ARTICLE IN PRESS

Digital Communications and Networks xxx (xxxx) xxx

Contents lists available at ScienceDirect

Digital Communications and Networks

journal homepage: www.keaipublishing.com/dcan

Temporal pattern mining from user generated content

Adnan Ali [a,*], Jinlong Li [a], Huanhuan Chen [a], Ali Kashif Bashir [b,c,d]

[a] School of Computer Science and Technology, University of Science and Technology of China, Hefei, 230026, Anhui, China
[b] Department of Computing and Mathematics, Manchester Metropolitan University, Manchester, M15 6BH, United Kingdom
[c] Department of Computer Science and Mathematics, Lebanese American University, Beirut, Lebanon
[d] Woxsen School of Business, Woxsen University, India

ARTICLE INFO

Keywords:
Social media analysis
Collaborative computing
Social data
Twitter data
Temporal patterns mining
Dynamic graphs

ABSTRACT

Faster internet, IoT, and social media have reformed the conventional web into a collaborative web resulting in enormous user-generated content. Several studies are focused on such content; however, they mainly focus on textual data, thus undermining the importance of metadata. Considering this gap, we provide a temporal pattern mining framework to model and utilize user-generated content's metadata. *First*, we scrap 2.1 million tweets from Twitter between Nov-2020 to Sep-2021 about 100 hashtag keywords and present these tweets into 100 User-Tweet-Hashtag (UTH) dynamic graphs. *Second*, we extract and identify four time-series in three timespans (Day, Hour, and Minute) from UTH dynamic graphs. *Lastly*, we model these four time-series with three machine learning algorithms to mine temporal patterns with the accuracy of 95.89%, 93.17%, 90.97%, and 93.73%, respectively. We demonstrate that user-generated content's metadata contains valuable information, which helps to understand the users' collective behavior and can be beneficial for business and research. Dataset and codes are publicly available; the link is given in the dataset section.

1. Introduction

People post their opinions, sentiments, views, ideologies, reviews about products [1], locations [2], and exchange their thoughts on various topics, such as politics [3], health, education, and current affairs on social media platforms, which led to massive user-generated content in recent years. Mining valuable insights and information from this user-generated content is compelling and beneficial for business and research [4]. Collaborative computing is providing state of the art solutions for such information fusion, we contribute in this process with temporal pattern mining from user-generated data by utilizing the Twitter's data. Twitter is a famous social networking platform with 229 million daily active users[1] communicating with each other via tweets and messages and creating enormous amount of data every second. Tweets are 140 characters (before October 2018) or 280 characters (so far) long text, *hashtags*, URLs and multi media content posted by users. Twitter and its content have aroused the interest of researchers in various computer science domains, including interest mining [5], hashtag recommendation [6], text mining [7,8], sentimental analysis [2,9,10], textual analysis [7], cognitive analysis [11], disaster analysis [12], intention mining [13], and community detection [14]. In this research, we mainly focus on hashtags, tweets and users, because users post tweets containing hashtags.

A hashtag starting with the # symbol contains alphanumerical characters without white space. Users create new hashtags by mentioning them in their tweets, or use existing hashtags as keywords to index topics on Twitter. The hashtags in Twitter are part of metadata while metadata is the data of Twitter object[2] including twitter id, creation date, user information, time, etc. The hashtags are employed to express views, feelings, and sentiments [15] toward products [16,17], topics, events [7], incidents [12], places [2], politics [18,19], health [20], or routine life, which makes hashtags interesting for business, non-business, government and non-government organizations. The famous and top hashtags on Twitter (depending on many factors) are called Trends or Trending topics.[3] How does a hashtag become a trend? Any answer will be vague since Twitter algorithms are confidential. Our analysis shows hashtags' popularity on Twitter may last hours to days, rather than weeks. Considering that the popularity of content will grow and fade over time,

* Corresponding author.
E-mail addresses: adnanali@ustc.edu.cn (A. Ali), jlli@ustc.edu.cn (J. Li), hchen@ustc.edu.cn (H. Chen), dr.alikashif.b@ieee.org (A.K. Bashir).

[1] https://www.statista.com/statistics/970920/monetizable-daily-active-twitter-users-worldwide.
[2] https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model/tweet.
[3] https://help.twitter.com/en/using-twitter/twitter-trending-faqs.

https://doi.org/10.1016/j.dcan.2022.11.008
Received 28 February 2022; Received in revised form 26 October 2022; Accepted 1 November 2022
Available online xxxx

the active periods of various trends are different [21]. Finding and utilizing these temporal patterns is a prompting perspective.

The time-variant properties are called temporal patterns, which help to capture events, opinions, and synonyms [22]. Furthermore, tweets are continuous streams of data, and hashtags' frequency of occurrence fluctuates over time as content's popularity is time-dependent [21]. Some hashtags could be interlinked based on temporal similarity regardless of their textual differences [23]. The following are some temporal patterns-based questions: when did a particular hashtag start or end? What is the lifetime of a hashtag? How many users participated in any hashtag in the given time? How many tweets were posted about any given hashtag on a particular day or hour? Answers to these questions are not yet explored in any research so far, yet these answers can lead to a better understanding of user-generated content.

We present a short case study to explain the significance of hashtags' time-variant properties. A hashtag *#BoycottNetflix* was trending on 22-November-2020 in India. A total of 59317 tweets were posted in one day; however, there were less than 100 tweets per day before the mentioned date. After eight days, Netflix India started a hashtag *#NetflixStreamFest* on 01-December-2020 and announced that it would provide free Netflix for Indian users for two days (December 5–6). We are not sure if there exists a connection between the two hashtags. Still, organizations are criticized by customers in such scenarios, organizations may be interested in knowing more about the hashtags, so that countermeasures can be taken to prevent their reputations from being tarnished. We develop a platform to help stakeholders and researchers to understand users' content and behavior better by crossing the limits of textual analysis. With this research and framework, stakeholders will be able to see hidden patterns in Twitter data, which can lead to better decisions for organizations.

Since the same hashtag might refer to different events or contexts at another time, the textual analysis-based research is not enough for temporal-based content [23]. For example, the hashtag #WCFifaFinal does not indicate which year the final was held; without time, the analysis would be insufficient. Researchers used temporal patterns [24, 25] to solve cluster finding, information diffusion patterns, anomalies, and outliers. However, there is very little focus on hashtag temporal patterns regardless of their significance in real life. Marketing companies use Twitter content, such as user feedback and take decisions on sentimental analysis [16], further including temporal properties will be a productive step. We use graphs to analyze hashtags from a temporal point of view.

Graphs are a natural choice to present social media data and have been adopted by Refs. [15,26,27] to model Twitter's content. However, most research only focuses on tweets, users, or hashtags individually, or any two as nodes. In fact, the three elements are interconnected and essential in Twitter, as users post tweets, and tweets contain the hashtags. Therefore, not only some even all elements can be presented as nodes and their relationships as edges. Another most important aspect is time as hashtag mention time indicates hashtag's temporal behavior. This work uses a heterogeneous graph with three types of nodes to store metadata and two types of edges to store time. The hashtag has some important properties in contrast to user and tweet, and this is the only element contributed by multiple users and contains more temporal patterns as hashtags evolve. Presenting hashtags as nodes and incoming edges with *hashtag mention time* information has answered many unanswered questions that we explore in this paper.

Below are the contributions of this paper.

- We provide a temporal pattern mining framework to mine from user-generated content. To mine temporal patterns, it creates the dynamic graph from Tweets' metadata, with twitter objects as input and dynamic graph and scalar properties as outputs.
- We interactively present Twitter data as the graph, which contains three types of nodes: Author, Tweet, Hashtag, and two types of edges. The nodes contain metadata as properties, and the edges contain the temporal information.
- We utilize our framework to mine four different temporal patterns from the dynamic graph. Temporal patterns are used to predict the hashtag's Peak Time (top time of life) and lifetime. The graph is rich in information, and can be used to further mine metadata properties.
- We create 100 dynamic graphs dataset consisting of 1.1 m Users, 2.1 m Tweets and 230K hashtags. We make this dataset publicly available.

The rest of the paper is arranged as follows. We review the literature in Section 2. Section 3 introduces the methodology to explain the time series, and how to create graphs form Twitter data. Results and related discussions are presented in Section 4, and finally, the conclusion and future directions are in Section 5.

2. Literature review

This research consists of three sub-topics graphs, temporal patterns, and hashtags. We discuss the literature of these topics in the below subsections.

2.1. Hashtags

A hashtag is a word or unspaced phrase starting with the # symbol. It can be a single word, such as #Earth or multiple words #Earthday, #Earthday2021. The different uses of hashtags in Twitter can be estimated from Ref. [28], where authors analyzed 60 million hashtags with the findings that half of the hashtags consist of multiple words. Hashtag in research is a well-studied area, and this research can be grouped into *Content-based* and *Content's Properties-based* applications.

Content-based refers to applications that use the text content of hashtags for research purposes. One of the earliest usages is in text mining research, starting from the sentimental analysis. The authors used hashtags in the paper [29], while in another paper [30], they used hashtag and smileys for sentimental analysis. The later referenced article used the #sarcasm hashtag to recognize sarcastic tweets, and the former one proposed a framework for sentiment classification. In another early work [15] on hashtags, authors categorized the hashtags into three types: topic, sentiment, and sentiment-topic hashtags. A model is proposed in Ref. [31] to discover the topics in tweets, and the effectiveness of clustering and classification of hashtags is evaluated. In addition, the use of hashtags is the hashtag recommendation of tweets is presented in recent papers [6,32,33].

Content's properties-based applications are related to hashtag's expansion, evolution, and predictions. Hashtags, called memes in Ref. [34], are extracted from more than 400 million tweets, whose spreading is predicted with linear regression model. Another article [35] proposes hashtag popularity prediction in the next day as a contrast to a weekly basis [34]. The paper [19] about political hashtags studied the factors of the wide adaptation of hashtags during the presidential debate in the US election, and the paper [18] discussed the evolution of political hashtags over time. STRM [36] proposes to model the popularity of hashtags and the accuracy of tweets' popularity prediction. Hashtags are still a famous topic of research, and we can see some of the latest hashtag applications [6,20,37,38] in multiple domains.

There are many articles related to the content, and it is not necessary to mention those unhelpful or irrelevant. On the contrary, metadata did not attract much research attention. However, it also includes valuable information, which needs to be explored and exploited. In this research, hashtags are a node-type of the directed graph, and their metadata is used as node properties.

2.2. Temporal patterns

Time varying properties are called *Temporal Patterns*. Most of the

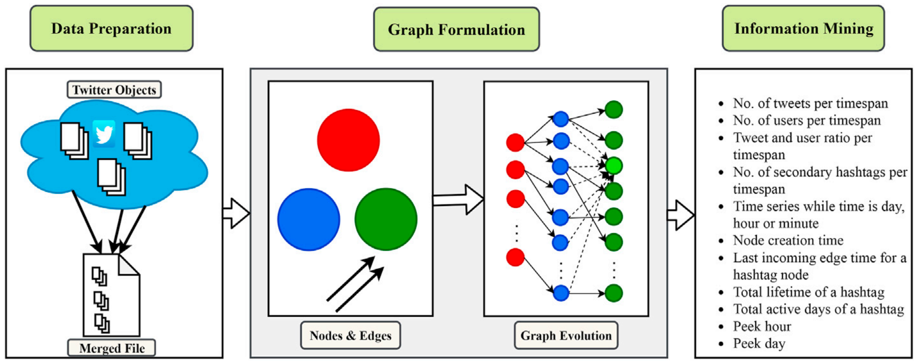

Fig. 1. Model diagram, red = author, blue = tweets, green = secondary hashtags, light green = primary hashtag.

Temporal Patterns research focuses on their applications or uses them as features, and there is no work to explain the process of mining patterns. First, there are some clustering applications. For example, K-Spectral Centroid (K-SC) is a temporal pattern-based clustering algorithm presented in Ref. [21]. The paper's primary focus is temporal patterns and timeseries clustering on Twitter and websites data. A model (SPIKEM) and its implications based on information diffusion patterns are proposed in Ref. [39], and the model explains the rising and falling patterns. A hashtag sense clustering algorithm (SAX*) was proposed in Ref. [23] based on temporal similarity, and authors described that hashtags can have multiple meanings and contexts. Furthermore, the temporal patterns of known events are used to find the related hashtags. A recent paper [22] about temporal patterns was based one the hashtag clustering method, although authors claimed the method as a novel method. The literature mentioned above also used temporal patterns for hashtag clustering.

Second, there are some other applications of temporal patterns. The authors in the paper [40] proposed “Rest, Sleep and Comment”, which matches the four discovered patterns, and can also detect bots and outliers with 94% bots detection accuracy. The paper [34] predicted the idea-spreading on Twitter by combining the social graph topology and content's hybrid approach. The authors used temporal patterns as a feature, but they failed to explore the temporal spreading of the patterns. Another paper [34] focused on longer time frames (around 30 weeks), we argue that trends does not last that long. We further argue that, temporal patterns can be used for predictions, however, no noteworthy work is available, and the discovered patterns are not suitable for prediction. The patterns we found can predict the hashtag lifetime and peak time. We use the real-time Twitter data and mine patterns from it; we also explain the process to mine patterns that can be used for further applications and research.

2.3. Graphs

The graphs on Twitter can be traced back to the beginning of Twitter, but the dynamic graphs on Twitter, in particular, need further research. There are several research areas where graphs are used with Twitter; we will explain the work one by one.

2.3.1. Text mining

In text mining approaches, one of the earliest works on graphs with Twitter's hashtags is paper [16], which is about sentimental analysis based on the hashtags graph model. Hashtags data is collected about ten topics and labeled as positive and negative. Results are related to sentimental analysis, which is not the domain of our paper. Another paper [41] inspired by Ref. [15] used graphs, hashtags, and smileys for sentimental analysis. The author of paper [27] used a bipartite graph, while nodes are users and tweets. It shows the diffusion model to show people's interest in large-scale events, influential users, and the most popular tweet node on the network, although authors did not put the cases in a promising way. Another text mining-based paper [42], where the main goal is to find similarity among tweets and cluster them. A graph-based method for summarizing the tweets and graph is used to show the similarity among tweets. Similar URLs, hashtags, usernames, cosine similarity, and Levenshtein distance are used to calculate the likeness between two tweets and construct a weighted graph, where similarity scores are weight between two tweets nodes. The mentioned similarity is further used to cluster the related nodes in the graph. The dataset of the mentioned work is a bit small, with only 2921 tweets.

2.3.2. Clustering

Here are some-clustering based methods. The maximum *k*-clique method used to find strongly-related groups is proposed in Ref. [24] where frequent words are nodes. An opinion community detection method and opinion leader detection are proposed in the paper [43] with content and time similarity and the topology structure for users. User interactions are presented in graphs where nodes are users, and their replies are edges. A graph-based approach to detect possible spam in tweets is presented in Ref. [25], where nodes are the name entities of tweets, assuming that tweets related to the same topic will have the same entities, and edges are relationships between name entities.

2.3.3. Extraction and prediction

The third group is trend extraction, link prediction, and hashtag prediction. The authors of [44] used the weighted graph and hashtags for trend extraction on Twitter. Nodes are tweets, words, and hashtags. In Ref. [45] vertices are hashtags, and edges are linked to hashtags co-occurrence in the specific tweet. Vertices and edges are weighted according to the number of appearances of hashtags and the number of simultaneous tweets with the hashtags. These weights are used for link prediction. The [34] used a directed graph, where 241K users are nodes and 680K edges are the times that other users mention a user. The paper is about hashtag prediction using undirected hashtag graph where nodes are hashtags and edges are the explicit relationship between nodes.

Table 1
Topics of hashtags.

Topic	Description
Events	Any sudden or planned event. Natural disasters, accidents, Labor Day, spring festival, a national day of any country etc.
Entertainment	Hashtag related to movies, music, TV shows, and celebrities.
Health	Hashtags related to health diseases, medicine, surgeries, hospitals, and viruses.
Politics	Hashtags with political leaders or country names, elections, governments.
Action-Oriented	Hashtags demanding action against any individual, agency, government, or organization.
Criticism-based	Hashtags criticizing any individual, government, agency, or organization.
Sports	Hashtags related to sports activities, players and matches

Table 2
Primary hashtags H_p (each primary hashtag is one dataset).

Hashtag	Coverage	Topic	Hashtag	Coverage	Topic
AFLSaintsDees	Sports	Aus	justiceforbushrarajpar	Action	Pak
alisadpara	Event	Pak	JusticeforUsamaNadeemSatti	Action	Pak
BilalSaeed	Entertainment	Pak	NetflixStreamFest	Entertainment	Ind
ChristineHolgate	Politics	Aus	PAKvsSA	Sports	Pak
CoronavirusPak	Health	Pak	earthquake	Event	World
Hazaragenocide	Criticism	Pak	WorldHealthDay2021	Event	World
WearAMask	Health	World			

2.3.4. Dynamic graphs

The forth group is influence, influencers and dynamic graphs. The research [46] measured the influence of users on Twitter by in-degree, retweets, and mentions. They also did a temporal analysis to point out how different types of influential individuals interact with the audience. For dynamic graphs, the most prominent research area is to find influencers in dynamic graphs. The paper [47] tracks influential individuals in the dynamic graphs with different datasets and Twitter. Although the paper is about influential nodes and different from our work in pattern recognition, it is one of the earliest works in tracking influential nodes in dynamic graphs. Another research work on influence maximization in dynamic graphs is [48], which used the Twitter dataset. In this work, we are using dynamic graphs, as social media's dynamic behavior and its temporal properties are convenient to manage and mine in/from graphs.

This section concludes the published work where hashtag, temporal patterns and dynamic graphs' usage is explained. We summarize that the published work only focuses on the content of tweets, and ignores the tweeting patterns. We argue that different hashtags can be important for different organizations when a group of users is posting a specific topic, and that group can be just bots. Text mining-based approaches are not enough and exploring beyond the text of a tweet will add new dimensions to user-generated content mining. In this work, we contribute to explore the user-generated content in terms of metadata.

3. Methodology

This section elaborates the working structure of the framework as illustrated in Fig. 1. The model diagram includes three components: *Data Preparation*, *Graph Formulation*, and *Information Mining*. In the *Data Preparation*, Twitter objects are downloaded with Twitter API. These objects contain the tweets' content and metadata information, then objects are merged into a file for further processing. The *Graph Formulation* has two steps. Firstly, the objects are processed, and each object yields a tweet node, an author node, hashtag nodes, and metadata information. Secondly, the notes generate a temporally evolving dynamic graph, where nodes contain metadata information as attributes, and edges contain tweet timestamps as edge weight. Finally, in the *Information Mining*, time-based properties are mined from the dynamic graph, and these properties can be user-focused, tweet-focused, and hashtag-focused.

Data preparation comprised of two steps: *Hashtag Selection* and *Dataset Construction*. In the first step, we choose the hashtags to be finalized; in the second step, we create the dataset from the chosen hashtags.

3.1. Hashtag selection

Since, Twitter has millions of hashtags [28], choosing a representative hashtag is challenging. We define two types of hashtags:

1. Primary hashtags (H_p): They are used as the search query to download tweets. These are our representative hashtags for this paper in Table 2. We download a hundred datasets, and each dataset is a *collection of tweets objects* about one primary hashtag. One dataset has only one primary node in it. If two hashtags are related to same topic we chose the one with more tweets.
2. Secondary hashtags (H_s): These are the hashtags other than primary hashtags mentioned in the tweets. These hashtags are not used in the search query, and they co-exist in the same tweets with H_p. There is no upper or lower limit of these hashtags in the dataset.

To choose diverse H_p, hashtag popularity, topics and usage locations are considered and to have the diversity of temporal patterns, data download time is also different. The website *getdaytrends*[4] is used to choose the top hashtags, and Twitter[5] is used to select the less famous and unpopular hashtags. The list of topics is given in Table 1, while Table 2 shows the H_p. As different countries have a different number of Twitter and internet users; for diversity we consider Pakistan, India, United States, England, and Australia. A H_p trending in more than one country is marked as worldwide.

3.2. Hashtag selection for downloading

1. Find hashtags from *getdaytrends* and Twitter. Candidate hashtags may or may not be trending.
2. Match if the hashtag topic is from Table 1.
3. Check the country/region where the hashtag is used, which can be facilitated by *getdaytrends*.
4. Check the number of tweets (both websites provide tweet count against hashtag).

3.3. Dataset construction

Here are the steps to download the selected hashtags.

Function $\delta(\cdot)$ takes H_P as the input, downloads tweet objects[6] containing the search query H_p and outputs a downloaded file j_p, as shown by Eq. (1).

$$j_p = \delta(H_p) \tag{1}$$

where the downloaded file $j_p = \{t_{p_1}, t_{p_2}, \ldots, j_{p_e}\}$ contains p_e tweet objects, and each tweet is comprised of the tweet content. Furthermore, all tweet objects are merged as a set J, as shown in Eq. (2).

$$J = \cup\{j_p\} \tag{2}$$

3.3.1. Graph formulation

To represent the relationships among Twitter objects, we convert the tweet objects in J into the heterogeneous yet directed acyclic dynamic graph. First, we define the graph, its components and notions, then dynamic and temporal properties of graphs that explain how the graph evolves from the first node to the last edge.

3.3.1.1. User-Tweet-Hashtag (UTH) dynamic graph. The UTH dynamic graph $\boldsymbol{G} = (N, E, T_n, T_e)$ is 4-tuple, including nodes $\boldsymbol{N} = \{n_1, n_{2,3}, \ldots, _w\}$,

4 https://getdaytrends.com.
5 https://twitter.com.
6 https://developer.twitter.com/en/docs/twitter-api/v1/data-dictionary/object-model/tweet.

Table 3
Metadata of tweet as attributes of all three types of nodes.

Author node (red colored)		Tweet node (blue colored)	
Attribute	Description	Attribute	Description
ID	Twitter user ID	ID	Tweet's unique ID across Twitter
Name	Username on Twitter	User ID	User who posted the tweet
Verification	User is verified by Twitter or not	Retweet Count	Number of retweets
Followers	Number of people following the user	Favorite Count	Number of favorites on tweet
Following	Number of people followed by user	Node Type	Blue
Node Type	Red		
Account Date	Date of Twitter account creation		
Hashtag node (green colored)			
ID	Hashtag ID	Name	Hashtag Text
Status	Active/Not Active	Node Type	Green

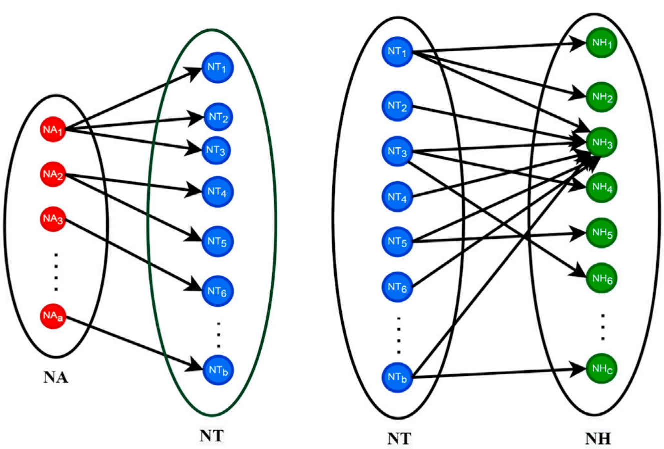

Fig. 2. Nodes relation from one set of nodes to other set. Set ***NA*** has one-to-many relation with ***NT*** (One user can post many tweets), and ***NT*** has many-to-many relation with ***NH*** (multiple tweets can have multiple hashtags).

edges $\boldsymbol{E} = \{e_1, e_2, e_3, \ldots, e_x\}$. The node category $\boldsymbol{T_n}$ and edge category $\boldsymbol{T_e}$ represent the category of node and edge, respectively. There are three categories of nodes: *Authors, Tweets, Hashtags*, and two categories of Edges: *Author to Tweet* and *Tweet to Hashtag*. The $\boldsymbol{G}$ is heterogeneous since it has node categories and directed acyclic behavior, and there is only one directed edge between each pair of nodes.

3.3.1.1.1. Nodes and edges. We identify the following three types of nodes from Twitter objects to create the graph. Every node has metadata as attributes, which are mentioned in Table 3, and time values are saved in edges $\boldsymbol{E}$. To exemplify the nodes and the edges relation of the graph, we visually divide it into two sections in Fig. 2.

Authors

The authors are Twitter users, and any user posting a tweet can be part of the graph. They are indicated as red color nodes in graph, Fig. 2. $\boldsymbol{NA}$ a is set of all author nodes $\boldsymbol{NA} = \{na_1, na_2, na_3, \ldots, na_a\}$, while each node presents one user.

Tweets

Tweets refers to the written content that consists of text, URLs, and hashtags. A tweet should have at least one hashtag to be qualified to be in the graph. It is indicated as blue color nodes in Fig. 2. $\boldsymbol{NT}$ is the set of all tweet nodes $\boldsymbol{NT} = \{nt_1, nt_2, nt_3, \ldots, nt_b\}$, where each tweet is one node.

Hashtags

Hashtags are words, phrases, or characters string starting with # in tweets. A hashtag should be mentioned in two tweets to make its place in the graph. It is indicated as green color nodes in Fig. 2. $\boldsymbol{NH}$ is the set of all hashtag nodes $\boldsymbol{NH} = \{nh_1, nh_2, nh_3, \ldots, nh_c\}$, and each hashtag is one node.

$$N = NA \cup NT \cup NH \tag{3}$$

where $\boldsymbol{N}$ is union of authors, tweets, and hashtag nodes. There are two types of Directed Edges $\boldsymbol{E}$, while every pair of nodes have only one edge connecting them. These edges present the relations between nodes.

Author to tweet E_{AT}

It refers to the edge between the authors and tweet nodes ***E(NA, NT)***. Any author node $\boldsymbol{na_a}$ can have more than one edge to $\boldsymbol{NT}$ nodes, so this is a 1 to n mapping as one author can post multiple tweets, Fig. 2. $\boldsymbol{E_{AT}}$ is set of all edges $\boldsymbol{E_{AT}} = \{e_{at1}, e_{at2}, e_{at3}, \ldots, e_{at\ d}\}$

Tweet to hashtag E_{TH}

It is the edge between tweets and hashtag nodes ***E(NT, NH)***. A n to n mapping given that many tweets can have many hashtags, Fig. 2. $\boldsymbol{E_{TH}}$ is the set of all edges $\boldsymbol{E_{TH}} = \{e_{th1}, e_{th2}, e_{th3}, \ldots, e_{th\ e}\}$.

$$E = E_{AT} \cup E_{TH} \tag{4}$$

$$G = N \cup E \tag{5}$$

3.3.1.1.2. Graph evolution. Once a user posts the tweet, temporal graphs evolves. It causes the upgrade in the graph, and the following two updates including five operations:

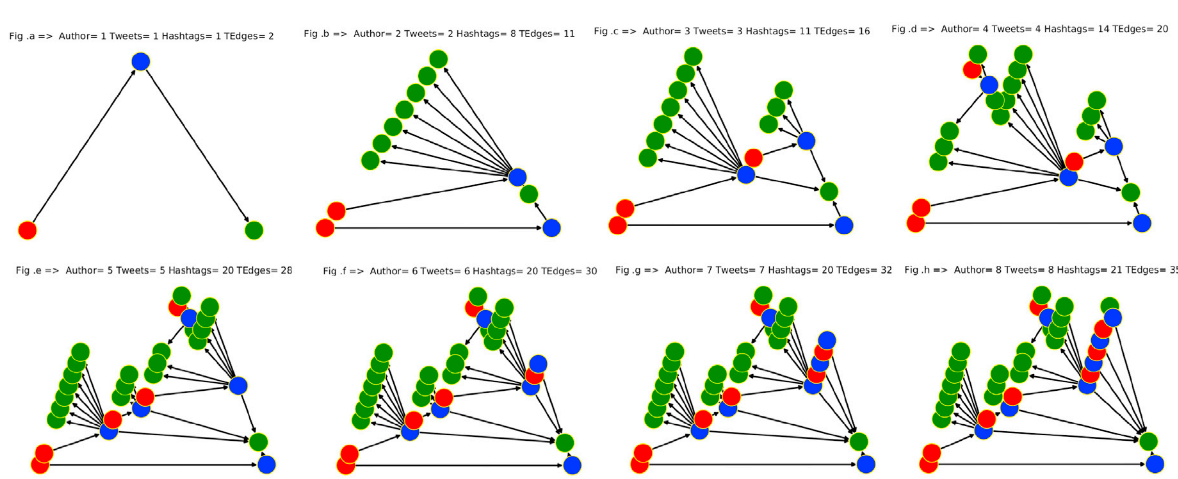

Fig. 3. Graph evolution from 1 tweet to 8 in T time, where red nodes (users), blue nodes (tweets) and green nodes (hashtags) increasing with different pace.

1. Three necessary operations are: at-least one new tweet node $\boldsymbol{NT}$ and two edges $\boldsymbol{E}_{AT} = \boldsymbol{E}(\boldsymbol{NA}, \boldsymbol{NT})$, $\boldsymbol{E}_{TH} = \boldsymbol{E}(\boldsymbol{NT}, \boldsymbol{NH})$ created in the graph.

$$\boldsymbol{NT} \cup \{nt\},\ \boldsymbol{E}_{AT} \cup \{e_{at}\},\ \boldsymbol{E}_{TH} \cup \{e_{th}\}$$

2. Two possible operations are: one new user node $\boldsymbol{NA}$ and n hashtag nodes $\boldsymbol{NH}$ insertion in the graph. They may or may not happen depending on whether the nodes are already in the graph.

$$\boldsymbol{NA} \cup \{na\},\ \boldsymbol{NH} \cup \{nh\}$$

Fig. 3 demonstrates the working structure of all five operations mentioned above (three nodes and two edges) in Fig. 3(a). In 3(b), one hashtag node already exists, so the only edge to that node is created while other actions are the same as 3(a).

3.3.2. Information mining from graph

UTH Graph G is a dynamic graph and contains two types of information: *Temporal Series* and *Temporal Scalars*. Some events continue to occur over time, such as new tweets posted by users. We present these events in the *Temporary Series*. On the contrary, some events have only one value, such as the *peak tweet posting hour* of a hashtag, such which will be presented in *Temporal Scalars*.

3.3.2.1. Temporal series. Time/temporal series refer to the sequence of events with equal time intervals, which can be extracted from the dynamic graph and we extract four timeseries as follows:

1. Percentage of the number of tweet nodes
2. Quadrants of tweets nodes
3. Percentage of tweets and user nodes
4. Percentage of H_s nodes

For all timeseries, we denote a timespan τ, in which the number of events that occurred is counted. The value of τ can be a day, an hour, or a minute.

3.3.2.1.1. Percentage of the number of tweet nodes. Nodes insertion in graphs varies with time. To display the changes in the number of tweet nodes as time passes, we define the timeseries as the *number of tweet nodes per timespan* τ by Eq. (7). In order to do that first, we find the relative value TF_i^τ of tweet nodes in τ.

$$TF_i^\tau : \frac{f_t^i \times 100}{max\{f_t^1, f_t^2, \ldots\}} \tag{6}$$

Different hashtags nodes have different number of tweet nodes that mentioned the hashtag. To make it general for every graph, we define relative tweet nodes frequency, where f_t^i is the tweet nodes in the graph (G) added at the i-th timespan. This is a *relative Tweet-frequency* percentage with the ratio of i-th value and maximum value. That is why TF^τ value will be between 0 and 100. The equation for timeseries of tweet nodes per time TD^τ is:

$$TD^\tau = \{TF_1^\tau, TF_2^\tau, TF_3^\tau, \ldots\} \tag{7}$$

where TD^τ is series of TF^τ variations. An entry is appended in series after every τ to obtain a time-series of events, as shown in Eq. (7).

3.3.2.1.2. Quadrants of the number of tweet nodes. In smaller time-spans (minutes), quantity of new tweet nodes fluctuates faster. This irregularity causes irregularity in timeseries, and grouping these tweet node frequencies in quarters leads to different behavior in timeseries. Here we define the quadrants of the number of tweet nodes. As TF_i^τ values range from 0 to 100, we define it as q_n, which means the quarter with 25 differences and n is between 1 and 4.

$$Q_i^\tau = \begin{cases} q_1 & TF_n^\tau \in \{1, 2, \ldots, 25\} \\ q_2 & TF_n^\tau \in \{26, 27, 28, \ldots, 50\} \\ q_3 & TF_n^\tau \in \{51, 52, 53, \ldots, 75\} \\ q_4 & TF_n^\tau \in \{76, 77, 78, \ldots, 100\} \end{cases} \tag{8}$$

where $Q_i^\tau := q_n$ and $(QP)^\tau$ are a sequence of quarters derived from Eq. (7). The following is the timeseries QP_i^τ of quadrants with timespan τ.

$$QP_i^\tau = \{Q_1^\tau, Q_2^\tau, Q_3^\tau, \ldots\} \tag{9}$$

The lengths of series 7 and series 9 will be equal for the same time-span. As it is a series of quadrants, resulting a smoother curve to mine patterns and make predictions.

3.3.2.1.3. Tweet and user nodes ratio. Many users post many tweets, leading to *many-to-many* relation between users and tweets. One user node can have more tweet child nodes, while a tweet node can not have multiple parents or authors. The ratio between tweets and users is:

$$R_i^\tau = \frac{f_t^i}{f_u^i}$$

where f_u^i is the number of user nodes that causes at least one new tweet node addition, f_t^i is the number of tweet nodes, and R_i^τ is the ratio, which varies per τ timespan. Considering that tweets-set will be always greater than or equal to users ($\boldsymbol{NT} \geq \boldsymbol{NA}$), R_i^τ value will always be greater than or equal to one ($R_i^\tau \geq 1$). The ratios are presented in the timeseries in Eq. (10).

$$TUR^\tau = \{RP_1^\tau, RP_2^\tau, RP_3^\tau, \ldots\} \tag{10}$$

where TUR^τ is the timeseries of tweets and user nodes ratio. The length of the timeseries depends on the author nodes set $\boldsymbol{NA}$. As long as new users continue to post, the values will continue to be appended. However, if only old users post tweets, new ratios will remain zero.

3.3.2.1.4. Hashtags mention timeseries. Every time a user mentions a new hashtag, a hashtag node is created. Although, if the hashtag already exists in UTH dynamic graph G, just a new edge is created. Here we focus on the new hashtag nodes. Let f_h^i be the number of hashtag nodes created at i-th timespan. The *New Hashtags Mention* value:

$$HF_i^\tau : \frac{f_i^h \times 100}{max\{f_1^h, f_2^h, f_3^h, \ldots\}} \tag{11}$$

where HF_τ is the hashtag ratio relative to maximum f_i^h value. The timeseries of HF_τ is defined as follows:

$$HP^\tau = \{HF_\tau^1, HF_\tau^2, HF_\tau^3, \ldots, HF_\tau^i\} \tag{12}$$

where HP^τ is the series of HF^τ per τ timespan. Initially, it is an empty set $\emptyset$; however, $HF\tau$ is keep appended in HP_τ as τ timespan and results as Eq. (12). As long as users continue to use new hashtags, the length of the timeseries will keep changing, and the graph will be updated.

The above-defined timeseries are for three timespans; the effect of the timespan will be the change in length of timeseries and number of occurrences. For example, if we have seven days of data, the size of the timeseries will be 7 days, 168 hours, or 10,080 minutes. In the case of the number of occurrences, values per timespan will decrease as the

Table 4
Temporal scalars from the UTH dynamic graph G.

Scalar	Notion	Description
Peek Hour	P_h	Hour when Hashtag was most active.
Hashtag Creation Time	NH_{time}	Time when hashtag node was created.
Last Edge Creation Time	LE_{time}	Time when last edge was created to top Hashtag.
Total Days	T_d	Total number of days of Hashtag.
Active Days	A_d	No. of days when tweet nodes were more than 100.

Table 5
Dataset stats.

Property	Description
Software Package	NetworkX
Python Version	3.9.5
Dataset Start Date	Nov-2020
Dataset End Date	Sep-2021
Time Span in Days	294
Tweets Nodes	2.1 m
User Nodes	1.1 m
Hashtag Nodes	230K
Total Nodes	3.44 m
Total Graphs	100
Graph Format	.gpickle

timespan downgrades.

3.3.2.2. Temporal scalars. The UTH dynamic graph G has time based scalar properties, we refer as temporal scalars. These temporal scalars contain useful information to understand user-generated content and its expansion. Table 4 presents the temporal-based scalar values we extract from the dynamic graph. These temporal scalar values help us to understand topics and location-based behavior leading to the hypothesis that *timeseries have the topics and location-based temporal patterns*.

4. Results and discussion

In this section, we first present our dataset as dynamic graphs. Based on it, four kinds of timeseries and four kinds of temporal patterns are mined from them are demonstrated. To verify the functions of our framework, we apply three different models, and tune their parameters to fit with data. There are multiple timeseries models which can be used for data fitting. However, the scope of this paper is pattern mining rather than the best model selection. So, to keep it simple, we use the *Polynomial Regression (PR), Auto-Regressive Integrated Moving Average Model (ARIMA)* and *Long Short-Term Memory Network (LSTM)* for timespan day, hour and minute, respectively. To model the timeseries, time is an independent variable, and the number of occurrences of the event is used as the dependent variable. It is to be noticed that making predictions is not the goal of this paper; rather, we perform it to elaborate our research's significance.

4.1. Dataset stats

The downloaded dataset from Twitter is converted into UTH dynamic graphs with NetworkX [49] and python. NetworkX is a python package to create complex and dynamic graphs and networks, while it is used here for dynamic graphs creation, processing and temporal pattern mining. It supports multiple graph formats, and we create graphs in these formats (Graph Exchange XML Format), Pickled Graphs, GraphML, and Pajek. We

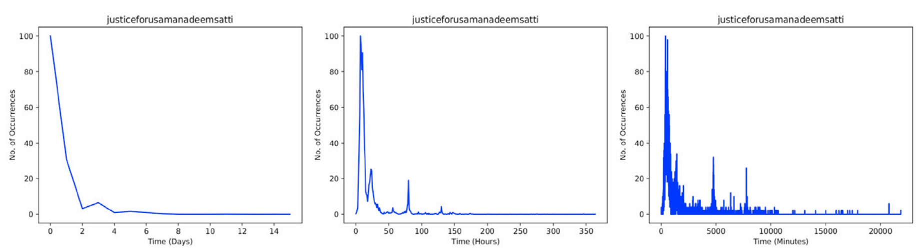

Fig. 4. Relative number of occurrences of a hashtag (#JusticeForUsamaNadeemSatti) while τ = Day, Hour, and Minute.

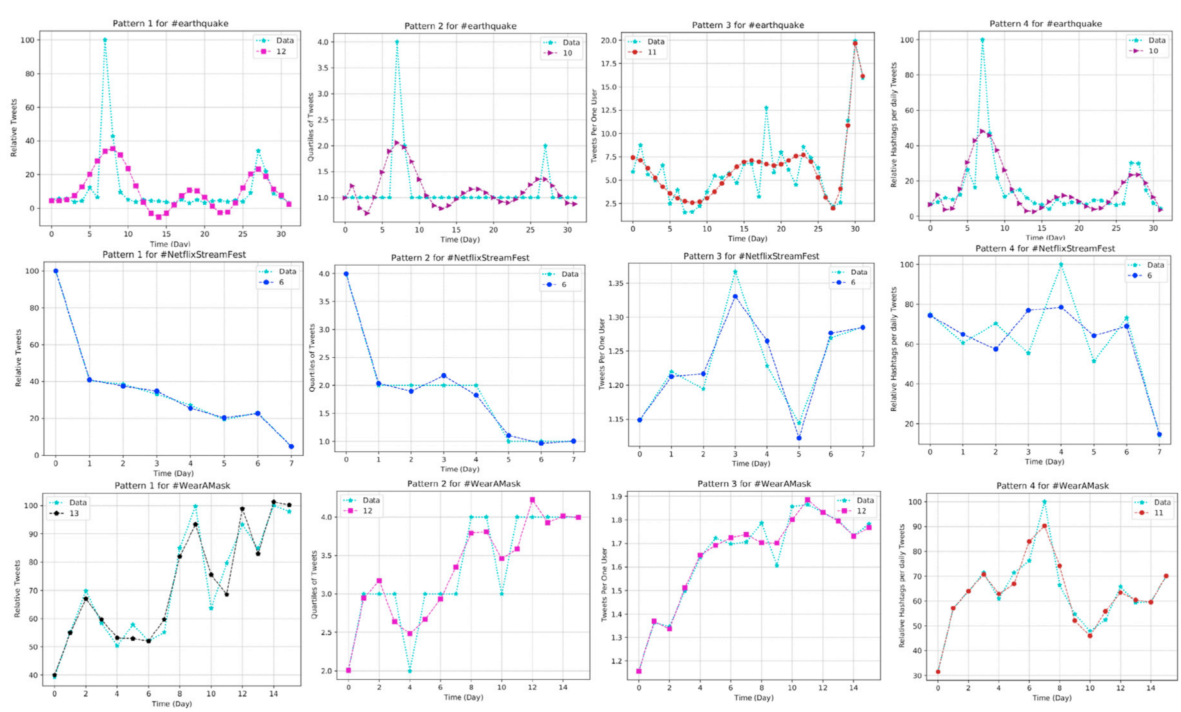

Fig. 5. Polynomial regression for *#Earthquake*, *#NetflixStreamFest* and *#WearAMask* where $\tau = Day$, Number mentioned in legends of each picture is PR degree, and it varies from 6 to 12. Each row consists on one hashtag and four patterns.

Table 6
Polynomial regression accuracy for all 4 patterns (P1,P2,P3,P4) with and degree against each hashtag while $\tau = Day$.

Hashtag	Days	P1, D	P2, D	P3, D	P4, D
#AFLSaintsDees	8	87.42, 06	88.50, 06	98.09, 06	81.30, 05
#alisadpara	32	82.72, 10	84.50, 10	89.64, 11	87.63, 10
#BilalSaeed	11	99.99, 15	98.25, 09	99.36, 09	99.08, 07
#ChristineHolgate	9	99.14, 07	99.75, 07	99.54, 07	96.32, 07
#Coronaviruspakistan	13	97.87, 11	97.25, 11	99.66, 11	100.0, 12
#earthquake	32	86.63, 10	88.75, 10	97.57, 11	88.10, 10
#Hazaragenocide	16	99.40, 11	97.00, 12	95.77, 11	97.24, 12
#justiceforbushrarajpar	13	99.93, 11	98.00, 11	95.19, 13	100.0, 12
#JusticeforUsamaNadeemSatti	16	99.14, 11	97.50, 12	91.98, 11	86.08, 11
#NetflixStreamFest	8	98.99, 06	97.50, 06	98.54, 06	87.29, 06
#PAKvsSA	8	91.10, 07	93.25, 07	95.13, 07	95.66, 06
#WearAMask	16	95.68, 12	93.25, 12	96.00, 12	95.88, 11
#WorldHealthDay2021	13	99.96, 11	100.0, 11	99.00, 13	99.98, 11

use Pickled Graphs (.gpickle) format for further processing.

Table 5 illustrates the general descriptions of dataset, while a comprehensive description of dataset is presented in Table 9. Table 9 presents the hashtags list, topics, coverage area and the number of nodes for each node type. Each hashtag has a graph, and all graphs in the formats mentioned above are available at MendeleyData.[7] We also use the same Table 9 for *Scalar Temporal Properties*. However, to avoid repetition, the table is included in section 4.4.

4.2. Experimental setup

To validate our framework's effectiveness, we create a hundred hashtags dataset, which stats are mentioned in Table 5. All of the experiments are performed on CPU AMD Ryzen 5 3600 6-Core Processor, and no special equipment is required to regenerate the results other than Python. Python codes are uploaded to download tweets, create graphs from tweets, create time-series, and mine temporal patterns at MendeleyData[10].

4.3. Series temporal patterns

The timeseries is the number of occurrences of any event per time. This paper's considered events of *Number of Tweet Nodes*, *Quadrants of Number of Tweet Nodes*, *Tweet and User Nodes Ratio*, and *Hashtags Mention*, and the time is τ. In addition, these events occurrences variate as time τ passes. The time-variant properties are called temporal patterns. Modeling these patterns leads to predicting the future occurrences of events.

Fig. 4 illustrates the *timeseries* of one hashtag *#JusticeForUsamaNadeemSatti*, while y-axis is the *Number of Tweet Nodes* and x-axis are three timespans (τ) days, hours, and minutes. For each timespan τ value, we use different model *Polynomial Regression (PR)* for days, *Auto-Regressive Integrated Moving Average Model (ARIMA)* for hours and *Long Short-Term Memory Network (LSTM)* for minutes to model the four patterns. In the following subsections, we mention them as:

- Pattern-1: Percentage of the number of tweet nodes.
- Pattern-2: Quadrants of the number of tweet nodes.
- Pattern-3: Tweet and user nodes ratio.
- Pattern-4: H_s mentions.

4.3.1. Timespan: Day

Fig. 5 illustrates the four timeseries and the mined patterns for three hashtags *#Earthquake*, *#NetflixStreamFest*, and *#WearAMask*, while the remaining hashtags are presented in Table 6, where the timespan is *Day*. Polynomial regression is used to mine patterns and the dotted line represents data and dashed lines are patterns. The figure helps us to compare two scenarios: *pattern-to-pattern* comparison and *hashtag-to-hashtag* comparison. The *pattern-to-pattern* comparison is made when the hashtags are the same and have an equal length of timespan. For example, the number of days for *#Earthquake*, *#NetflixStreamFest*, and *#WearAMask* are 32, 8, and 16, respectively, for all four patterns. In addition, the *hashtag-to-hashtag* comparison is between hashtags, and the length of timespan and the number of occurrences of the events are different from each other.

In both scenarios, the polynomial regression degree varies, which is illustrated in Fig. 5; at the upper corner of each figure, the written number is the degree of polynomial regression. Regardless of the same timespan length and model, the degree of PR varies in each timeseries and hashtag due to different data points and patterns. The degree ranges from 6 to 14. We calculate Root Mean Squared Error(*RMSE*) for each pattern and hashtag, and present the accuracy in Table 6. The mean values of the accuracy for Pattern-1,Pattern-2,Pattern-3 and Pattern-4 are 95.22%, 94.88%, 96.11%, and 93.43%, respectively. This proves that polynomial regression is perfectly fit on timespan (τ) = day, and it can be used for predictions. Meanwhile, the polynomial regression's degree and *RMSE* variance express that all patterns and hashtags are unique.

We also model this data with ARIMA and LSTM, but PR results are better. For example, for pattern-4 on hashtags *#WearAMask* and *#AFLSaintsDees*, results of ARIMA are 85.97%, 66.41%, and LSTM are 86.21%, 48.29% as compared to PR's accuracy *95.88%*, *81.30%*, respectively.

4.3.2. Timespan: Hour

ARIMA is a suitable model for $\tau = Hour$ among models used in this paper, while the length of hours for all hashtags is between 160 and 760 hours. The number of occurrences between each hour also varies, leading to unique patterns of each hashtag. The data is presented with the dotted line and the ARIMA's prediction with the dashed line in Fig. 6. For these experiments, ARIMA parameter values *p, d,* and *q* are ranging from 0 to 10, 0 to 3 and 0 to 3, respectively. Due to the limited space, it is not possible to present all hashtags in the figures, so we just demonstrate one hashtag, *#WearAMask* with the four patterns in Fig. 6. Also, we have already illustrated the comparison between hashtags in Fig. 5, thus we don't do the *hashtag-to-hashtag* comparison to avoid repetition. The hashtag *#WearAMask* is suitable for further analysis on the grounds that its lifetime is close to the mean of all hashtags. Furthermore, it has more spikes and fluctuations in its lifetime compared to other hashtags. These properties make it closer to real-life scenarios.

All hashtags are presented in Table 7 with Root Mean Squared Error (*RMSE*) based accuracy. The mean values of the accuracy of Pattern-1, Pattern-2, Pattern-3, and Pattern-4 are 94.28%, 90.51%, 89.58%, and 90.06%, respectively. The high accuracy of the results establishes the significance of ARIMA for this lifespan. We also test some sample hashtags' timeseries with polynomial regression and LSTM. For pattern-4 on *#WearAMask*, the accuracy of Polynomial Regression and LSTM is 86.04% and 86.82%, while the ARIMA is *87.42%*. For another hashtag *#AFLSaintsDees*, the accuracy is 93.37%, 85.56%, and *93.96%*, respectively. Between ARIMA and LSTM, there is a marginal accuracy difference with a significant processing time difference. In our experiments,

[7] https://data.mendeley.com/datasets/yx65w3tmc5/draft?a=1157b5e2-c65e-4c4b-a067-ab4aea63f149.

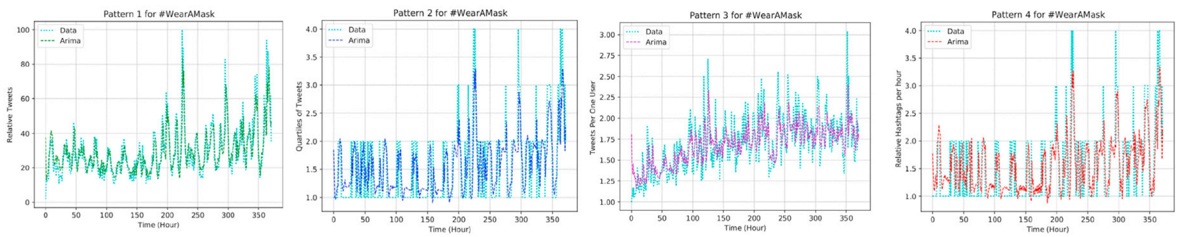

Fig. 6. ARIMA for #WearAMask (Four Patterns) where $\tau = Hour$.

Table 7
ARIMA for all 4 patterns (P1,P2,P3,P4) with accuracy against each hashtag while $\tau = Hour$.

Hashtag	Days	P1	P2	P3	P4
#AFLSaintsDees	167	93.96	91.38	90.00	86.39
#alisadpara	753	96.83	91.76	89.14	93.06
#BilalSaeed	242	96.35	92.06	89.04	89.50
#ChristineHolgate	195	93.26	89.74	88.46	87.78
#Coronaviruspakistan	282	95.15	87.94	85.44	92.06
#earthquake	758	96.33	92.00	90.23	96.26
#Hazaragenocide	369	96.38	91.50	85.25	94.02
#justiceforbushrarajpar	281	95.65	92.25	92.92	90.58
#JusticeforUsamaNadeemSatti	365	96.50	93.25	94.48	93.30
#NetflixStreamFest	177	89.15	89.00	88.20	78.28
#PAKvsSA	187	91.16	88.25	88.25	88.77
#WearAMask	372	91.00	87.75	91.78	87.10
#WorldHealthDay2021	287	93.93	89.75	91.36	93.62

Table 8
LSTM for all 4 patterns (P1,P2,P3,P4) with accuracy against each Hashtag while $\tau = Minute$.

Hashtag	Days	P1	P2	P3	P4
#AFLSaintsDees	07.9	99.21	96.25	86.67	98.22
#alisadpara	35.8	99.72	93.50	92.40	99.57
#BilalSaeed	11.6	99.30	98.50	91.00	99.57
#ChristineHolgate	09.3	98.52	90.00	78.47	98.59
#Coronaviruspakistan	13.3	98.32	92.75	83.63	97.90
#earthquake	36.0	99.67	92.00	71.48	97.53
#Hazaragenocide	17.7	99.23	95.50	96.07	99.19
#justiceforbushrarajpar	13.5	99.90	99.75	92.72	99.89
#JusticeforUsamaNadeemSatti	17.5	99.86	99.00	82.75	98.74
#NetflixStreamFest	08.5	94.95	92.75	80.33	93.78
#PAKvsSA	08.9	97.83	88.25	98.35	94.68
#WearAMask	17.8	90.03	89.25	90.50	92.81
#WorldHealthDay2021	13.7	99.70	96.00	89.44	99.53

ARIMA takes more processing time than LSTM, and sometimes the time ratio is 12 to 1 for ARIMA and LSTM. So we leave it on users of this model to choose the suitable option, considering the trade-off between accuracy and processing time.

4.3.3. Timespan: Minute

For timespan $\tau = Minute$, LSTM is suitable as it is a deep learning model and performs well on big data. Fig. 7 illustrates all four patterns for *#WearAMask*, where dotted lines are data and dashed lines are LSTM's predictions. The length of the timespan is long as one day has 1440 minutes, so we just present one hashtag's patterns in figure. All hashtags' *RMSE* based accuracy is presented in Table 8, where it can be observed that the accuracy is significantly great for LSTM. The mean values of accuracy for Pattern-1, Pattern-2, Pattern-3 and Pattern-4 are 98.17%, 94.12%, 87.22%, and 97.69%, respectively, which proves that LSTM is the best model for this timespan.

We also compare sample hashtag results with ARIMA and Polynomial Regression, and they both fail to outperform LSTM. For *#WearAMask*, the accuracy of pattern-4 Polynomial Regression and ARIMA are 92.86% and 87.42%, while that of LSTM is *92.81%*. For *#AFLSaintsDees*, the values are 97.80%, 97.78% and *99.21%*. In this timespan, although some differences are marginal, it is to be remembered that ARIMA takes too much time compared to LSTM (12:1). Polynomial regression is not very suitable if timespan's length is in thousands. It is faster than the other two techniques, but if its degree increases, so does its processing time.

The above three sections show each model's results, and Fig. 8 illustrates the concise results based on hashtags. Some hashtags perform well and some do not, which may depend on many factors, including topic, coverage area, trend time are some of many. The lowest result among hashtags is *#NetflixStreamFest*, with an accuracy rate of 90.73%, and the highest is #BilalSaeed with 96% accuracy. We also find pattern based average results for each pattern. Pattern 1 has the highest accuracy with an average of 95.89% among all patterns while patterns 4,2,3 are 93.73%, 93.17% and 90.97%, respectively. All results above 90% support our stance that these patterns can be used for predictions. Meanwhile, we also believe more timeseries and patterns can be mined from these graphs.

4.4. Scalar temporal properties of graph

There are some properties of graph and nodes which are temporal, but they are not based on timeseries. Examples are lifetime, active days, and peak hour of a hashtag node. These property values differ from one hashtag graph to another. In Table 9, we present all the hashtags used in this paper and their scalar temporal properties. Based on our observations, *Topic* and *Coverage area* affect the hashtag lifetime, the number of tweets and tweet-user ratio. These properties can be grouped into topic-based or coverage-based, and the grouping can lead to topic-based temporal patterns. Peak hour is also an interesting property as some hashtags peak in few hours, yet some take days to become trending. We can observe them with scalar temporal properties from Table 9.

- Action-oriented hashtags have peak hours in early days while sports hashtags takes time to be noticed.

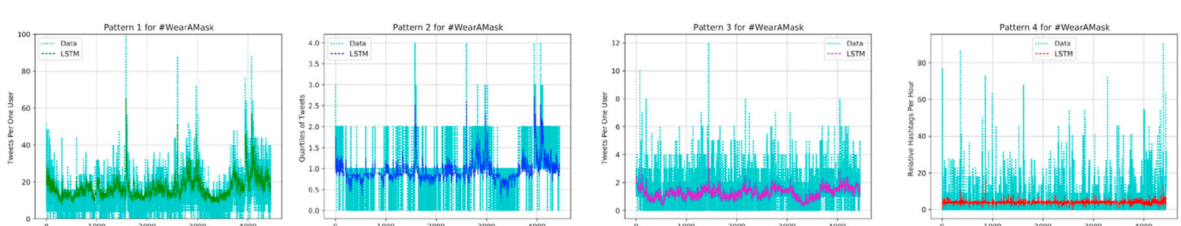

Fig. 7. LSTM for all 4 patterns for hashtag #WearAMask.

Table 9
Detailed Dataset Description and Temporal Scalars Values of all Graphs (All datasets including these and remaining are publicly available as Section 4.1

Sr#	Hashtag	Topic	Coverage	P_h	NH_{time}	LE_{time}	A_d	Tweet Nodes	User Nodes	Hashtag Nodes
1	BoycottNetflix	Action	India	8	11/25/2020 0:16	12/7/2020 7:07	13	1132	962	432
2	DelayMDCAT2020	Action	Pakistan	11	11/25/2020 1:23	11/25/2020 15:12	1	35007	3248	577
3	NetflixStreamFest	Entertainment	India	6	12/7/2020 0:06	12/14/2020 8:48	8	1644	1364	2055
4	justiceforusamanadeemsatti	Action	Pakistan	8	1/2/2021 7:35	1/17/2021 11:51	16	16148	8411	1200
5	hazaragenocide	Criticism	Pakistan	17	1/3/2021 0:31	1/18/2021 8:48	16	25472	12495	2200
6	BilalSaeed	Entertainment	Pakistan	48	2/2/2021 17:25	2/12/2021 18:39	11	5206	3741	573
7	OurBeautifulPakistan	Event	Pakistan	1	2/3/2021 19:00	2/8/2021 15:06	5	1426	721	224
8	earthquake	Event	Worldwide	182	2/5/2021 4:18	3/8/2021 17:14	32	152675	65322	11915
9	Alisadpara	Event	Pakistan	320	2/5/2021 5:07	3/8/2021 13:05	32	76965	37473	6484
10	PAKvsSA	Sports	Pakistan	60	2/5/2021 22:38	2/13/2021 16:37	8	31918	13143	2863
11	justiceforbushrarajpar	Action	Pakistan	27	2/12/2021 14:27	2/24/2021 6:54	12	9196	3876	512
12	AFLSaintsDees	Sports	Australia	124	3/22/2021 7:47	3/29/2021 6:03	7	1452	709	202
13	PakistanDayParade	Event	Pakistan	2	3/25/2021 10:12	3/25/2021 15:34	1	13627	6858	871
14	WearAMask	Action	Worldwide	225	4/4/2021 6:51	4/19/2021 17:45	16	52349	31620	13659
15	WorldHealthDay2021	Event	Worldwide	2	4/7/2021 8:08	4/19/2021 6:08	12	22684	18396	11510
16	Coronaviruspakistan	Health	Pakistan	78	4/8/2021 9:18	4/20/2021 2:41	12	3710	1912	1122
17	ChristineHolgate	Politics	Australia	35	4/12/2021 0:10	4/20/2021 2:37	9	15818	5391	1249
18	CovidIndia	Health	India	45	4/24/2021 10:32	4/26/2021 18:44	3	77137	53085	11029
19	londonprotest	Criticism	UK	2	5/29/2021 20:22	6/4/2021 13:44	6	15286	8923	1677
20	ImWithTheBand	Entertainment	UK	113	5/31/2021 6:06	6/7/2021 22:37	8	5338	1386	303
21	UFCVegas28	Sports	USA	131	5/31/2021 14:38	6/8/2021 11:48	8	35789	11543	1773
22	AFLDeesLions	Sports	Australia	83	6/1/2021 0:21	6/8/2021 7:17	8	3473	1341	361
23	SaveOurOcean	Action	Worldwide	174	6/1/2021 2:19	6/9/2021 4:33	9	3560	3073	1038
24	ClintEastwood	Politics	Australia	4	6/4/2021 7:40	6/4/2021 13:53	1	24	24	70
25	YachtShouldWeNameIt	Politics	UK	1	6/4/2021 21:21	6/6/2021 21:54	3	2665	1970	162
26	WorldEnvironmentDay	Event	Worldwide	6	6/7/2021 2:11	6/9/2021 5:22	3	18365	14345	5012
27	SplicerFashionShow	Entertainment	Worldwide	2	6/7/2021 21:37	6/8/2021 4:37	1	2620	2118	94
28	WorldOceansDay	Event	Worldwide	2	6/8/2021 20:13	6/11/2021 3:27	3	31282	25844	5556
29	SolarEclipse	Event	Worldwide	4	6/11/2021 13:22	6/15/2021 13:27	5	2189	1892	1485
30	EngvsNZ	Sports	UK	1	6/13/2021 14:19	6/15/2021 23:57	3	307	241	274
31	CleanAirDay	Event	Worldwide	2	6/17/2021 12:16	6/20/2021 8:09	3	4978	3468	1371
32	PSL6final	Sports	Pakistan	86	6/21/2021 7:24	6/29/2021 12:12	9	19783	9987	1839
33	WorldRefugeeDay	Event	Worldwide	2	6/21/2021 8:18	6/23/2021 12:07	3	17434	10864	2190
34	ENGvsGER	Sports	UK	166	6/22/2021 20:59	6/30/2021 4:38	8	25104	21264	2587
35	CancelAllBoardExams	Action	Pakistan	6	6/29/2021 22:26	6/30/2021 4:23	1	2045	670	198
36	WeStandWithSaeedGhani	Politics	Pakistan	50	7/2/2021 17:45	7/7/2021 3:09	5	21578	2680	359
37	CopaAmerica2021	Sports	Worldwide	26	7/3/2021 2:37	7/6/2021 2:24	3	37188	24380	7705
38	BanHumStyleAward	Action	Pakistan	20	7/4/2021 19:15	7/9/2021 1:42	5	24700	5581	1015
39	CancelAllExams	Action	Pakistan	4	7/5/2021 6:23	7/7/2021 3:54	2	69097	11533	1461
40	DilipKumar	Entertainment	India	5	7/7/2021 13:55	7/15/2021 14:30	9	39013	24687	4131
41	AUSvFRA	Sports	Australia	141	7/7/2021 15:02	7/15/2021 13:17	8	4724	1986	1262
42	CrackingCOVID	Health	Australia	133	7/7/2021 23:04	7/15/2021 13:04	8	1781	1092	177
43	PHXvsMIL	Sports	USA	165	7/8/2021 7:55	7/16/2021 3:05	8	10922	4968	794
44	GoBolts	Sports	USA	105	7/8/2021 8:48	7/13/2021 8:54	6	18054	8445	1502
45	ArrestTrumpNow	Politics	USA	2	7/12/2021 1:33	7/16/2021 3:20	5	147392	55328	3700
46	Euro2020Final	Sports	Worldwide	5	7/12/2021 3:01	7/13/2021 9:02	2	136548	96221	9498
47	covidnsw	Health	Australia	78	7/18/2021 20:15	7/27/2021 4:08	9	10698	5251	2893
48	SydneyOutbreak	Health	Australia	75	7/18/2021 23:33	7/27/2021 4:42	9	2838	1545	778
49	AJKElections	Politics	Pakistan	152	7/19/2021 5:18	7/27/2021 7:17	9	4851	3338	508
50	ADA31	Event	USA	174	7/19/2021 12:32	7/27/2021 8:56	8	7152	4965	1625
51	raisetheage	Action	Australia	145	7/20/2021 22:46	7/29/2021 3:35	9	2712	1860	360
52	JusticeForNoor	Action	Pakistan	7	7/21/2021 5:48	7/23/2021 14:41	3	25157	13551	1871
53	JusticeforAndaleeb	Action	Pakistan	32	7/23/2021 12:11	7/27/2021 4:07	4	4398	2799	693
54	YorkshireDay	Event	UK	146	7/26/2021 7:12	8/3/2021 4:12	8	17134	12162	3336
55	Tokyo2020	Event	Worldwide	2	7/27/2021 7:46	7/27/2021 8:53	1	55975	39478	5405
56	GBRvAUS	Sports	UK	63	7/27/2021 21:41	8/1/2021 21:11	5	7561	3145	818

(*continued on next page*)

Table 9 (*continued*)

Sr#	Hashtag	Topic	Coverage	P_h	NH_{time}	LE_{time}	A_d	Tweet Nodes	User Nodes	Hashtag Nodes
57	USAvMEX	Sports	USA	79	7/29/2021 21:31	8/3/2021 4:01	5	31526	19925	1337
58	lockdown6	Health	Australia	153	7/29/2021 22:51	8/6/2021 11:32	8	11342	7242	3816
59	BMXRacing	Sports	Australia	2	7/30/2021 3:59	8/2/2021 12:37	4	12553	9323	1431
60	womenhockeyindia	Sports	India	163	7/30/2021 9:39	8/7/2021 8:01	8	22102	16301	2254
61	TheSuicideSquad	Entertainment	USA	83	7/30/2021 16:33	8/7/2021 8:01	8	171742	74858	8377
62	insiders	Politics	Australia	160	8/1/2021 8:35	8/9/2021 10:15	9	21815	6303	2065
63	NRLTitansCowboys	Sports	Australia	121	8/3/2021 6:00	8/9/2021 7:17	7	583	227	104
64	CyclingTrack	Sports	UK	27	8/3/2021 7:55	8/6/2021 9:52	4	44008	21513	1811
65	AFLHawksPies	Sports	Australia	101	8/4/2021 2:30	8/9/2021 7:02	6	1234	458	153
66	14augustazadiday	Event	Pakistan	116	8/8/2021 20:29	8/17/2021 3:21	9	30375	12644	2616
67	IndiaAt75	Event	India	2	8/15/2021 15:31	8/18/2021 8:29	3	31709	22563	4498
68	RakshaBandhan	Event	India	156	8/15/2021 17:08	8/23/2021 7:42	8	138667	91454	30137
69	SpiderManNoWayHomeTrailer	Entertainment	Worldwide	44	8/30/2021 21:20	9/8/2021 3:34	9	2701	2321	972
70	NationalConventionOfFarmers	Criticism	India	8	8/30/2021 21:24	9/7/2021 15:57	8	133	113	16
71	NoFarmers$_{N}$*oFuture*	Criticism	India	162	8/30/2021 21:30	9/8/2021 3:02	9	245	197	136
72	NoMoreLockdowns	Criticism	UK	169	8/30/2021 21:31	9/8/2021 4:17	9	4774	3444	1659
73	UFCFightNight	Sports	Worldwide	120	8/30/2021 21:35	9/8/2021 4:10	9	1826	1093	613
74	NarayanRane	Politics	India	153	8/30/2021 23:30	9/8/2021 4:14	9	486	399	435
75	NationalInsurance	Criticism	UK	166	8/31/2021 15:58	9/9/2021 3:27	9	8492	6052	1762
76	WheelchairTennis	Sports	Australia	94	8/31/2021 15:58	9/8/2021 23:15	9	8938	5369	481
77	PaulWoodley	Entertainment	Worldwide	4	8/31/2021 15:59	9/9/2021 1:18	9	783	595	522
78	NWA73	Entertainment	USA	57	8/31/2021 16:04	9/9/2021 0:50	9	1193	676	243
79	SASWhoDaresWins	Entertainment	UK	125	8/31/2021 16:05	9/8/2021 23:35	9	853	523	150
80	IStandWithDan	Politics	Australia	156	8/31/2021 16:24	9/9/2021 2:34	9	5811	2731	760
81	PMforNSW	Politics	Australia	150	8/31/2021 22:15	9/9/2021 3:00	9	2214	1193	365
82	FSUvsND	Sports	USA	125	8/31/2021 23:58	9/8/2021 19:57	8	7436	4502	1002
83	TaxTheRich	Criticism	UK	159	9/1/2021 1:41	9/9/2021 5:00	9	25566	19742	1942
84	ThreatenedSpeciesDay	Event	Australia	142	9/1/2021 5:51	9/9/2021 2:03	8	3729	2364	346
85	AUSvCHN	Sports	Australia	34	9/1/2021 9:00	9/7/2021 17:05	7	1204	571	112
86	SocialCare	Criticism	UK	138	9/1/2021 19:38	9/10/2021 4:08	9	25090	15109	4286
87	WomensSafetySummit	Event	Australia	100	9/1/2021 21:43	9/10/2021 4:34	9	18397	6534	1580
88	SackHunt	Politics	Australia	133	9/2/2021 20:46	9/9/2021 14:15	7	8439	3323	809
89	ModiStopMisleadingFarmers	Criticism	India	71	9/4/2021 6:37	9/8/2021 0:56	4	87	61	71
90	AEWRampage	Entertainment	USA	7	9/4/2021 7:17	9/8/2021 10:46	5	11609	6366	1636
91	IStandWithBiden	Politics	USA	61	9/5/2021 2:38	9/8/2021 5:44	4	2626	1980	295
92	HappyTeachersDay	Event	India	3	9/5/2021 13:44	9/9/2021 12:49	4	7332	6211	2312
93	InternationalDogDay	Event	Worldwide	26	9/6/2021 16:44	9/9/2021 2:25	3	252	141	84
94	SoccerAid	Sports	UK	10	9/6/2021 22:01	9/9/2021 1:01	3	170	139	113
95	UseGenericMedicine	Health	India	1	9/7/2021 4:03	9/9/2021 12:47	3	7754	2861	239
96	JusticeForRabiya	Action	India	35	9/7/2021 5:03	9/9/2021 8:17	3	4537	2729	371
97	TexasWarOnWomen	Criticism	USA	9	9/7/2021 13:05	9/9/2021 7:32	2	16234	13166	1741
98	JusticeForKarnalFarmers	Criticism	India	2	9/8/2021 2:37	9/8/2021 5:57	1	20579	3873	369
99	TrillerFightClub	Sports	UK	2	9/12/2021 5:24	9/15/2021 3:22	3	1968	1465	491
100	EmmaRaducanu	Sports	Worldwide	2	9/13/2021 11:12	9/14/2021 3:57	1	1205	1091	1072
	Total				294 Days			2,103,600	1,112,420	230,142

- Entertainment based hashtags have more active days than action based hashtags.
- Entertainment based hashtags have more user-involvements than action oriented hashtags.
- A trend active in the USA and India have more tweets than one active in Pakistan.

Since there are six topics and eighteen graphs, and above all the points based on observations of dataset, that are not enough to prove or disprove these points. Therefore, more research is needed to prove them and our proposed framework can be used for the analysis.

5. Conclusion

User-generated content has transformed from *conventional web* into *social web*, which leads to content diversity. Available research mostly

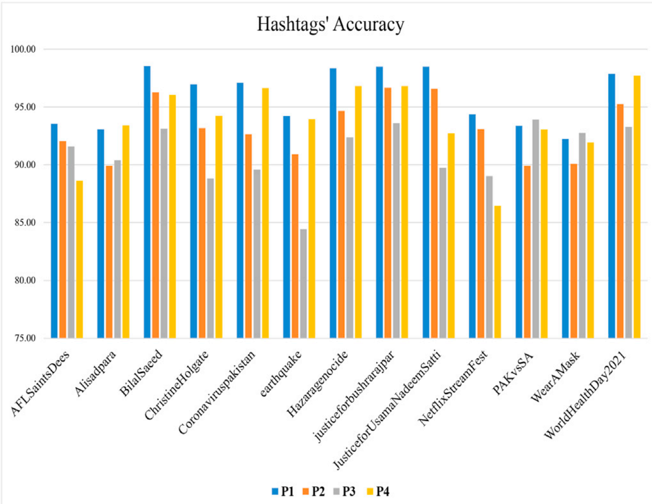

Fig. 8. Primary hashtags (H_p) results against each pattern.

focuses on textual content, which undermines the metadata of the content. Metadata is as important as content, as it keeps information useful to analyze throughout the behavior of the content and make future predictions. This paper presents a framework that uses user-generated content (Twitter) metadata and finds patterns from it. *First*, it takes Twitter raw data and converts it into a dynamic graph. As a result, a hundred dynamic graphs are made publicly available. *Second*, it extracts four types of timeseries from the graph by considering three timespans: *Day, Hour*, and *Minute*. *Third*, it applies three different models to the subsequent timespan-based timeseries and finds temporal patterns. *Finally*, the temporal patterns lead to the lifetime predictions of a hashtag, user involvement ratio predictions, and hashtag influence spreading. At last, we present our results, usability analysis, and finding in results, analysis, and finding sections.

Future directions

For future directions, according to our analysis, hashtags belonging to the same topics may have similar patterns. As political and entertainment hashtags tend to live longer than action-oriented and criticism-based hashtags. These are our observations and analysis. However, more research is required to prove or disprove them. Additionally, there are some non-temporal properties of the UTH graph, varying from one graph to another. For example, tweet-to-user ratio, hashtag-to-tweet ratio, user-to-hashtag ratio, number of hashtags, number of tweets, number of users, and last but not least influential user node of the graph. Expanding the graph via linking the same type of nodes will lead to more complex and information-rich graphs. As a result, more features can be considered to complete the task. Exploring these graphs and properties mentioned above can lead us to a better understanding of user-generated content. We suggest using the same machine learning models used in this paper, for these temporal patterns, but if the graph is updated, then the model updates may also be required.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This Research is supported by National Natural Science Foundation of China (grant no. 61573328). The authors gratefully acknowledge the joint PhD fellowship provided by the Chinese Academy of Sciences (CAS) and The World Academic of Sciences (TWAS) as the first author is sponsored by CAS- TWAS President's Fellowship for Ph.D at School of Computer Science and Technology, University of Science and Technology China. He would like to thanks his supervisor Dr. Li and CAS-TWAS for the opportunity. On the behalf of all of authors, we are grateful to reviewers' excellent suggestions to improve the quality of this work. At last, we appreciate *Twitter* for providing free access to data and *diagrams.net* for free access to their wonderful graphics drawing software.

References

[1] L.A. Kothamasu, E. Kannan, Sentiment analysis on twitter data based on spider monkey optimization and deep learning for future prediction of the brands, Concurrency Comput. Pract. Ex. 34 (21) (2022), e7104.
[2] V. Ramanathan, T. Meyyappan, Twitter text mining for sentiment analysis on people's feedback about Oman tourism, in: 2019 4th MEC International Conference on Big Data and Smart City, ICBDSC), 2019, pp. 1–5.
[3] R. Buettner, K. Buettner, A systematic literature review of twitter research from a socio-political revolution perspective, in: 2016 49th Hawaii International Conference on System Sciences, HICSS), 2016, pp. 2206–2215.
[4] M. Shafiq, Z. Tian, A.K. Bashir, A. Jolfaei, X. Yu, Data mining and machine learning methods for sustainable smart cities traffic classification: a survey, Sustain. Cities Soc. 60 (2020), 102177.
[5] R. Sharma, S. Uniyal, V. Gera, Performing interest mining on tweets of twitter users for recommending other users with similar interests, in: B. Pati, C.R. Panigrahi, S. Misra, A.K. Pujari, S. Bakshi (Eds.), Progress in Advanced Computing and Intelligent Engineering, Springer Singapore, Singapore, 2019, pp. 593–603.
[6] Y. Djenouri, A. Belhadi, G. Srivastava, J.C.-W. Lin, Toward a cognitive-inspired hashtag recommendation for twitter data analysis, IEEE Trans. Comput. Soc. Sys. (2022) 1–10.
[7] J.R. Saura, A. Reyes-Menendez, P. Palos-Sanchez, Are black friday deals worth it? mining twitter users' sentiment and behavior response, J. Open Innov.: Technol. Market. Complex. 5 (3) (2019) 58.
[8] E. Georgiadou, S. Angelopoulos, H. Drake, Big data analytics and international negotiations: sentiment analysis of brexit negotiating outcomes, Int. J. Inf. Manag. 51 (2020), 102048.
[9] A. Alamoodi, B. Zaidan, A. Zaidan, O. Albahri, K. Mohammed, R. Malik, E. Almahdi, M. Chyad, Z. Tareq, A. Albahri, H. Hameed, M. Alaa, Sentiment analysis and its applications in fighting covid-19 and infectious diseases: a systematic review, Expert Syst. Appl. 167 (2021), 114155.
[10] S.-U. Hassan, N.R. Aljohani, N. Idrees, R. Sarwar, R. Nawaz, E. Martínez-Cámara, S. Ventura, F. Herrera, Predicting literature's early impact with sentiment analysis in twitter, Knowl. Base Syst. 192 (2020), 105383.
[11] Guest editorial: cognitive analytics of social media for industrial manufacturing, IEEE Trans. Ind. Inf. 17 (4) (2021) 2899–2901.
[12] M. Martínez-Rojas, M. del Carmen Pardo-Ferreira, J.C. Rubio-Romero, Twitter as a tool for the management and analysis of emergency situations: a systematic literature review, Int. J. Inf. Manag. 43 (2018) 196–208.
[13] A. Sharma, M.O. Shafiq, A comprehensive artificial intelligence based user intention assessment model from online reviews and social media, Appl. Artif. Intell. 36 (1) (2022), 2014193.
[14] W. Deitrick, B. Valyou, W. Jones, J. Timian, W. Hu, Enhancing sentiment analysis on twitter using community detection, Commun. Netw. (2013) 192–197, 05 (03).
[15] X. Wang, F. Wei, X. Liu, M. Zhou, M. Zhang, Topic sentiment analysis in twitter: a graph-based hashtag sentiment classification approach, in: Proceedings of the 20th ACM International Conference on Information and Knowledge Management, CIKM '11, Association for Computing Machinery, New York, NY, USA, 2011, pp. 1031–1040.
[16] N.F. Ibrahim, X. Wang, A text analytics approach for online retailing service improvement: evidence from twitter, Decis. Support Syst. 121 (2019) 37–50.
[17] A.K. Rathore, P.V. Ilavarasan, Pre- and post-launch emotions in new product development: insights from twitter analytics of three products, Int. J. Inf. Manag. 50 (2020) 111–127.
[18] K. Booten, Hashtag drift: tracing the evolving uses of political hashtags over time, in: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, CHI '16, Association for Computing Machinery, New York, NY, USA, 2016, pp. 2401–2405.
[19] Y. Lin, D. Margolin, B. Keegan, A. Baronchelli, D. Lazer, #bigbirds never die: understanding social dynamics of emergent hashtags, in: E. Kiciman, N.B. Ellison, B. Hogan, P. Resnick, I. Soboroff (Eds.), Proceedings of the Seventh International Conference on Weblogs and Social Media, ICWSM 2013, The AAAI Press, Cambridge, Massachusetts, USA, 2013. July 8-11, 2013.
[20] A.E. Sharma, Z. Mann, R. Cherian, J.B. Del Rosario, J. Yang, U. Sarkar, Recommendations from the twitter hashtag #doctorsaredickheads: qualitative analysis, J. Med. Internet Res. 22 (10) (2020), e17595.
[21] J. Yang, J. Leskovec, Patterns of temporal variation in online media, in: Proceedings of the Fourth ACM International Conference on Web Search and Data Mining, WSDM '11, Association for Computing Machinery, New York, NY, USA, 2011, pp. 177–186.
[22] B. Cai, G. Huang, S. Yang, Y. Xiang, C.-H. Chi, Clustering hashtags using temporal patterns, in: Z. Huang, W. Beek, H. Wang, R. Zhou, Y. Zhang (Eds.), Web

Information Systems Engineering – WISE 2020, Springer International Publishing, Cham, 2020, pp. 183–195.

[23] G. Stilo, P. Velardi, Hashtag sense clustering based on temporal similarity, Comput. Ling. 43 (1) (2017) 181–200.

[24] T.-Y. Kim, J. Kim, J. Lee, J.-H. Lee, A Tweet Summarization Method Based on a Keyword Graph, ICUIMC '14, Association for Computing Machinery, New York, NY, USA, 2014.

[25] R. Paudel, P. Kandel, W. Eberle, Detecting spam tweets in trending topics using graph-based approach, in: K. Arai, R. Bhatia, S. Kapoor (Eds.), Proceedings of the Future Technologies Conference (FTC) 2019, Springer International Publishing, Cham, 2020, pp. 526–546.

[26] C. Zheng, Q. Zhang, G. Long, C. Zhang, S.D. Young, W. Wang, Measuring time-sensitive and topic-specific influence in social networks with lstm and self-attention, IEEE Access 8 (2020) 82481–82492.

[27] R.E. Bacha, T. Thi Zin, Ranking of influential users based on user-tweet bipartite graph, in: 2018 IEEE International Conference on Service Operations and Logistics, and Informatics, SOLI, 2018, pp. 97–101.

[28] A. Çelebi, A. Özgür, Segmenting hashtags and analyzing their grammatical structure, J. Assoc. Inf. Sci. Technol. 69 (5) (2018) 675–686.

[29] D. Davidov, O. Tsur, A. Rappoport, Enhanced sentiment learning using twitter hashtags and smileys, in: Proceedings of the 23rd International Conference on Computational Linguistics: Posters, COLING '10, Association for Computational Linguistics, USA, 2010, pp. 241–249.

[30] D. Davidov, O. Tsur, A. Rappoport, Semi-supervised recognition of sarcastic sentences in twitter and amazon, in: Proceedings of the Fourteenth Conference on Computational Natural Language Learning, CoNLL '10, Association for Computational Linguistics, USA, 2010, pp. 107–116.

[31] Y. Wang, J. Liu, Y. Huang, X. Feng, Using hashtag graph-based topic model to connect semantically-related words without co-occurrence in microblogs, IEEE Trans. Knowl. Data Eng. 28 (7) (2016) 1919–1933.

[32] A. Belhadi, Y. Djenouri, J.C. Lin, A. Cano, A data-driven approach for twitter hashtag recommendation, IEEE Access 8 (2020) 79182–79191.

[33] F. Godin, V. Slavkovikj, W. De Neve, B. Schrauwen, R. Van de Walle, Using topic models for twitter hashtag recommendation, in: Proceedings of the 22nd International Conference on World Wide Web, WWW '13 Companion, Association for Computing Machinery, New York, NY, USA, 2013, pp. 593–596.

[34] O. Tsur, A. Rappoport, What's in a hashtag? content based prediction of the spread of ideas in microblogging communities, in: Proceedings of the Fifth ACM International Conference on Web Search and Data Mining, WSDM '12, Association for Computing Machinery, New York, NY, USA, 2012, pp. 643–652.

[35] Z. Ma, A. Sun, G. Cong, Will this hashtag be popular tomorrow?, in: Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '12 Association for Computing Machinery, New York, NY, USA, 2012, pp. 1173–1174.

[36] B. Samanta, A. De, N. Ganguly, Strm: a sister tweet reinforcement process for modeling hashtag popularity, in: IEEE INFOCOM 2017 - IEEE Conference on Computer Communications, 2017, pp. 1–9.

[37] D. Sipocz, J.D. Freeman, J. Elton, A toxic trend?": generational conflict and connectivity in twitter discourse under the BoomerRemover hashtag, Gerontol. 61 (2) (2020) 166–175.

[38] A.S. Bradshaw, # doctorsspeakup, Exploration of hashtag hijacking by anti-vaccine advocates and the influence of scientific counterpublics on twitter, Health Commun. (2022) 1–11, 0 (0).

[39] Y. Matsubara, Y. Sakurai, B.A. Prakash, L. Li, C. Faloutsos, Rise and fall patterns of information diffusion: model and implications, in: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '12, Association for Computing Machinery, New York, NY, USA, 2012, pp. 6–14.

[40] A. Ferraz Costa, Y. Yamaguchi, A. Juci Machado Traina, C. Traina, C. Faloutsos, Rsc: mining and modeling temporal activity in social media, in: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '15, Association for Computing Machinery, New York, NY, USA, 2015, pp. 269–278.

[41] E. Dritsas, I.E. Livieris, K. Giotopoulos, L. Theodorakopoulos, An Apache spark implementation for graph-based hashtag sentiment classification on twitter, in: Proceedings of the 22nd Pan-Hellenic Conference on Informatics, PCI '18, Association for Computing Machinery, New York, NY, USA, 2018, pp. 255–260.

[42] S. Dutta, S. Ghatak, M. Roy, S. Ghosh, A.K. Das, A graph based clustering technique for tweet summarization, in: 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), 2015, pp. 1–6.

[43] C. Li, J. Bai, L. Zhang, H. Tang, Y. Luo, Opinion community detection and opinion leader detection based on text information and network topology in cloud environment, Inf. Sci. 504 (2019) 61–83.

[44] Z. Majdabadi, B. Sabeti, P. Golazizian, S.A. Ashrafi Asli, O. Momenzadeh, R. Fahmi, Twitter trend extraction: a graph-based approach for tweet and hashtag ranking, utilizing no-hashtag tweets, in: Proceedings of the 12th Language Resources and Evaluation Conference, European Language Resources Association, Marseille, France, 2020, pp. 6213–6219.

[45] L. Praznik, G. Srivastava, C. Mendhe, V. Mago, Vertex-weighted measures for link prediction in hashtag graphs, in: Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM '19, Association for Computing Machinery, New York, NY, USA, 2019, pp. 1034–1041.

[46] M. Cha, H. Haddadi, F. Benevenuto, K.P. Gummadi, Measuring user influence in twitter: the million follower fallacy, in: ICWSM '10: Proceedings of International AAAI Conference on Weblogs and Social, 2010.

[47] Y. Yang, Z. Wang, J. Pei, E. Chen, Tracking influential individuals in dynamic networks, IEEE Trans. Knowl. Data Eng. 29 (11) (2017) 2615–2628.

[48] V.K. Yalavarthi, A. Khan, Steering top-k influencers in dynamic graphs via local updates, in: 2018 IEEE International Conference on Big Data, Big Data, 2018, pp. 576–583.

[49] A.A. Hagberg, D.A. Schult, P.J. Swart, Exploring network structure, dynamics, and function using networkx, in: G. Varoquaux, T. Vaught, J. Millman (Eds.), Proceedings of the 7th Python in Science Conference, 2008, pp. 11–15. Pasadena, CA USA.