาย, อ่าน), (ตาย, โลก), (ตาย, หญิง), (ตาย, สัปดาห์), (ตาย, GT), (ตาย, ชัชชาติ), (ตาย, ปี), (ตาย, ติด), (ตาย, ใบ), (ตาย, แตงโม), (ตาย, รัฐบาล), (ตาย, ประชาชน), (ตาย, ขวิด), (ตาย, พล), (ตาย, โพสต์), (ตาย, เครื่อง), (ตาย, แม่น้ำ), (ตาย, ศึก), (ตาย, ล้าน), (ตาย, แชง), (ตาย, สำเร็จ), (ตาย, ชาติ), (ตาย, เนชั่น), (ตาย, ตัว), (ตาย, ขยะ), (ตาย, คดี), (ตาย, มาโน่), (ตาย, เสนอ), (รัฐบาล, ปี), (รัฐบาล, ประชาชน), (รัฐบาล, ติด), (รัฐบาล, ใบ), (รัฐบาล, มิถุนายน), (รัฐบาล, เจอ), (รัฐบาล, แตงโม), (รัฐบาล, วอลเลย์บอล), (รัฐบาล, GT), (รัฐบาล, ชัชชาติ), (รัฐบาล, ขวิด), (รัฐบาล, จำนวน), (รัฐบาล, พล), (รัฐบาล, โพสต์), (รัฐบาล, เครื่อง), (รัฐบาล, โลก), (รัฐบาล, หญิง), (รัฐบาล, เตรียม), (รัฐบาล, สัปดาห์), (รัฐบาล, แม่น้ำ), (รัฐบาล, ศึก), (รัฐบาล, ล้าน), (รัฐบาล, อ่าน), (รัฐบาล, จีน), (รัฐบาล, แชง), (รัฐบาล, สำเร็จ), (รัฐบาล, ชาติ), (รัฐบาล, คว้าแชมป์), (รัฐบาล, เซ็ต), (รัฐบาล, เนชั่น), (รัฐบาล, ตัว), (รัฐบาล, ขยะ), (รัฐบาล, คดี), (รัฐบาล, มาโน่), (รัฐบาล, เสนอ), (เซ็ต, อ่าน), (เซ็ต, โลก), (เซ็ต, คว้าแชมป์), (เซ็ต, วอลเลย์บอล), (เซ็ต,

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

		เปิดหัว), (เซ็ต, ชาติ), (เซ็ต, เนชั่น), (เซ็ต, สัปดาห์), (เซ็ต, GT), (เซ็ต, ชัชชาติ), (เซ็ต, ปี), (เซ็ต, ติด), (เซ็ต, ใบ), (เซ็ต, มิถุนายน), (เซ็ต, เจอ), (เซ็ต, แตงโม), (เซ็ต, ประชาชน), (เซ็ต, ขวิด), (เซ็ต, จำนวน), (เซ็ต, พล), (เซ็ต, โพสต์), (เซ็ต, เครื่อง), (เซ็ต, หญิง), (เซ็ต, เตรียม), (เซ็ต, แม่น้ำ), (เซ็ต, ศึก), (เซ็ต, ล้าน), (เซ็ต, จีน), (เซ็ต, แซง), (เซ็ต, สำเร็จ), (เซ็ต, ตัว), (เซ็ต, ขยะ), (เซ็ต, คดี), (เซ็ต, มาโน่), (เซ็ต, เสนอ), (คดี, แตงโม), (คดี, ฟ้อง), (คดี, ออกความเห็น), (คดี, อ่าน), (คดี, เตรียม), (คดี, จำนวน), (คดี, จีน), (คดี, เจอ), (คดี, โลก), (คดี, หญิง), (คดี, สัปดาห์), (คดี, วอลเลย์บอล), (คดี, GT), (คดี, ชัชชาติ), (คดี, ปี), (คดี, ติด), (คดี, ใบ), (คดี, มิถุนายน), (คดี, ประชาชน), (คดี, ขวิด), (คดี, พล), (คดี, โพสต์), (คดี, เครื่อง), (คดี, แม่น้ำ), (คดี, ศึก), (คดี, ล้าน), (คดี, แซง), (คดี, สำเร็จ), (คดี, ชาติ), (คดี, คว้าแชมป์), (คดี, เนชั่น), (คดี, ตัว), (คดี, ขยะ), (คดี, มาโน่), (คดี, เสนอ), (แตงโม, ปี), (แตงโม, ติด), (แตงโม, ใบ), (แตงโม, มิถุนายน), (แตงโม, เจอ), (แตงโม, วอลเลย์บอล), (แตงโม, GT), (แตงโม, ชัชชาติ), (แตงโม, ประชาชน), (แตงโม, ขวิด), (แตงโม, จำนวน), (แตงโม, พล), (แตงโม, โพสต์), (แตงโม, เครื่อง), (แตงโม, โลก), (แตงโม, หญิง), (แตงโม, เตรียม), (แตงโม, สัปดาห์), (แตงโม, แม่น้ำ), (แตงโม, ศึก), (แตงโม, ล้าน), (แตงโม, อ่าน), (แตงโม, จีน), (แตงโม, แซง), (แตงโม, สำเร็จ), (แตงโม, ชาติ), (แตงโม, คว้าแชมป์), (แตงโม, เนชั่น), (แตงโม, ตัว), (แตงโม, ขยะ), (แตงโม, มาโน่), (แตงโม, เสนอ), (เตรียม, พล), (เตรียม, จีน), (เตรียม, ศบค), (เตรียม, มาโน่), (เตรียม, เสนอ), (เตรียม, ฟ้อง), (เตรียม, ขึ้นไป), (เตรียม, ชัชชาติ), (เตรียม, สัปดาห์), (เตรียม, โลก), (เตรียม, โพสต์), (เตรียม, หญิง), (เตรียม, แซง), (เตรียม, เจอ), (เตรียม, แม่น้ำ), (เตรียม, วอลเลย์บอล), (เตรียม, GT), (เตรียม, ปี), (เตรียม, ติด), (เตรียม, ใบ), (เตรียม, มิถุนายน), (เตรียม, ประชาชน), (เตรียม, ขวิด), (เตรียม, จำนวน), (เตรียม, เครื่อง), (เตรียม, ศึก), (เตรียม, ล้าน), (เตรียม, อ่าน), (เตรียม, สำเร็จ), (เตรียม, ชาติ), (เตรียม, คว้าแชมป์), (เตรียม, เนชั่น), (เตรียม, ตัว), (เตรียม, ขยะ), (ปี, ตั้งเป้า), (ปี, ส่งผล), (ปี, ติด), (ปี, ใบ), (ปี, มิถุนายน), (ปี, เจอ), (ปี, วอลเลย์บอล), (ปี, GT), (ปี, ชัชชาติ), (ปี, ประชาชน), (ปี, ขวิด), (ปี, จำนวน), (ปี, พล), (ปี, ล้าน), (ปี, โพสต์), (ปี, เครื่อง), (ปี, โลก), (ปี, หญิง), (ปี, สัปดาห์), (ปี, แม่น้ำ), (ปี, ศึก), (ปี, จีน), (ปี, แซง), (ปี, สำเร็จ), (ปี, ชาติ), (ปี, คว้าแชมป์), (ปี, เนชั่น), (ปี, ตัว), (ปี, ขยะ), (ปี, มาโน่), (ปี, เสนอ), (โพสต์, เสนอ), (โพสต์, ชัชชาติ), (โพสต์, จีน), (โพสต์, สัปดาห์), (โพสต์, โลก), (โพสต์, เครื่อง), (โพสต์, วอลเลย์บอล), (โพสต์, หญิง), (โพสต์, แซง), (โพสต์, GT), (โพสต์, ติด), (โพสต์, ใบ), (โพสต์, มิถุนายน), (โพสต์, เจอ), (โพสต์, ประชาชน), (โพสต์, ขวิด), (โพสต์, จำนวน), (โพสต์, พล), (โพสต์, แม่น้ำ), (โพสต์, ศึก), (โพสต์, ล้าน), (โพสต์, อ่าน), (โพสต์, สำเร็จ), (โพสต์, ชาติ), (โพสต์, คว้าแชมป์), (โพสต์, เนชั่น), (โพสต์, ตัว), (โพสต์, ขยะ), (โพสต์, มาโน่), (ขวิด, ตัว), (ขวิด, เจอ), (ขวิด, จำนวน), (ขวิด, พล), (ขวิด, ชัชชาติ), (ขวิด, วอลเลย์บอล), (ขวิด, GT), (ขวิด, ติด), (ขวิด, ใบ), (ขวิด, มิถุนายน), (ขวิด, ประชาชน), (ขวิด, เครื่อง), (ขวิด, โลก), (ขวิด, หญิง), (ขวิด, สัปดาห์), (ขวิด, แม่น้ำ), (ขวิด, ศึก), (ขวิด, ล้าน), (ขวิด, อ่าน), (ขวิด, จีน), (ขวิด, แซง), (ขวิด, สำเร็จ), (ขวิด, ชาติ), (ขวิด, คว้าแชมป์), (ขวิด, เนชั่น), (ขวิด, ขยะ), (ขวิด, มาโน่), (ขวิด, เสนอ), (ตัว, เจอ), (ตัว, ติด), (ตัว, ใบ), (ตัว,

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

		มิถุนายน), (ตัว, วอลเลย์บอล), (ตัว, GT), (ตัว, ชัชชาติ), (ตัว, ประชาชน), (ตัว, จำนวน), (ตัว, พล), (ตัว, เครื่อง), (ตัว, โลก), (ตัว, หญิง), (ตัว, สัปดาห์), (ตัว, แม่น้ำ), (ตัว, ศึก), (ตัว, ล้าน), (ตัว, อ่าน), (ตัว, จีน), (ตัว, แชง), (ตัว, สำเร็จ), (ตัว, ชาติ), (ตัว, คว้าแชมป์), (ตัว, เนชั่น), (ตัว, ขยะ), (ตัว, มาโน่), (ตัว, เสนอ), (ล้าน, จำนวน), (ล้าน, แสน), (ล้าน, ใบ), (ล้าน, กระสอบ), (ล้าน, ติด), (ล้าน, มิถุนายน), (ล้าน, เจอ), (ล้าน, วอลเลย์บอล), (ล้าน, GT), (ล้าน, ชัชชาติ), (ล้าน, ประชาชน), (ล้าน, พล), (ล้าน, เครื่อง), (ล้าน, โลก), (ล้าน, หญิง), (ล้าน, สัปดาห์), (ล้าน, แม่น้ำ), (ล้าน, ศึก), (ล้าน, อ่าน), (ล้าน, จีน), (ล้าน, แชง), (ล้าน, สำเร็จ), (ล้าน, ชาติ), (ล้าน, คว้าแชมป์), (ล้าน, เนชั่น), (ล้าน, ขยะ), (ล้าน, มาโน่), (ล้าน, เสนอ), (ติด, มิถุนายน), (ติด, ใบ), (ติด, เจอ), (ติด, วอลเลย์บอล), (ติด, GT), (ติด, ชัชชาติ), (ติด, ประชาชน), (ติด, จำนวน), (ติด, พล), (ติด, เครื่อง), (ติด, โลก), (ติด, หญิง), (ติด, สัปดาห์), (ติด, แม่น้ำ), (ติด, ศึก), (ติด, อ่าน), (ติด, จีน), (ติด, แชง), (ติด, สำเร็จ), (ติด, ชาติ), (ติด, คว้าแชมป์), (ติด, เนชั่น), (ติด, ขยะ), (ติด, มาโน่), (ติด, เสนอ), (คว้าแชมป์, หวด), (คว้าแชมป์, เฟรนช์), (คว้าแชมป์, วอลเลย์บอล), (คว้าแชมป์, โลก), (คว้าแชมป์, GT), (คว้าแชมป์, ชัชชาติ), (คว้าแชมป์, ใบ), (คว้าแชมป์, มิถุนายน), (คว้าแชมป์, เจอ), (คว้าแชมป์, ประชาชน), (คว้าแชมป์, จำนวน), (คว้าแชมป์, พล), (คว้าแชมป์, เครื่อง), (คว้าแชมป์, หญิง), (คว้าแชมป์, สัปดาห์), (คว้าแชมป์, แม่น้ำ), (คว้าแชมป์, ศึก), (คว้าแชมป์, อ่าน), (คว้าแชมป์, จีน), (คว้าแชมป์, แชง), (คว้าแชมป์, สำเร็จ), (คว้าแชมป์, ชาติ), (คว้าแชมป์, เนชั่น), (คว้าแชมป์, ขยะ), (คว้าแชมป์, มาโน่), (คว้าแชมป์, เสนอ), (หวด, เฟรนช์), (ขยับ, ขึ้นไป), (หญิง, วอลเลย์บอล), (หญิง, โลก), (หญิง, จีน), (หญิง, สัปดาห์), (หญิง, แชง), (หญิง, เจอ), (หญิง, GT), (หญิง, ชัชชาติ), (หญิง, ใบ), (หญิง, มิถุนายน), (หญิง, ประชาชน), (หญิง, จำนวน), (หญิง, พล), (หญิง, เครื่อง), (หญิง, แม่น้ำ), (หญิง, ศึก), (หญิง, อ่าน), (หญิง, สำเร็จ), (หญิง, ชาติ), (หญิง, เนชั่น), (หญิง, ขยะ), (หญิง, มาโน่), (หญิง, เสนอ), (แม่น้ำ, ขยะ), (แม่น้ำ, ประชาชน), (แม่น้ำ, โลก), (แม่น้ำ, เจอ), (แม่น้ำ, ใบ), (แม่น้ำ, มิถุนายน), (แม่น้ำ, วอลเลย์บอล), (แม่น้ำ, GT), (แม่น้ำ, ชัชชาติ), (แม่น้ำ, จำนวน), (แม่น้ำ, พล), (แม่น้ำ, เครื่อง), (แม่น้ำ, สัปดาห์), (แม่น้ำ, ศึก), (แม่น้ำ, อ่าน), (แม่น้ำ, จีน), (แม่น้ำ, แชง), (แม่น้ำ, สำเร็จ), (แม่น้ำ, ชาติ), (แม่น้ำ, เนชั่น), (แม่น้ำ, มาโน่), (แม่น้ำ, เสนอ), (นักตบ, จีน), (จีน, ล้ม), (จีน, แชง), (จีน, นักบินอวกาศ), (จีน, สัปดาห์), (จีน, โลก), (จีน, จำนวน), (จีน, เจอ), (จีน, อ่าน), (จีน, วอลเลย์บอล), (จีน, GT), (จีน, ชัชชาติ), (จีน, ใบ), (จีน, มิถุนายน), (จีน, ประชาชน), (จีน, พล), (จีน, เครื่อง), (จีน, ศึก), (จีน, สำเร็จ), (จีน, ชาติ), (จีน, เนชั่น), (จีน, ขยะ), (จีน, มาโน่), (จีน, เสนอ), (สัปดาห์, ส่งผล), (สัปดาห์, เป็นมา), (สัปดาห์, โลก), (สัปดาห์, มาโน่), (สัปดาห์, แชง), (สัปดาห์, เจอ), (สัปดาห์, วอลเลย์บอล), (สัปดาห์, GT), (สัปดาห์, ชัชชาติ), (สัปดาห์, ใบ), (สัปดาห์, มิถุนายน), (สัปดาห์, ประชาชน), (สัปดาห์, จำนวน), (สัปดาห์, พล), (สัปดาห์, เครื่อง), (สัปดาห์, ศึก), (สัปดาห์, อ่าน), (สัปดาห์, สำเร็จ), (สัปดาห์, ชาติ), (สัปดาห์, เนชั่น), (สัปดาห์, ขยะ), (สัปดาห์, เสนอ), (พล, จำนวน), (พล, ชัชชาติ), (พล, วอลเลย์บอล), (พล, GT), (พล, ใบ), (พล, มิถุนายน), (พล, เจอ), (พล, ประชาชน), (พล, เครื่อง), (พล, โลก), (พล, ศึก),

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

		(พล, อ่าน), (พล, แชง), (พล, สำเร็จ), (พล, ชาติ), (พล, เนชั่น), (พล, ขยะ), (พล, มาโน่), (พล, เสนอ), (เจอ, ใบ), (เจอ, มิถุนายน), (เจอ, โลก), (เจอ, วอลเลย์บอล), (เจอ, GT), (เจอ, ชัชชาติ), (เจอ, ประชาชน), (เจอ, จำนวน), (เจอ, เครื่อง), (เจอ, ศึก), (เจอ, อ่าน), (เจอ, แชง), (เจอ, สำเร็จ), (เจอ, ชาติ), (เจอ, เนชั่น), (เจอ, ขยะ), (เจอ, มาโน่), (เจอ, เสนอ), (GT, จำนวน), (GT, เครื่อง), (GT, วอลเลย์บอล), (GT, ชัชชาติ), (GT, ใบ), (GT, มิถุนายน), (GT, ประชาชน), (GT, โลก), (GT, ศึก), (GT, อ่าน), (GT, แชง), (GT, สำเร็จ), (GT, ชาติ), (GT, เนชั่น), (GT, ขยะ), (GT, มาโน่), (GT, เสนอ), (ชัชชาติ, วอลเลย์บอล), (ชัชชาติ, ใบ), (ชัชชาติ, มิถุนายน), (ชัชชาติ, ประชาชน), (ชัชชาติ, จำนวน), (ชัชชาติ, เครื่อง), (ชัชชาติ, โลก), (ชัชชาติ, ศึก), (ชัชชาติ, อ่าน), (ชัชชาติ, แชง), (ชัชชาติ, สำเร็จ), (ชัชชาติ, ชาติ), (ชัชชาติ, เนชั่น), (ชัชชาติ, ขยะ), (ชัชชาติ, มาโน่), (ชัชชาติ, เสนอ), (จำนวน, เครื่อง), (จำนวน, กระสอบ), (จำนวน, วอลเลย์บอล), (จำนวน, ใบ), (จำนวน, มิถุนายน), (จำนวน, ประชาชน), (จำนวน, โลก), (จำนวน, ศึก), (จำนวน, อ่าน), (จำนวน, แชง), (จำนวน, สำเร็จ), (จำนวน, ชาติ), (จำนวน, เนชั่น), (จำนวน, ขยะ), (จำนวน, มาโน่), (จำนวน, เสนอ), (ชาติ, เปิดหัว), (ชาติ, เนชั่น), (ชาติ, ศึก), (ชาติ, อ่าน), (ชาติ, โลก), (ชาติ, วอลเลย์บอล), (ชาติ, ใบ), (ชาติ, มิถุนายน), (ชาติ, ประชาชน), (ชาติ, เครื่อง), (ชาติ, แชง), (ชาติ, สำเร็จ), (ชาติ, ขยะ), (ชาติ, มาโน่), (ชาติ, เสนอ), (เนชั่น, วอลเลย์บอล), (เนชั่น, เปิดหัว), (เนชั่น, ศึก), (เนชั่น, โลก), (เนชั่น, ใบ), (เนชั่น, มิถุนายน), (เนชั่น, ประชาชน), (เนชั่น, เครื่อง), (เนชั่น, อ่าน), (เนชั่น, แชง), (เนชั่น, สำเร็จ), (เนชั่น, ขยะ), (เนชั่น, มาโน่), (เนชั่น, เสนอ), (นักบินอวกาศ, อ่าน), (วอลเลย์บอล, โลก), (วอลเลย์บอล, ใบ), (วอลเลย์บอล, มิถุนายน), (วอลเลย์บอล, ประชาชน), (วอลเลย์บอล, เครื่อง), (วอลเลย์บอล, ศึก), (วอลเลย์บอล, อ่าน), (วอลเลย์บอล, แชง), (วอลเลย์บอล, สำเร็จ), (วอลเลย์บอล, ขยะ), (วอลเลย์บอล, มาโน่), (วอลเลย์บอล, เสนอ), (โลก, อ่าน), (โลก, ร้อน), (โลก, แชง), (โลก, ใบ), (โลก, มิถุนายน), (โลก, ประชาชน), (โลก, เครื่อง), (โลก, ศึก), (โลก, สำเร็จ), (โลก, ขยะ), (โลก, มาโน่), (โลก, เสนอ), (ศึก, เปิดหัว), (ศึก, ใบ), (ศึก, มิถุนายน), (ศึก, ประชาชน), (ศึก, เครื่อง), (ศึก, อ่าน), (ศึก, แชง), (ศึก, สำเร็จ), (ศึก, ขยะ), (ศึก, มาโน่), (ศึก, เสนอ), (เครื่อง, ใบ), (เครื่อง, มิถุนายน), (เครื่อง, ประชาชน), (เครื่อง, อ่าน), (เครื่อง, แชง), (เครื่อง, สำเร็จ), (เครื่อง, ขยะ), (เครื่อง, มาโน่), (เครื่อง, เสนอ), (มิถุนายน, ใบ), (มิถุนายน, ประชาชน), (มิถุนายน, อ่าน), (มิถุนายน, แชง), (มิถุนายน, สำเร็จ), (มิถุนายน, ขยะ), (มิถุนายน, มาโน่), (มิถุนายน, เสนอ), (เสนอ, ศบค), (เสนอ, อ่าน), (เสนอ, ใบ), (เสนอ, ประชาชน), (เสนอ, แชง), (เสนอ, สำเร็จ), (เสนอ, ขยะ), (เสนอ, มาโน่), (ประชาชน, ขยะ), (ประชาชน, ใบ), (ประชาชน, อ่าน), (ประชาชน, แชง), (ประชาชน, สำเร็จ), (ประชาชน, มาโน่), (เฟรนช์, โอเพน), (โอเพน, สำเร็จ), (สำเร็จ, อ่าน), (สำเร็จ, แชง), (สำเร็จ, ใบ), (สำเร็จ, ขยะ), (สำเร็จ, มาโน่), (อ่าน, ร้อน), (อ่าน, แชง), (อ่าน, ใบ), (อ่าน, ขยะ), (อ่าน, มาโน่), (แชง, ใบ), (แชง, ขยะ), (แชง, มาโน่), (แสน, ใบ), (ใบ, ขยะ), (ใบ, มาโน่), (มาโน่, ขยะ)
	Overlapped edges	(ระบุ, ชัชชาติ), (ระบุ, ไทย), (ชิว, ไทย), (ระบุ, โพสต์), (ชิว, อ่าน), (ระบุ, อ่าน), (ระบุ, เผย), (ปี, อ่าน), (ไทย, โพสต์), (ไทย, อ่าน), (สลาก, ล้าน), (ไทย, เผย)

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

	Added nodes	ขยับ, จำนวน, ฟ้อง, ศึก, ขึ้นไป, แม่น้ำ, ออกความเห็น, ตาย, เซ็ต, ร้อน, เฟรนช์, ส่งผล, หญิง, นักบินอวกาศ, เครื่อง, เป็นมา, อ่าน, นักตบ, คดี, ศบค, ใบ, ขยะ, GT, ชัย, ติด, คว้าแชมป์, ตัว, แสน, โลก, ตั้งเป้า, สำเร็จ, เตรียม, เปิดหัว, แซง, ทีม, หวด, รัฐบาล, วอลเลย์บอล, กระสอบ, เจอ, มาโน่, เอาชนะ, เรื่อง, สร้าง, สัปดาห์, โอเพน, เนชั่น, ขวิด, พล, ชาติ, มิถุนายน, นู, ชาย, ผ่อนคลาย, จีน, ชัด, ล้ม, เสนอ
	Overlapped nodes	ล้าน, ประชาชน, อ่าน, เผย, ชัชชาติ, ไทย, โพสต์, คลอง, สลาก, เสียชีวิต, ปี, แตงโม, คน, ระบุ, ชิว

Table 2: Added and overlapped nodes/edges

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

4. Evaluation

To assess the experiment result, the performance will be evaluated based on precision, recall, and f1-score. Precision measures the number of relevant samples out of the total number of relevant predictions. However, a recall is a fraction of correctly relevant predictions and the total number of actual relevant samples. F-score is the harmonic mean between precision and recall. Therefore, these indicators are used for the performance evaluation between our framework and baseline approaches. Searching for any information, we typically input a query that we would like to find. Thus, a query term is required as a reference input before retrieving the information.

In our framework, we initially take a query term to find the edges in the simplified graph. Within 2-hop neighborhoods connected to the vertex are considered as a set of possible edges. Furthermore, we extend a pruning process to aid in keeping only relevant 2-hop vertices by utilizing the top 5 vertices of DC and CF. Then, the final set of edges is used to find word co-occurrence in the sample tweets.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

The first baseline approach, Word Mover's Distance (WMD), is a novel approach that measures the distance that the embedded words of one document require to travel to the embedded words of another document [44]. The embedded query term measures a similarity among the embedded tweet messages. The WMD similarity is implemented using Gensim Python library [45] which provides a useful library to compute the similarity between documents. Tweets that have a similarity score greater than a given threshold will be considered the retrieved tweets.

The second baseline approach called Soft Cosine Measure (SCM) is a state-of-the-art approach to the semantic text similarity task. SCM assesses a similarity

between words using word2vec embeddings [46]. Also, the SCM similarity utilizes the Gensim library to compute a score regarding the query term. Only similarity scores greater than a given threshold will be considered in the retrieved tweets.

Latent Dirichlet Allocation, also known as LDA, is a generative probabilistic model of a text corpus that has been widely studied in diverse areas of natural language processing [47]. LDA is one of the most renowned techniques in topic modeling which facilitates semantic mining and topic discovery among documents. Using cosine similarity to compare a query term and model topics derives a similarity score. Also, a given threshold is defined as a relevant tweet if a similarity score is higher.

To discover tweets related to a query term, different approaches use diverse retrieval methods. After a collection of tweets have retrieved, human-annotated judgment is essential in this study to examine the relevant tweets. Three annotators considered each retrieved tweet message, and their opinions are operated using a majority vote to finalize the result.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

5. Experimental results

5.1 Performance experiment

To assess the performance in the retrieval of relevant documents, we downloaded tweets messages from two popular Thai news agency accounts in June-July 2022. A 14-day window size was used for tracing the propagation of news topics. In the experiment, we evaluated the performance measured by precision, recall, and F-score from 10 news trending topics. A set of parameter configurations in the experiment were given in Table 3.

Parameters	Configurations
Selected top N vertices (Simplified graph)	30
Selected to N vertices (Pruned neighborhoods)	5
Number of topics used in topic models	10
Similarity threshold	0.85

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Table 3: Experimental configurations

The performance from 10 selected topics in the experiment showed that our framework significantly outperformed baseline approaches on all metrics as shown in **Error! Reference source not found.**. To compare among baseline approaches, SCM was a leading of the average recall, and WMD slightly performed better in terms of the average precision. LDA method showed the most F-score on average. However, the average F-score of our framework gained 29.78% compared to the LDA model. Also, the average precision and recall were increased by 31% and 19% respectively.

With the benefits of graph evolution, our framework could widely retrieve tweet messages where the situations had evolved to the others that related to a similar topic. On the other hand, using a similarity score of the word vector embedded approaches might not find relevant tweets when the situations have changed. Furthermore, simplifying and pruning processes assisted to preserve important information and increased the recall score.

Topic No.	Metrics	WMD	SCM	LDA	Our framework
1	Precision	0.48919	0.54304	0.5279	0.99536
	Recall	0.5	0.83535	0.60341	0.75423
	F-score	0.49453	0.52698	0.53294	0.83475
2	Precision	0.50041	0.51564	0.54819	0.90242
	Recall	0.50073	0.66892	0.59732	0.93514
	F-score	0.49916	0.30753	0.56206	0.91811
3	Precision	0.49358	0.50201	0.5065	0.51432
	Recall	0.5	0.51656	0.5482	0.61082
	F-score	0.49677	0.11922	0.49191	0.50285
4	Precision	0.55218	0.51466	0.51975	0.63529
	Recall	0.51989	0.66396	0.60612	0.77829
	F-score	0.52776	0.48295	0.52036	0.68022
5	Precision	0.49367	0.50339	0.50026	0.84116
	Recall	0.5	0.54896	0.50136	0.60653
	F-score	0.49681	0.44813	0.49037	0.66082
6	Precision	0.65879	0.5301	0.5179	0.61858
	Recall	0.50757	0.57057	0.52888	0.62419
	F-score	0.49477	0.28426	0.51732	0.62129
7	Precision	0.54378	0.52918	0.5142	0.66545
	Recall	0.57539	0.56432	0.52867	0.91237
	F-score	0.55189	0.16913	0.51289	0.71907
8	Precision	0.48991	0.51024	0.54985	0.73692
	Recall	0.49943	0.5945	0.72727	0.82098
	F-score	0.49463	0.22009	0.56533	0.77223

Topic No.	Metrics	WMD	SCM	LDA	Our framework
9	Precision	0.53286	0.53376	0.51591	0.83205
	Recall	0.54226	0.5444	0.51944	0.83515
	F-score	0.53577	0.19889	0.51679	0.83359
10	Precision	0.50103	0.51986	0.5096	0.88013
	Recall	0.50508	0.78947	0.53558	0.89423
	F-score	0.48941	0.43507	0.50783	0.88705

Table 4: Overall performance comparison

5.2 Our improvement over the ordinary approaches

Retrieving relevant tweet messages using word-embedded similarity approaches could show poor performance when the query term did not exist in the tweet message or finding related tweets if the situations have developed. With the use of centrality measurement and graph evolutionary properties, our framework performed an outstanding retrieval performance that could not be found in the related tweets using similarity score. Also, a big improvement is when the situation expanded to further topics, our framework could capture the related situation change.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

The examples in the experiment illustrate how our framework improved the retrieval of relevant tweets over the others. The following sections are some of the examples we did in the experiment where similarity-based methods cannot find the related tweets, but our framework does. The first given query example is “ชัชชาติ” who was the talk of the town at that time because he became the Bangkok governor in 2022. When we observe these tweets (1), (2), and (3), there were related to the given query and the query term did not exist in any tweet messages.

(1) “สำรวจราคาสินค้าในตลาดสดพื้นที่กทม. พบต้นหอมปรับราคาขึ้นต่อเนื่อง ขณะที่แม่ค้าขนมจีนโอดต้นทุนแพง ขายทั้งวันได้กำไรเพียง 100 บาท แทบไม่พอใช้จ่ายเลี้ยงครอบครัว”

(2) “เกิดเหตุเพลิงไหม้บ้านเรือนประชาชนภายในชุมชนบ่อนไก่ ถนนพระราม 4 เขตปทุมวัน เพลิงลุกลามอย่างรวดเร็ว เสียหายแล้ว 56 หลังคาเรือน เจ้าหน้าที่ดับเพลิงเร่งฉีดน้ำสกัดไม่ให้ลุกลาม แต่เป็นไปอย่างยากลำบาก เนื่องจากที่เกิดเหตุเป็นซอยแคบ”

(3) “กรุงเทพโพลล์เผยคนกรุงโอดรถเมล์ขาดระยะ ต้องรอรถนาน กระทบหนักทั้งไปเรียนทำงานสายฝากถาม ขสมก.จะต้องทนกับปัญหารถขาดระยะอีกนานแค่ไหน”

Secondly, the pandemic of COVID-19 has spread throughout the world and thousands of people were infected exponentially. Therefore, we selected the word “vaccine” which was the most conversational topic afterward. (4) and (5) are some examples that are considered relevant tweets, but the similarity-based approaches didn’t.

(4) “ศบค ไฟเขียวไฟเซอร์ Marooncap กว่า 3 ล้านโดส สำหรับฉีดให้เด็กอายุ 6 เดือน 5 ปี ซึ่งอยู่ระหว่างรอรับมอบ ด้านนายก ห่วงการระบาดโควิด19 ในโรงเรียน กำชับสธ.กทม.ให้ดูแลอย่างใกล้ชิด”

(5) “ผู้ว่า ชัชชาติ เผยสถานการณ์โควิด19 ในกทม. ยังไม่น่าเป็นห่วง ส่วนใหญ่เป็นผู้ป่วยอาการสีเขียว พบผู้ป่วยสีแดงและเหลืองเพียงร้อยละ 3 เท่านั้น ขอประชาชนฉีดบู๊ทเตอร์โดส ลดความเสี่ยง ป้องกันป่วยหนัก”

Another popular topic that has been mentioned is the prime minister of Thailand. We then set his first name, the word “ประยุทธ์”, as a query term and ran the experiment compared between word-embedded similarity approaches. As shown in (6), (7), (8), (9), and (10), not any tweets contained the first name of the prime minister. However, we can notice that all examples below shared the same word which was his alias name. Our framework could also retrieve related tweets, even though they were not exactly matched to the given query.

(6) “บิ๊กตู่ชื่นชมเยาวชน ภาษาไทยดีเด่น คนรุ่นใหม่อนุรักษ์ภาษา”

(7) “ฝ่ายค้านยอมทุ่ม 30 ชั่วโมง ซักฟอกบิ๊กตู่คนเดียว เพื่อไทยลั่นพร้อมแล้ว เรียงคิวถลก 11 รมต อภิปรายไม่ไว้วางใจ”

(8) “ข่าวหน้า1 ชัชชาติบิ๊กตู่ลุย 4 ภารกิจ ป้องท่วม กทม.ช่วยเหลือคนกรุง”

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

(9) “บิ๊กตู่ เผย ครม.เคาะ คนละครึ่งเฟส 5 คนละ 800 บาท”

(10) “บิ๊กป้อม ยิ้มกริ่ม ชิงฝ่ายค้านไม่เคยมีเพื่อนดี ปมยืมนาฬิกาหรู โบ้ย นายก มือปฏิวัติคนเดียว ลั่นพูดไปเรื่อย 3 ป 3 เปอ ขณะบิ๊กตู่ ยกมือรับ ปฏิวัติเอง”

Therefore, using centrality measurement and graph evolutionary properties took a big advantage over similarity-based methods. Because there were a variety of words that shared the similar meaning which might not be found by computing similarity score. As we mentioned earlier, not only exactly matched tweets, but related tweets and time-evolved situations will also be retrieved.

As shown in Table 5, we would like to show the comparison of actual positive prediction which significantly outperformed any of the ordinary approaches. The f-score on average showed that our framework could find more relevant tweets compared to the others. Relatively, the precision was gained by 81% over LDA method on average. SCM was the leading average recall score that could retrieve the actual related tweets. However, the performance of SCM was not good because it was a lack of precision. The average precision of SCM was the poorest compared to the others. There was the topic no. 3 that showed a very low precision on both baseline approaches and our framework. This is because the total number of true positive classes on the experimental sampled data was relatively small. When considering the tradeoff between precision and recall to better retrieve more relevant tweets is very crucial, using our framework could achieve 88% average f-score, 53.6% average precision, and 58.7% average recall. The average recall of our framework was also increased by 85.86% and 52.13% over WMD and LDA approaches respectively.

Topic No.	Metrics	WMD	SCM	LDA	Our framework
1	Precision	0.05	0.07	0.04	1
	Recall	0.05	0.69	0.1	0.51
	F-score	0.05	0.13	0.05	0.67
2	Precision	0.02	0.04	0.11	0.81
	Recall	0.04	0.96	0.35	0.88
	F-score	0.03	0.07	0.16	0.84
3	Precision	0	0.02	0.04	0.04
	Recall	0	1	0.21	0.33
	F-score	0	0.03	0.06	0.07
4	Precision	0.1	0.04	0.04	0.28
	Recall	0.04	0.51	0.38	0.58
	F-score	0.06	0.07	0.08	0.37
5	Precision	0	0.02	0.05	0.69
	Recall	0	0.38	0.26	0.21
	F-score	0	0.04	0.08	0.33
6	Precision	0.12	0.12	0.22	0.3
	Recall	0.08	0.87	0.22	0.31
	F-score	0.1	0.21	0.22	0.3
7	Precision	0.13	0.06	0.16	0.34
	Recall	0.26	1	0.21	0.92
	F-score	0.17	0.11	0.18	0.49
8	Precision	0.06	0.02	0.05	0.48
	Recall	0.09	0.87	0.33	0.65
	F-score	0.07	0.04	0.09	0.55
9	Precision	0.15	0.1	0.21	0.42
	Recall	0.18	0.96	0.33	0.69
	F-score	0.16	0.17	0.25	0.52
10	Precision	0.02	0.04	0.08	1
	Recall	0.09	0.91	0.42	0.79
	F-score	0.03	0.09	0.13	0.88

Table 5: Predicted positive class performance

6. Conclusion

In this study, we explored the benefits of an evolutionary in graph-based representation for obtaining relevant tweet messages. With the use of a simplification-based technique in graph-based summarization, our framework maintains important information in the network based on centrality measurement and clustering coefficient. Meanwhile, the diffusion of relations is significant to find more relevant information using graph evolution. In the experiment, the proposed framework showed an outstanding performance against word-embedded similarity methods. When the events have developed to the others, using word-embedded vector approaches can only find related tweets on the current situation, but the expanded situations corresponding to the same topic might not be extensively retrieved. In our approach, not only the current situations will be considered as related tweets, but more relevant tweet messages with time-evolved situations will also be retrieved which significantly improved the recall.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

REFERENCES

1. *Most popular social networks worldwide as of October* 2021*, ranked by number of active users*. 2021; Available from: https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/.
2. *News Consumption Across Social Media in* 2021. 2021; Available from: https://www.pewresearch.org/journalism/2021/09/20/news-consumption-across-social-media-in-2021/.
3. *The Most Followed Accounts on Twitter*. 2021; Available from: https://www.brandwatch.com/blog/most-twitter-followers/.
4. *News on Twitter: Consumed by Most Users and Trusted by Many.* 2021.
5. Harary, F. and G. Gupta, *Dynamic graph models.* Mathematical and Computer Modelling, 1997. 25(7): p. 79-87.
6. *Twitter API*. 2022-06-15]; Available from: https://developer.twitter.com/en/docs/twitter-api.
7. Allahyari, M., et al., *Text summarization techniques: a brief survey.* arXiv preprint arXiv:1707.02268, 2017.
8. Mihalcea, R. and P. Tarau, *Textrank: bringing order into texts. Association for Computational Linguistics.* EECS News, 2004.
9. Yongkiatpanich, C. and D. Wichadakul. *Extractive Text Summarization Using Ontology and Graph-Based Method*. in 2019 *IEEE 4th International Conference on Computer and Communication Systems (ICCCS)*. 2019.
10. Joshi, M.L., N. Joshi, and N. Mittal, *SGATS: Semantic Graph-based Automatic Text Summarization from Hindi Text Documents.* ACM Trans. Asian Low-Resour. Lang. Inf. Process., 2021. 20(6): p. Article 102.
11. Natesh, A.A., S.T. Balekuttira, and A.P. Patil, *Graph based approach for automatic text summarization.* International Journal of Advanced Research in Computer and Communication Engineering, 2016. 5(2): p. 6-9.
12. Khan, A., et al., *Abstractive text summarization based on improved semantic graph approach.* International Journal of Parallel Programming, 2018. 46(5): p. 992-1016.
13. Li, W., et al., *Leveraging graph to improve abstractive multi-document summarization.*

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

arXiv preprint arXiv:2005.10043, 2020.

14. Tan, J., X. Wan, and J. Xiao. *Abstractive document summarization with a graph-based attentional neural model*. in *Proceedings of the* 55*th Annual Meeting of the Association for Computational Linguistics (Volume* 1*: Long Papers)*. 2017.
15. Belwal, R.C., S. Rai, and A. Gupta, *A new graph-based extractive text summarization using keywords or topic modeling.* Journal of Ambient Intelligence and Humanized Computing, 2021. 12(10): p. 8975-8990.
16. El-Kassas, W.S., et al., *EdgeSumm: Graph-based framework for automatic text summarization.* Information Processing & Management, 2020. 57(6): p. 102264.
17. Uçkan, T. and A. Karcı, *Extractive multi-document text summarization based on graph independent sets.* Egyptian Informatics Journal, 2020. 21(3): p. 145-157.
18. Ullah, S. and A.A.A. Islam. *A framework for extractive text summarization using semantic graph based approach*. in *Proceedings of the* 6*th international conference on networking, systems and security*. 2019.
19. Liu, Y., et al., *Graph Summarization Methods and Applications: A Survey.* ACM Comput. Surv., 2018. 51(3): p. Article 62.
20. Song, B., et al. *Aggregation-Based Attributed Graph Summarization*. in 2020 *IEEE* 5*th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*. 2020.
21. LeFevre, K. and E. Terzi. *GraSS: Graph structure summarization*. in *Proceedings of the* 2010 *SIAM International Conference on Data Mining*. 2010. SIAM.
22. Riondato, M., D. García-Soriano, and F. Bonchi. *Graph Summarization with Quality Guarantees*. in 2014 *IEEE International Conference on Data Mining*. 2014.
23. Yang, S., et al., *Distributed aggregation-based attributed graph summarization for summary-based approximate attributed graph queries.* Expert Systems with Applications, 2021. 176: p. 114921.
24. Maccioni, A. and D.J. Abadi, *Scalable Pattern Matching over Compressed Graphs via Dedensification*, in *Proceedings of the* 22*nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016, Association for Computing Machinery: San Francisco, California, USA. p. 1755–1764.
25. Chierichetti, F., et al., *On compressing social networks*, in *Proceedings of the* 15*th ACM*

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

SIGKDD international conference on Knowledge discovery and data mining. 2009, Association for Computing Machinery: Paris, France. p. 219–228.

26. Zeqian, S., M. Kwan-Liu, and T. Eliassi-Rad, *Visual Analysis of Large Heterogeneous Social Networks by Semantic and Structural Abstraction.* IEEE Transactions on Visualization and Computer Graphics, 2006. 12(6): p. 1427-1439.
27. Li, C. and S. Lin. *Egocentric Information Abstraction for Heterogeneous Social Networks*. in 2009 *International Conference on Advances in Social Network Analysis and Mining*. 2009.
28. Mehmood, Y., et al. *Csi: Community-level social influence analysis*. in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. 2013. Springer.
29. Hennessey, D., et al. *A Simplification Algorithm for Visualizing the Structure of Complex Graphs*. in 2008 12*th International Conference Information Visualisation*. 2008.
30. Li, Y., Q. Zhang, and T. Reps, *Fast graph simplification for interleaved Dyck-reachability*, in *Proceedings of the* 41*st ACM SIGPLAN Conference on Programming Language Design and Implementation*. 2020, Association for Computing Machinery: London, UK. p. 780–793.
31. Ruan, N., R. Jin, and Y. Huang. *Distance Preserving Graph Simplification*. in 2011 *IEEE* 11*th International Conference on Data Mining*. 2011.
32. Naik, S., et al. *Tweet Summarization: A New Approch*. in 2018 *Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*. 2018.
33. Dusart, A., K. Pinel-Sauvagnat, and G. Hubert, *TSSuBERT: Tweet Stream Summarization Using BERT.* arXiv preprint arXiv:2106.08770, 2021.
34. Chakraborty, R., et al., *Tweet Summarization of News Articles: An Objective Ordering-Based Perspective.* IEEE Transactions on Computational Social Systems, 2019. 6(4): p. 761-777.
35. Panagiotou, N., I. Katakis, and D. Gunopulos, *Detecting events in online social networks: Definitions, trends and challenges.* Solving Large Scale Learning Tasks. Challenges and Algorithms, 2016: p. 42-84.
36. Cordeiro, M. and J. Gama, *Online social networks event detection: a survey*, in *Solving Large Scale Learning Tasks. Challenges and Algorithms*. 2016, Springer. p. 1-41.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

37. Dou, W., et al. *Event detection in social media data*. in *IEEE VisWeek workshop on interactive visual text analytics-task driven analytics of social media content*. 2012.

38. Hansen, D.L., et al., *Chapter* 3 - *Social network analysis: Measuring, mapping, and modeling collections of connections*, in *Analyzing Social Media Networks with NodeXL (Second Edition)*, D.L. Hansen, et al., Editors. 2020, Morgan Kaufmann. p. 31-51.

39. Srinivasan, S., et al., *Chapter Three - Machine learning techniques for fractured media*, in *Advances in Geophysics*, B. Moseley and L. Krischer, Editors. 2020, Elsevier. p. 109-150.

40. Brandes, U., *On variants of shortest-path betweenness centrality and their generic computation.* Social Networks, 2008. 30(2): p. 136-145.

41. Perez, C. and R. Germon, *Chapter* 7 - *Graph Creation and Analysis for Linking Actors: Application to Social Data*, in *Automating Open Source Intelligence*, R. Layton and P.A. Watters, Editors. 2016, Syngress: Boston. p. 103-129.

42. Saramäki, J., et al., *Generalizations of the clustering coefficient to weighted complex networks.* Physical Review E, 2007. 75(2): p. 027105.

43. Junlong, Z. and L. Yu. *Degree Centrality, Betweenness Centrality, and Closeness Centrality in Social Network*. in *Proceedings of the* 2017 2*nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM*2017). 2017. Atlantis Press.

44. Kusner, M.J., et al., *From word embeddings to document distances*, in *Proceedings of the* 32*nd International Conference on International Conference on Machine Learning - Volume* 37. 2015, JMLR.org: Lille, France. p. 957–966.

45. *Word Mover's Distance*. Available from: https://radimrehurek.com/gensim/auto_examples/tutorials/run_wmd.html.

46. Sidorov, G., et al., *Soft similarity and soft cosine measure: Similarity of features in vector space model.* Computación y Sistemas, 2014. 18(3): p. 491-504.

47. Jelodar, H., et al., *Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey.* Multimedia Tools Appl., 2019. 78(11): p. 15169–15211.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

VITA

NAME	Patta Yovithaya
DATE OF BIRTH	7 June 1996
PLACE OF BIRTH	Chonburi, Thailand
INSTITUTIONS ATTENDED	B.Sc. (Hons) Information and Communication Technology, Mahidol University
HOME ADDRESS	27/31 Thawi watthana, Bangkok, 10170

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY