Making Sense of Tweets Using Sentiment Analysis on Closely Related Topics

Sarvesh Bhatnagar (sarveshbhatnagar08@gmail.com)
Mukesh Patel School of Technology Management and Engineering - Shirpur Campus
https://orcid.org/0000-0003-2391-0341

Nitin Choubey
Mukesh Patel School of Technology Management and Engineering - Shirpur Campus

Research Article

Keywords: Community Detection, Sentiment Analysis, Social Network Analysis, Online Social Networks, Trend Analysis

Posted Date: April 12th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-385908/v1

Version of Record: A version of this preprint was published on May 3rd, 2021. See the published version at https://doi.org/10.1007/s13278-021-00752-0.

Noname manuscript No.
(will be inserted by the editor)

Making Sense of Tweets Using Sentiment Analysis on Closely Related Topics

Sarvesh Bhatnagar · Nitin Choubey

Received: date / Accepted: date

Abstract Microblogging has taken a considerable upturn in recent years, with the growth of microblogging websites like Twitter people have started to share more of their opinions about various pressing issues on such online social networks. A broader understanding of the domain in question is required to make an informed decision. With this motivation, our study focuses on finding overall sentiments of related topics with reference to a given topic. We propose an architecture that combines sentiment analysis and community detection to get an overall sentiment of related topics. We apply that model on the following topics: shopping, politics, covid19 and electric vehicles to understand emerging trends, issues and its possible marketing, business and political implications.

Keywords Community Detection · Sentiment Analysis · Social Network Analysis · Online Social Networks · Trend Analysis

1 Introduction

Online social networks(OSNs) have been burgeoning in recent years [2]. This rapid growth of social network, combined with easily accessible data and discussions on multitude of topics provides great research potential for customer analysis, product analysis, sector analysis and digital marketing. Different data science and Machine learning techniques such as Clustering, association rule mining, ensemble models, deep learning, sentiment analysis, etc are used in conjunction with digital marketing and product analysis [58, 3]. People use social networks to discuss wide ranging topics and share opinions on them [67].

S. Bhatnagar
SVKM's NMIMS, MPSTME, Shirpur
E-mail: sarveshbhatnagar08@gmail.com

N. Choubey
SVKM's NMIMS, MPSTME, Shirpur
E-mail: nitin.choubey@nmims.edu

Given the scale of information on OSNs, there arises a need to apply different data mining techniques to get actionable insights from them[15, 60].

OSNs such as Facebook, Twitter, Instagram, and so on encourages people to participate and collaborate, forming virtual online communities [40, 54]. This encouragement is in various forms such as likes, shares, retweets, use of hashtags, comments, mentions, etc. In these OSNs, authors write about their life, share opinions on various topics and discuss wide-ranging issues [67]. Also, the use of collaboration features, as discussed above facilitates studies like community detection by allowing formation of multidimensional networks based on friend/follower network[17], network based on hashtags[68, 42], sentiment based network[70], and so on. Multidimensional networks are networks that may have multiple connections between any pair of nodes [7] for this purpose, multidimensional analysis is required to gain valuable insights from them. Among different OSNs, Twitter is one of the most studied OSN for social network research [38]. One of the main advantages of platforms like Twitter for research is that, on these platforms, users are organized in networks, which makes it possible to investigate groups of people, or communities, united by common interests, rather than individual profiles or personalities which is enabled by extensive use of hashtags, mentions and retweets that forms a complex network [30]. Which in-turn is important for big data analysis and digital marketing [32, 58, 29].

To gauge profitability of a product or a business one need to consider two main things: 1. Attractiveness of a business or product, 2. Competitiveness level [12]. Finding attractiveness of a product is important as with time, trend changes. These changes in trends often demand changes to existing business models in order to sustain in the market and to alleviate the inevitable risks involved [19]. As an example, recent trend to use sustainable energy led to the growth of electrical vehicles shifting the focus from petrol or diesel based vehicles [28, 25]. A growth in trend is generally accompanied with positive opinion for that topic consequently text analysis techniques are often used to identify public opinion about a trend [26]. In many cases it is often important to get a broader understanding of the topic to understand key players and overall opinion for that sector. Broader understanding of a topic also allows us to better understand emerging trends and public opinion about them. For this purpose traditional methods tend to be more time consuming as it involves finding relevant topic and then applying sentiment analysis over it [11]. This takes time because topic modeling is a slow process and often involves qualitative human intervention [55]. Also, existing topic modeling methods does not allow changes to generality of the found topics[10, 11, 55, 59]. With this as our motivation, In this paper we propose a framework that can be used to get related trends and topics accompanied by their overall public sentiment along with a parameter that can be used to change generality of the found topics. Finding recent trends and topics is important for businesses, politicians and marketing agencies alike. We can use the results from our proposed model to answer questions like what is the overall sentiment for a given topic?, What are the emerging issues and public opinion about them? How is a product faring

compared to other products?, what is the general market trend? and who are the key players for a given trend?. For this purpose we apply our model over wide ranging topics like shopping, electric vehicles, covid19 and politics. We then compare our results with recent exploratory analysis in these topics.

In conclusion, the contributions of this paper include: 1. Overall Topic Sentiment Classifier model; 2. Evaluation of different trends based on our model findings. In this our goal being to propose a model that can be used to effectively find related topics along with their overall sentiment from a given topic on twitter platform. Furthermore, In our model, we provide a key metric that can be used to vary how general the resultant related topics are with respect to our given topic. We also note that user generated content are qualitative and, therefore, should be used for exploratory analysis [34,50].

1.1 Background

Rapid growth of OSNs and massive data flow through social networks have given rise to research on the analysis of social networks [2,65,37]. OSNs have also changed the dynamics of how consumers buy products and interact with one another [39]. This change of dynamics combined with modern data mining techniques have led to its use in digital marketing and targeted customer analysis[58]. In particular, sentiment analysis for opinion mining[18] and community detection for customer targeting, segmentation and topic modeling[32] are widely studied. We use sentiment analysis to understand how people orient themselves about a topic given a piece of text [71]. Particularly we aim to determine weather the given text is of positive connotation, negative connotation or neutral connotation [35]. It helps us understand public opinion given a text corpus of a given topic. As an example we can try to identify public opinion about covid-19 based on tweets about it[10].

Another task for understanding broader market dynamics is to retrieve closely related subtopics for which we can perform the above mentioned analysis [11]. To get closely related subtopics, we can use community detection over a topic based network [42]. We define community detection as a way in which we attempt to find a set of clusters such that it minimizes intraconnection between them and maximizes interconnection within the cluster in a given set [22]. Another possible approach is using Latent Dirichlet Allocation(LDA) over corpus of text to find embedded topics within them [59]. Although LDA is a good choice to detect themes discussed in a set of text corpus, it fails to incorporate intrinsic twitter feature "hashtag" that inherently is used for expressing the topic that particular tweet is about [38,16]. Furthermore, there is no means using which we can induce the generality metric to find related topics for the given tweets.

In real world, networks are often multidimensional. To get actionable insights from such networks, we require multidimensional analysis to distinguish among different kinds of interactions or equivalently look at interaction from different perspectives. Dimensions can either be explicit that directly reflect

interactions such as friend-follower network or it can be implicit that reflect interesting qualities of interactions that can be inferred from the available data for instance hashtag network [7]. In our work we focus on multidimensional network with two explicit dimensions 1. Hashtags and 2. Opinions about topics. There can be different interactions between two users. They can be connected to each other with same set of topics with similar opinion. They can be connected to each other with same set of topics with different opinion. They can also be connected to each other with partial set of topics with same or different opinion.

1.2 Organization

The rest of the paper is organized as follows. In Section 2, we discuss prior works on community detection and sentiment analysis. In Section 3, we take a look at community detection and sentiment analysis. The architecture for Overall topic sentiment classifier(OTSC) is described in Section 4, mention our results in Section 5 and Discuss it in Section 6. In Section 7, we conclude by stating our contributions, discuss managerial implications and practical/social implications for marketers. Finally we discuss Limitations and Future Research in Section 8.

2 Related Work

It is important to understand emerging trends and their public opinion for making informed business and political decision [6, 4, 51]. Interactions among people in OSN leads to formation of a multidimensional complex network. [7] paper lays foundations of multidimensional network and its analysis.

Among different OSNs, Twitter is one of the most studied OSN [30]. [48, 36, 38] analyzes twitter tweets using sentiment analysis. Furthermore, there are many works related to development and discussion about different sentiment analysis models [35, 8, 72, 71]. These models can be used to understand sentiment of a given text. To train a model for classification purposes, we need labelled data. For this purpose some automatic data collection methods have been researched, for instance [53] used emoji's to collect and label data while [16] uses hashtags for the same.

Other major topic of research for OSNs includes community detection[32]. Community detection is not a well defined topic and requires some degree of arbitrariness and/or common sense. Given that, [23] performed a thorough review of community detection algorithms. [32] studies various applications of community detection such as criminology, public health, politics, customer segmentation, smart advertising, targeted marketing, network summarization, social network analysis, recommendation systems, link prediction and community evolution prediction. Furthermore, in rare instances, community detection and sentiment analysis have been combined, for instance [17] focuses on using

sentiment analysis to enhance community detection. They use different twitter specific features to further enhance the detected community.

[59] proposes a three stage method for text mining using LDA for topic modeling followed by sentiment analysis which is followed by application of text mining techniques. Application of similar procedure can be found in [54, 60, 11, 10]. [55] used model proposed in [59] to analyze and understand business implication of #metoo in twitter, further highlighting key takeaways for businesses and advertisers.

3 Community Detection and Sentiment Analysis

Community Detection and Sentiment analysis are core components of our architecture. There are various preprocessing tasks required for both stages, and in this Section, along with discussing our choice of algorithm, we will also look at preprocessing steps involved for that particular stage.

3.1 Community Detection

In this section, we discuss community detection briefly in the context of social network analysis. For a more detailed introduction to community detection, refer to [23].

3.1.1 Preprocessing for Community Detection

To apply a community detection algorithm, we need to model tweets as mathematical graphs. Generally, a friend follower network is selected because community detection [62, 43], in general, is used to detect closely related groups. We are more interested in closely related topics than groups; hence we use hashtags to model our network [68]. Hashtags by nature represent the topic of discussion in that tweet, thus giving considerable information about that tweet [38]. To form communities based on hashtags, we first extract hashtags from the tweet and lowercase it to retain meaning irrespective of capitalization. After successfully extracting hashtags, we form combinations of 2 and link all the combinations together. This process is repeated on all tweets to make a network of hashtags. To make this network of hashtags(hashmap), we need general topic G. G would be used to collect data from twitter such that the hashmap generated from it would cover wide ranging topics.

3.1.2 Community Detection Algorithm

Communities are defined as sets of vertices which are densely interconnected whereas sparsely connected with the rest of the vertices[49]. We have various community detection algorithms such as Newmans leading eigenvector [46], Label Propagation [52], Louvian method for community detection [9],

infomap [57] and many more [13]. For our purposes, we want to find k communities instead of some random number of communities. Using k we can change the generality of our result i.e. the higher the value of k, the more specific a community would be. This is due to the fact that k denotes number of communities found and if there are lesser number of communities, the more general a community is. Hence to find k communities, we use the Fluid Communities algorithm as it allows us to provide insights into the graph structure at different levels of granularity[49].

Fluid communities is a propagation-based algorithm that is capable of identifying variable number of communities in a network. It is based on the idea of introducing number of communities within a non-homogeneous environment where communities will expand and compete until a stable state is reached. Given a graph $G = (V, E)$ where V is set of vertices and E is set of edges in the graph, Fluid communities algorithm initializes k communities i.e. $C = \{c_1, c_2, .., c_k\}$, where $0 < k < |V|$. each community is initialized in a different random vertex and is associated with density d as described in equation 1.

$$d = \frac{1}{|v \in c|} \tag{1}$$

Fluid community algorithm operates in supersteps and updates communities using an update rule until assignment of a vertex to that community does not change for two consecutive supersteps [49].

3.2 Sentiment Analysis

In this section, we take a look at sentiment analysis and the preprocessing steps required for performing sentiment analysis on tweets.

3.2.1 Preprocessing for Sentiment Analysis

To train sentiment classifier, we require labeled tweets which we gather using method described in [24]. In this method we use emoji's to collect data and label them based on polarity of that emoji. In this method we assume that tweets containing happy emojis like ':-), :), :D' will correlate to a positive tweet and tweets containing sad emoji's like ':-(, :(, =(' will correlate to a negative tweet. After data gathering is done, we first filter our tweets by converting text to lowercase, removing all hashtags, removing retweet designations ("RT"), usernames, and URLs. After this step, we remove all the stopwords from the NLTK corpus, and we perform tokenization and remove punctuations from the collected tweets.

3.2.2 Sentiment Analysis Algorithm

After preprocessing, we need to extract features from our tweets, and for that, we apply TFIDF Vectorization(refer equation 4) [61]. In equation 2, $f_{t,d}$ is the

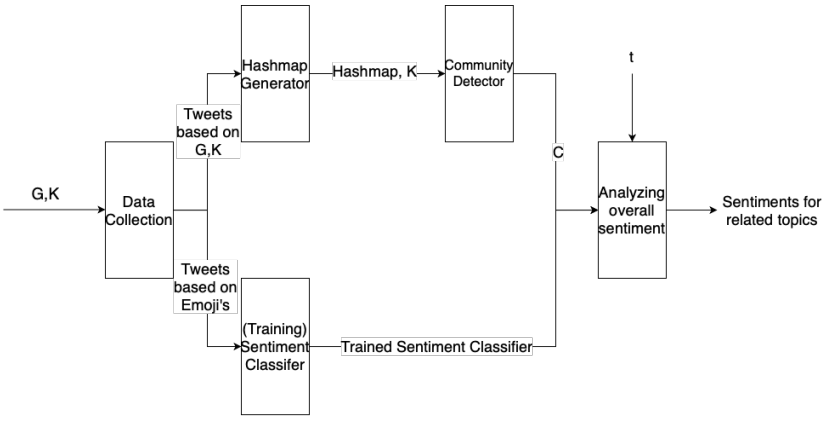

Fig. 1 Architecture for Overall Topic Sentiment Classifier model

raw count of a term t in the document d. In equation 3, N is the total number of Documents, i.e. $|D|$. The resulting features are passed to the Multinomial Naive Bayes Classifier[33], which classifies tweets positively or negatively.

$$\mathrm{tf}(t,d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}} \tag{2}$$

$$\mathrm{idf}(t,D) = \log \frac{N}{|\{d \in D : t \in d\}|} \tag{3}$$

$$\mathrm{tfidf}(t,d,D) = \mathrm{tf}(t,d) \cdot \mathrm{idf}(t,D) \tag{4}$$

Naive Bayes classifier is based on Bayes theorem [5] where s is a sentiment, M is a Twitter message. Because we have equal sets of positive and negative tweets we can simplify the equation as:

$$P(s|M) = \frac{P(s) \cdot P(M|s)}{P(M)} \tag{5}$$

$$P(s|M) = \frac{P(M|s)}{P(M)} \tag{6}$$

$$P(s|M) \sim P(M|s) \tag{7}$$

After training our sentiment classifier on our training data for sentiment classification, it performs with 77% accuracy. When we calculate f1 score, we get 76% for Positive labels and 78% for Negative labels.

4 Overall Topic Sentiment Classifier

Fig. 1 gives a general overview of our architecture. The first step involves collecting data and preprocessing it for training our sentiment classifier and generating a hashmap based on a general topic G (Refer section 3.1.1). After preprocessing is done, we use the hashmap to detect communities using fluid community detection algorithm. It takes in k that determines how many communities should be formed [49]. By default, we divide our hashmap into ten communities (i.e. $k = 10$). We can fine-tune these hyperparameters based on our requirements. After finding communities and training our classifier, the trained sentiment classifier and detected communities(C) are passed for analyzing overall sentiments for related topics. In this step we pass a topic t that we use along with C to find related topics and perform sentiment analysis for those topics.

4.1 Analyzing overall sentiments

This module takes in a general topic G, detected communities C, hashmap from the previous step and a topic t. It first uses a hashmap and t to calculate the most valued, directly related topics R_t. We can find this by looking at neighboring nodes of t, and we pick at max ten topics with the highest weight. We apply equation 8 to find suitable community (S_t) for t.

$$S_t = max(R_t \cap c) \quad \forall c \in C \tag{8}$$

In equation 8, C is set of communities we detect using FluidC algorithm, and R_t is set of directly connected topics we select using hashmap and topic t.

$$Q_f = \frac{w * d}{w + d} \tag{9}$$

After this step, we apply equation 9 to all the nodes within S_t and pick 10% nodes with highest Quality factor Q_f. It ensures that the topics we pick in a community are of high quality i.e. have high degree and its combined weight with adjacent edges is high. In equation 9, w is total weight of node with its adjacent nodes and d is the degree of the node under consideration. We pass those 10% selected topics along with a value n to our sentiment classifier. n denotes the number of tweets to fetch for each topic for sentiment analysis. To demonstrate, we take $n = 1000$ and use Twitter API to fetch tweets for our selected topics. The sentiment classifier then classifies sentiment for each tweet. To keep track of overall sentiment, we initialize T with 0 and increment it by 1 for every positive tweet and decrement by 1 for every negative tweet. To normalize, we divide T by n. Finally, the result we get is in the range -1 to 1 where 1 denotes every post encountered is a positive post and -1 denotes every post encountered is a negative post. The greater the output sentiment, the more positive it is.

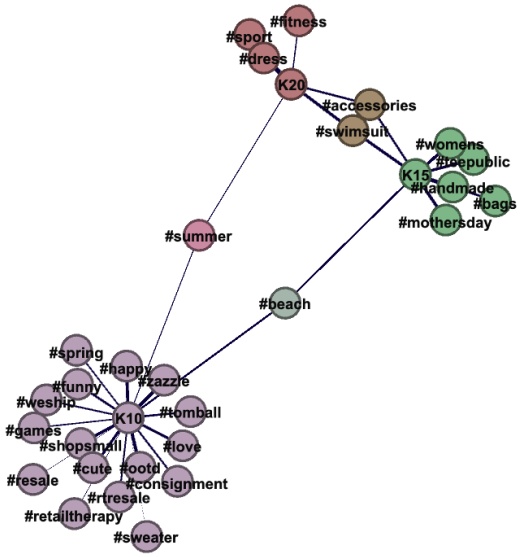

Fig. 2 Resultant closely related topics for $G = \#shopping$ and $t = \#summer$ (k=10, k=15 and k=20)

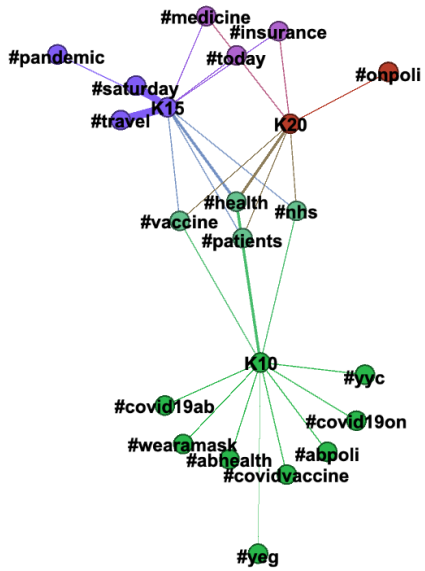

Fig. 3 Resultant closely related topics for $G = \#covid19$ and $t = \#vaccine$ (k=10, k=15 and k=20)

	summer-shopping	summer-sentiments	covid-vaccine	covid-sentiments
0	#love	0.80	#vaccine	-0.72
1	#shopsmall	0.85	#health	0.22
2	#spring	0.63	#covidvaccine	-0.29
3	#ootd	0.91	#wearamask	-0.04
4	#cute	0.81	#covid19ab	-0.40
5	#summer	0.54	#patients	-0.37
6	#happy	0.89	#yeg	0.05
7	#zazzle	0.94	#yyc	-0.15
8	#funny	0.75	#abpoli	-0.14
9	#games	0.60	#nhs	-0.24
10	#retailtherapy	0.47	#covid19on	-0.14
11	#sweater	0.40	#abhealth	-0.41
12	#beach	0.70	#pandemic	-0.13
13	#consignment	0.68	#saturday	0.56
14	#weship	0.66	#travel	0.62
15	#resale	0.45	#medicine	-0.45
16	#tomball	0.72	#today	-0.12
17	#rtresale	0.64	#insurance	-0.48
18	#womens	0.88	#onpoli	-0.24
19	#accessories	0.73		
20	#handmade	0.99		
21	#swimsuit	0.86		
22	#bags	0.73		
23	#teepublic	0.83		
24	#mothersday	0.88		
25	#dress	0.93		
26	#fitness	0.58		
27	#sport	0.59		

Table 1 Overall Sentiment Table for G='#summer', t='#shopping' and G='#covid19', t='#vaccine'

5 Results

Twitter users post messages about a range of topics unlike other sites which are designed for a specific topic. Users use hashtags (#) to mark topics a tweet talks or is related about [38]. We propose an OTSC model that uses this feature of twitter to make a hashmap of a general topic G and use this hashmap to find closely related topics of a given topic t. Furthermore, it finds overall sentiments of top 10% topics among the found topics. We use our proposed OTSC model and apply it to G=#summer and t=#shopping with k set to 10, 15 and 20. Similarly we apply our model to G=#covid19 and t=#vaccine, G=#politics and t=#issues and G=#electricvehicles and t=#tesla with k set to 10, 15 and 20.

The found topics can be referred in Fig. 2, 3, 4 and 5. In each of these figures, we added 3 nodes k10, k15 and k20. All the topics connected to k10 are found when $k = 10$, similarly, k15 corresponds to topics found when $k = 15$ and k20 for $k = 20$. Sentiments related to corresponding topics can be referred to in Table 1 and 2. In Fig. 2, node k10 is connected with 18 topics, k15 is connected with 8 topics and k20 is connected with 6 topics. k10 and k15 have

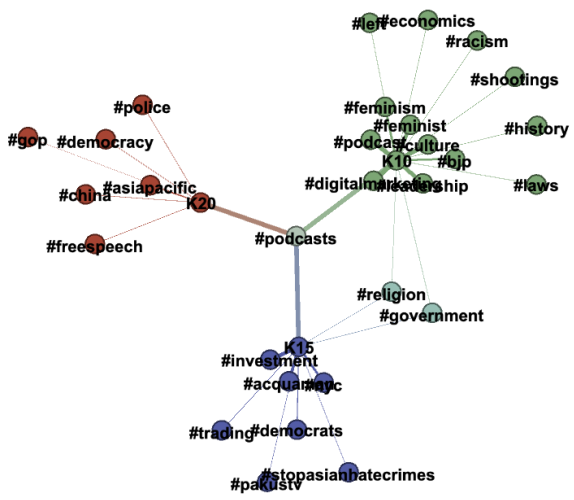

Fig. 4 Resultant closely related topics for $G = \#politics$ and $t = \#issues$ (k=10, k=15 and k=20)

1 topic in common, k10 and k20 have 1 topic in common whereas k15 and k20 have 2 topics in common. In Fig. 3, node k10 is connected with 12 topics, k15 is connected with 10 topics and k20 is connected with 8 topics. k10 and k15 have 4 topics in common, k10 and k20 have 4 topics in common whereas k15 and k20 have 7 topics in common. In Fig. 4, node k10 is connected with 16 topics, k15 is connected with 10 topics and k20 is connected with 7 topics. k10 and k15 have 3 topics in common, k10 and k20 have 1 topic in common whereas k15 and k20 have 1 topic in common. Finally for Fig. 5, node k10 is connected with 17 topics, k15 is connected with 11 topics and k20 is connected with 9 topics. k10 and k15 have 10 topics in common, k10 and k20 have 8 topics in common whereas k15 and k20 have 7 topics in common.

We find most topics positively viewed for Fig. 2 and most topics negatively viewed for Fig. 3 (Refer Table 1). For Fig. 4 some topics are positive and some are negative whereas for Fig. 5 most topics are positive while some being negative (Refer Table 2).

6 Discussion

It is important to understand emerging trends and public opinion about them to make informed decision and gain actionable insights [58,3]. These trend can be used for marketing analysis as used by [55] to understand implications of emerging trend and how advertisements can be made while keeping them

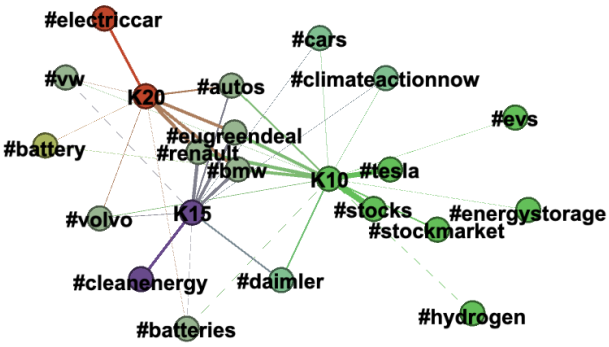

Fig. 5 Resultant closely related topics for $G = \#electricvehicles$ and $t = \#tesla$ (k=10, k=15 and k=20)

in mind. Similar to this, much research have been done to analyze emerging trends and their marketing implications [31, 1, 60, 34, 10, 54, 42, 11]. We can also observe policy changes of government to incorporate and promote sustainable development [41, 27, 21].

Using these research we can understand the importance of analysis of emerging trends. Given that, most of the work done to find topics are based of Latent Dirichlet Allocation(LDA) [59], which works well but it fails to incorporate hashtags that are primarily used for marking topics in a tweet [38]. Furthermore, there is no means using which we can change generality of the found topics when we use LDA. To include these features we use Fluid Community Detection algorithm that allows us to find k communities using which we can change granularity of the resultant communities [49].

We applied our model to find trends in the following (G, t) pairs: (#shopping,#summer),(#covid19,#vaccine), (#politics,#issues),(#electricvehicles , #tesla) for $k = 10, 15\&20$. In each cases, we found maximum number of topics for $k = 10$ and minimum number of topics for $k = 20$. From this we can infer that the topics found tends to be more specific due to the shrinkage of number of nodes in individual communities as we increase k(Number of communities) [49].

In Fig. 2 we found some interesting topics like #ootd which stands for outfit of the day and its corresponding sentiment (Refer Table 1) seems to be among the highest. This provides a potential advertising keyword and tech-

	politics-issues	politics-sentiments	ev-tesla	ev-sentiments
0	#government	-0.13	#evs	-0.24
1	#podcast	0.82	#tesla	0.53
2	#racism	-0.51	#cars	0.11
3	#history	-0.04	#battery	-0.36
4	#feminism	0.42	#bmw	0.35
5	#leadership	0.74	#vw	-0.04
6	#religion	0.09	#eugreendeal	0.35
7	#digitalmarketing	0.85	#renault	0.38
8	#bjp	0.38	#autos	0.23
9	#culture	0.75	#volvo	0.13
10	#shootings	-0.77	#batteries	-0.19
11	#feminist	0.69	#daimler	0.22
12	#laws	-0.18	#climateactionnow	0.11
13	#podcasts	0.84	#energystorage	-0.01
14	#left	-0.29	#hydrogen	-0.35
15	#economics	-0.71	#stocks	0.60
16	#democrats	0.18	#stockmarket	0.20
17	#investment	0.68	#cleanenergy	0.33
18	#nyc	0.42	#electriccar	0.31
19	#trading	0.12		
20	#acquaman	0.56		
21	#stopasianhatecrimes	-0.14		
22	#pakustv	-0.23		
23	#china	-0.20		
24	#democracy	-0.09		
25	#police	-0.52		
26	#gop	0.06		
27	#freespeech	-0.19		
28	#asiapacific	0.21		

Table 2 Overall Sentiment Table for G='#politics', t='#issues' and G='#electricvehicles' and t='#tesla'

nique that marketing agencies can use which is also supported by paper [14]. Also emerging topics like swimsuit, beach, accessories, handmade with above average sentiment suggests that people may tend to buy items related to these topics. This also opens a door for business opportunity in accessories, handmade products, bags, and custom designs made using zazzle. Topics such as love, cute, shopsmall, weship, mothersday and retail therapy can be used by marketing agencies to promote their products as they correspond to a positive public opinion. Given that, one should be careful to use the keyword "retail therapy" as its below average public opinion.

For topics relating to (#covid19, #vaccine) we found a general negative sentiment. Particularly for #vaccine which might point towards vaccine hesitancy. A research [56] suggests that about 31% of Americans wish to take wait and see approach and about 20% remain quite reluctant about it. Other reason for its negative sentiment might be because several European nations are suspending the use of Astrazenca covid-19 vaccine [44]. Furthermore, A negative sentiment in patients might indicate increasing number of covid patients. A positive travel sentiment paired with #yeg (Edmonton International Airport)

and #yyc (Calgary International Airport) might suggest ease of lockdown and possible investment opportunity in travel sector[45]. A positive sentiment on #health asserts that people are health conscious which provides opportunity related to healthcare and organic products [64]. A neutral sentiment of #wearamask suggests that many people have negative sentiment about it where our results match with a similar research which states that among 4099 respondents only 53.3% of symptotic participants reported wearing a mask in preceding week and about 62% people without symptoms did not wear a mask in prior week [20] pointing towards need for education about importance of mask.

Results from (#politics, #issues) gives us a list of pressing issues such as racism, shootings and freespeach. It also points out emerging asian hate (#stopasianhatecrimes) that recently emerged due to current covid19 pandemic[69]. Additionally, a positive viewpoint on OSN shows that people tend to view it positively suggesting that schemes promoting equality might be viewed positively[55]. Podcasts being connected to all three nodes (k10,k15 and k20) might suggest that people are often using podcasts to listen to or share opinions on politics and pressing issues. Advertisers might want to use that medium to advertise related content.

Finally Fig. 5 is one of the most overlapping graph among the topics we explored. These overlaps mostly points towards competitors of tesla such as renault, bmw, daimier, volvo and vw(volkswagen). Looking at the sentiments, we can infer that tesla might be one of the most positively viewed vehicle in electric vehicle domain [66]. Furthermore, the occurance of topics such as stocks and stockmarket indicate that people that talk about tesla or electric vehicles might also be related to investment community. A positive sentiment in #climateactionnow and #eugreendeal indicates that people are optimistic of sustainable alternatives[47] that can be considered while entering a business domain or while creating an advert. It also indicates that governments work on related policies might be positively viewed [41,27,47].

It is also important to note that in most cases overlapping topics are of greater importance as compared with non overlapping topics. e.g. beach, swimsuit and accessories for Fig. 2, vaccine and health and patients in Fig. 3, podcasts and government in Fig. 4 and different automakers such as renault, bmw, daimer, volvo and volkswagen in 5.

7 Conclusion

Analysis of emerging trends is important to both businesses and policy makers. In this paper we propose an OTSC model (Fig. 1) that can be used to get emerging trends along with their public sentiment based on Twitter tweets. We propose using Fluid Community Detection algorithm instead of generally used LDA for finding related topics in Twitter. This is because LDA fails to incorporate Twitters intrinsic features such as hashtags and does not provide metrics to change generality of the found topics. With the help of Fluid Com-

munity Detection algorithm, our model is capable of changing the granularity of underlying community thereby changing the generality of the found topics. We further assert this by applying our model to following (G, t) pairs: 1. (#summer,#shopping), 2. (#covid19,#vaccine), 3. (#politics,#issues) and 4. (#electricvehicles,#tesla) for k=10,15 and 20. Our resulting topics are presented in Fig. 2,3,4 and 5. Their corresponding sentiments are in Table 1 and 2. We found that for k=10, the number of topics found were maximum and for k=20 they were minimum pointing that as number of community increase, the topics tend to be more specific thereby less general. We further analyzed our results in context of business analysis and policy analysis and found that we can answer questions like what are the emerging trends, positively viewed keywords, key competitors, find pressing and emerging issues and general sentiment around a topic. This information can then be used to make informed business and political decisions.

7.1 Managerial Implications

Discovering emerging trends and issues have several applications in business analysis and management. Our research proposes an OTSC model to discover and analyze public sentiments about those emerging trends. Emerging trends can be used to understand a broader market picture potentially helping with business and managerial changes. One such example that we discussed is about emerging trends for shopping in summer that helps us understand how people are more positive about handmade products, swimwear and potentially fitness and sports products during summer. Furthermore, we can understand declining trends such as for sweater as compared with other trends for the same query. A closer look at electric vehicles and tesla points us that clean energy is positively viewed at. One can apply this knowledge to create a positive customer/client outlook by promoting environmental friendly approach. Given this information, it is important to understand how this might affect business and a prior knowledge about the domain is required. Prior knowledge is also necessary to set apart topics for business use and keywords for adverts in the found topics. Our model gives a broader view of the subject but to take decisions, one need to understand specifics of the topic of interest keeping in mind its broader implications.

7.2 Practical/Social Implications for marketers

Interactivity on the internet shifts the ways in which users perceive advertising. This research provides practical implications on how advertisers can use interactions among users to understand keywords that might give a better perception of their adverts. For instance we found ootd, love, cute, retailtherapy and shopsmall for shopping in summer. Based on user interactions, our model might also suggest relevant advertisement means (Podcasts for political advertisements) as found in Table 2. Other than that, corresponding sentiments

related to keywords can also help advertisers understand how a keyword might affect advertisement. For instance, retailtherapy have below average sentiment score as compared to other found topics. This may indicate that advertisers should take caution while using this keyword and better understanding of using this keyword might be required.

8 Future Work and Limitations

Our proposed model uses Fluid Community Detection algorithm and it does not always return the same result during each run [63]. Finding a suitable value of k requires trial and error furthermore results may vary at each run. We hand picked topics for analysis which might include some bias. Furthermore, topics found are open to interpretation and hence subjective. In Future we can try an iterative model that uses OTSC as a base and applies it for different values of k. This work can also be extended for analysis of different trends similar to the following works [10, 54, 42, 11]. This work can also be extended to better understand importance of overlapping topics for different values of k.

References

1. Alalwan, A.A., Rana, N.P., Dwivedi, Y.K., Algharabat, R.: Social media in marketing: A review and analysis of the existing literature. Telematics and Informatics **34**(7), 1177–1190 (2017)
2. Alamsyah, A., Rahardjo, B., et al.: Social network analysis taxonomy based on graph representation. arXiv preprint arXiv:2102.08888 (2021)
3. Alsini, A., Datta, A., Huynh, D.Q., Li, J.: Community aware personalized hashtag recommendation in social networks. In: Australasian Conference on Data Mining, pp. 216–227. Springer (2018)
4. Ansari, M.Z., Aziz, M., Siddiqui, M., Mehra, H., Singh, K.: Analysis of political sentiment orientations on twitter. Procedia Computer Science **167**, 1821–1828 (2020)
5. Anthony, J.H.: Probability and statistics for engineers and scientists. Thomson Brooks/Cole (2007)
6. Bello-Orgaz, G., Mesas, R.M., Zarco, C., Rodriguez, V., Cordón, O., Camacho, D.: Marketing analysis of wineries using social collective behavior from users' temporal activity on twitter. Information Processing & Management **57**(5), 102220 (2020)
7. Berlingerio, M., Coscia, M., Giannotti, F., Monreale, A., Pedreschi, D.: Multidimensional networks: foundations of structural analysis. World Wide Web **16**(5-6), 567–593 (2013)
8. Bhatnagar, S., Dixit, M., Prasad, N.: A review of common approaches to sentiment analysis and community detection. International Journal of Computer Applications **975**, 8887 (2020)
9. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment **2008**(10), P10008 (2008)
10. Boon-Itt, S., Skunkan, Y.: Public perception of the covid-19 pandemic on twitter: sentiment analysis and topic modeling study. JMIR Public Health and Surveillance **6**(4), e21978 (2020)
11. Chandrasekaran, R., Mehta, V., Valkunde, T., Moustakas, E.: Topics, trends, and sentiments of tweets about the covid-19 pandemic: Temporal infoveillance study. Journal of medical Internet research **22**(10), e22624 (2020)

12. Chevalier-Roignant, B., Trigeorgis, L.: Competitive Strategy: Options and Games, vol. 1, 1 edn. The MIT Press (2012). URL https://EconPapers.repec.org/RePEc:mtp:titles:0262015994
13. Chunaev, P.: Community detection in node-attributed social networks: a survey. Computer Science Review **37**, 100286 (2020)
14. Dar, T.M., Tariq, N.: Celebrities and influencers: Have they changed the game of online marketing? European Journal of Business and Management Research **6**(1), 106–111 (2021)
15. Davenport, T.: Big data at work: dispelling the myths, uncovering the opportunities. Harvard Business Review Press (2014)
16. Davidov, D., Tsur, O., Rappoport, A.: Enhanced sentiment learning using twitter hashtags and smileys. In: Coling 2010: Posters, pp. 241–249 (2010)
17. Deitrick, W., Hu, W.: Mutually enhancing community detection and sentiment analysis on twitter networks (2013)
18. Diamantini, C., Mircoli, A., Potena, D., Storti, E.: Social information discovery enhanced by sentiment analysis techniques. Future Generation Computer Systems **95**, 816–828 (2019)
19. Direction, S.: Firm capacity to manage new trends: Business model innovation can increase resilience (2021)
20. Egan, M., Acharya, A., Sounderajah, V., Xu, Y., Mottershaw, A., Phillips, R., Ashrafian, H., Darzi, A.: Evaluating the effect of infographics on public recall, sentiment and willingness to use face masks during the covid-19 pandemic: a randomised internet-based questionnaire study. BMC public health **21**(1), 1–10 (2021)
21. Elkerbout, M., Egenhofer, C., Núñez Ferrer, J., Catuti, M., Kustova, I., Rizos, V., et al.: The european green deal after corona: Implications for eu climate policy. CEPS Policy Insights (2020–06) (2020)
22. Fortunato, S.: Community structure in complex networks.
23. Fortunato, S.: Community detection in graphs. Physics reports **486**(3-5), 75–174 (2010)
24. Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. CS224N project report, Stanford **1**(12), 2009 (2009)
25. Hall, D., Lutsey, N.: Electric vehicle charging guide for cities. Consulting Report (2020)
26. Hassani, H., Beneki, C., Unger, S., Mazinani, M.T., Yeganegi, M.R.: Text mining in big data analytics. Big Data and Cognitive Computing **4**(1), 1 (2020)
27. Hishan, S.S., Khan, A., Ahmad, J., Hassan, Z.B., Zaman, K., Qureshi, M.I., et al.: Access to clean technologies, energy, finance, and food: environmental sustainability agenda and its implications on sub-saharan african countries. Environmental Science and Pollution Research **26**(16), 16503–16518 (2019)
28. Hsieh, I.Y.L., Pan, M.S., Green, W.H.: Transition to electric vehicles in china: Implications for private motorization rate and battery market. Energy Policy **144**, 111654 (2020)
29. Hu, X., Tang, L., Tang, J., Liu, H.: Exploiting social relations for sentiment analysis in microblogging. In: Proceedings of the sixth ACM international conference on Web search and data mining, pp. 537–546 (2013)
30. Hubert, M., Linzmajer, M., Riedl, R., Hubert, M., Kenning, P., Weber, B.: The use of psycho-physiological interaction analysis with fmri-data in is research–a guideline. Communications of the Association for Information Systems (CAIS) **40**(9), 181–217 (2017)
31. Iyengar, R., Van den Bulte, C., Valente, T.W.: Opinion leadership and social contagion in new product diffusion. Marketing Science **30**(2), 195–212 (2011)
32. Karataş, A., Şahin, S.: Application areas of community detection: A review. In: 2018 International congress on big data, deep learning and fighting cyber terrorism (IBIGDELFT), pp. 65–70. Ieee (2018)
33. Kibriya, A.M., Frank, E., Pfahringer, B., Holmes, G.: Multinomial naive bayes for text categorization revisited. In: Australasian Joint Conference on Artificial Intelligence, pp. 488–499. Springer (2004)
34. Kim, A.E., Hansen, H.M., Murphy, J., Richards, A.K., Duke, J., Allen, J.A.: Methodological considerations in analyzing twitter data. Journal of the National Cancer Institute Monographs **2013**(47), 140–146 (2013)

35. Kontopoulos, E., Berberidis, C., Dergiades, T., Bassiliades, N.: Ontology-based sentiment analysis of twitter posts. Expert systems with applications **40**(10), 4065–4074 (2013)
36. Kouloumpis, E., Wilson, T., Moore, J.: Twitter sentiment analysis: The good the bad and the omg! In: Fifth International AAAI conference on weblogs and social media. Citeseer (2011)
37. Kulshrestha, J., Zafar, M., Noboa, L., Gummadi, K., Ghosh, S.: Characterizing information diets of social media users. In: Proceedings of the International AAAI Conference on Web and Social Media, vol. 9 (2015)
38. Kumar, A., Sebastian, T.M.: Sentiment analysis on twitter. International Journal of Computer Science Issues (IJCSI) **9**(4), 372 (2012)
39. Lăzăroiu, G., Neguriţă, O., Grecu, I., Grecu, G., Mitran, P.C.: Consumers' decision-making process on social commerce platforms: Online trust, perceived risk, and purchase intentions. Frontiers in Psychology **11** (2020)
40. Leskovec, J., Lang, K.J., Dasgupta, A., Mahoney, M.W.: Statistical properties of community structure in large social and information networks. In: Proceedings of the 17th international conference on World Wide Web, pp. 695–704 (2008)
41. Li, Y., Zhan, C., de Jong, M., Lukszo, Z.: Business innovation and government regulation for the promotion of electric vehicle use: lessons from shenzhen, china. Journal of Cleaner Production **134**, 371–383 (2016)
42. Lorenz-Spreen, P., Wolf, F., Braun, J., Ghoshal, G., Conrad, N.D., Hövel, P.: Tracking online topics over time: understanding dynamic hashtag communities. Computational social networks **5**(1), 1–18 (2018)
43. Luo, L., Liu, K., Guo, B., Ma, J.: User interaction-oriented community detection based on cascading analysis. Information Sciences **510**, 70–88 (2020)
44. Mahase, E.: Covid-19: Who says rollout of astrazeneca vaccine should continue, as europe divides over safety (2021)
45. Nayak, J., Mishra, M., Naik, B., Swapnarekha, H., Cengiz, K., Shanmuganathan, V.: An impact study of covid-19 on six different industries: Automobile, energy and power, agriculture, education, travel and tourism and consumer electronics. Expert Systems (2021)
46. Newman, M.E.: Finding community structure in networks using the eigenvectors of matrices. Physical review E **74**(3), 036104 (2006)
47. O'Riordan, T.: Environmental science, sustainability and politics. Transactions of the Institute of British Geographers **29**(2), 234–247 (2004)
48. Pak, A., Paroubek, P.: Twitter as a corpus for sentiment analysis and opinion mining. In: LREc, vol. 10, pp. 1320–1326 (2010)
49. Parés, F., Gasulla, D.G., Vilalta, A., Moreno, J., Ayguadé, E., Labarta, J., Cortés, U., Suzumura, T.: Fluid communities: A competitive, scalable and diverse community detection algorithm. In: International Conference on Complex Networks and their Applications, pp. 229–240. Springer (2017)
50. Pfeffer, J., Mayer, K., Morstatter, F.: Tampering with twitter's sample api. EPJ Data Science **7**(1), 50 (2018)
51. Puthussery, A.: Digital marketing: An overview (2020)
52. Raghavan, U.N., Albert, R., Kumara, S.: Near linear time algorithm to detect community structures in large-scale networks. Physical review E **76**(3), 036106 (2007)
53. Read, J.: Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In: Proceedings of the ACL student research workshop, pp. 43–48 (2005)
54. Reyes-Menendez, A., Saura, J.R., Alvarez-Alonso, C.: Understanding# worldenvironmentday user opinions in twitter: A topic-based sentiment analysis approach. International journal of environmental research and public health **15**(11), 2537 (2018)
55. Reyes-Menendez, A., Saura, J.R., Filipe, F.: Marketing challenges in the# metoo era: Gaining business insights using an exploratory sentiment analysis. Heliyon **6**(3), e03626 (2020)
56. Rosenbaum, L.: Escaping catch-22—overcoming covid vaccine hesitancy (2021)
57. Rosvall, M., Bergstrom, C.T.: Maps of random walks on complex networks reveal community structure. Proceedings of the National Academy of Sciences **105**(4), 1118–1123 (2008)

58. Saura, J.R.: Using data sciences in digital marketing: Framework, methods, and performance metrics. Journal of Innovation & Knowledge (2020)
59. Saura, J.R., Bennett, D.R.: A three-stage method for data text mining: Using ugc in business intelligence analysis. Symmetry **11**(4), 519 (2019)
60. Saura, J.R., Reyes-Menendez, A., Bennett, D.R.: How to extract meaningful insights from ugc: A knowledge-based method applied to education. Applied Sciences **9**(21) (2019). DOI 10.3390/app9214603. URL https://www.mdpi.com/2076-3417/9/21/4603
61. Singh, A.K., Shashi, M.: Vectorization of text documents for identifying unifiable news articles. Int. J. Adv. Comput. Sci. Appl **10** (2019)
62. Solomon, R.S., Srinivas, P., Das, A., Gamback, B., Chakraborty, T.: Understanding the psycho-sociological facets of homophily in social network communities. IEEE Computational Intelligence Magazine **14**(2), 28–40 (2019)
63. Sun, Z., Sun, Y., Chang, X., Wang, Q., Yan, X., Pan, Z., Li, Z.p.: Community detection based on the matthew effect. Knowledge-Based Systems **205**, 106256 (2020)
64. Tandon, A., Dhir, A., Kaur, P., Kushwah, S., Salo, J.: Why do people buy organic food? the moderating role of environmental concerns and trust. Journal of Retailing and Consumer Services **57**, 102247 (2020)
65. Tavakolifard, M., Almeroth, K.C.: Social computing: an intersection of recommender systems, trust/reputation systems, and social networks. IEEE Network **26**(4), 53–58 (2012)
66. Thomas, V., Maine, E.: Market entry strategies for electric vehicle start-ups in the automotive industry–lessons from tesla motors. Journal of Cleaner Production **235**, 653–663 (2019)
67. Wu, S., Hofman, J.M., Mason, W.A., Watts, D.J.: Who says what to whom on twitter. In: Proceedings of the 20th international conference on World wide web, pp. 705–714 (2011)
68. Xiao, F., Noro, T., Tokuda, T.: Finding news-topic oriented influential twitter users based on topic related hashtag community detection. J. Web Eng. **13**(5&6), 405–429 (2014)
69. Xu, J., Sun, G., Cao, W., Fan, W., Pan, Z., Yao, Z., Li, H.: Stigma, discrimination, and hate crimes in chinese-speaking world amid covid-19 pandemic. Asian journal of criminology pp. 1–24 (2021)
70. Xu, K., Li, J., Liao, S.S.: Sentiment community detection in social networks. In: Proceedings of the 2011 iConference, pp. 804–805 (2011)
71. Yadav, A., Vishwakarma, D.K.: Sentiment analysis using deep learning architectures: a review. Artificial Intelligence Review **53**(6), 4335–4385 (2020)
72. Zhang, Q., Zhang, Z., Yang, M., Zhu, L.: Exploring coevolution of emotional contagion and behavior for microblog sentiment analysis: A deep learning architecture. Complexity **2021** (2021)

Figures

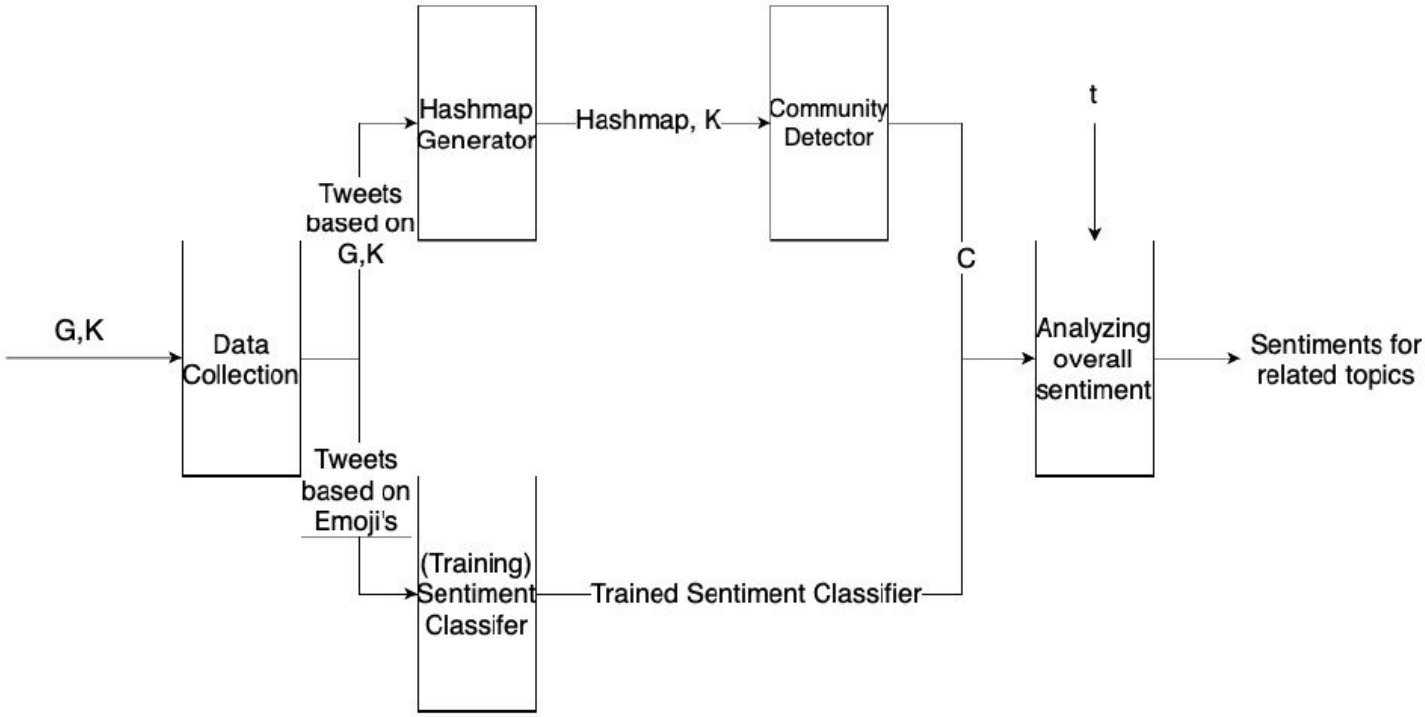

Figure 1

Architecture for Overall Topic Sentiment Classier model

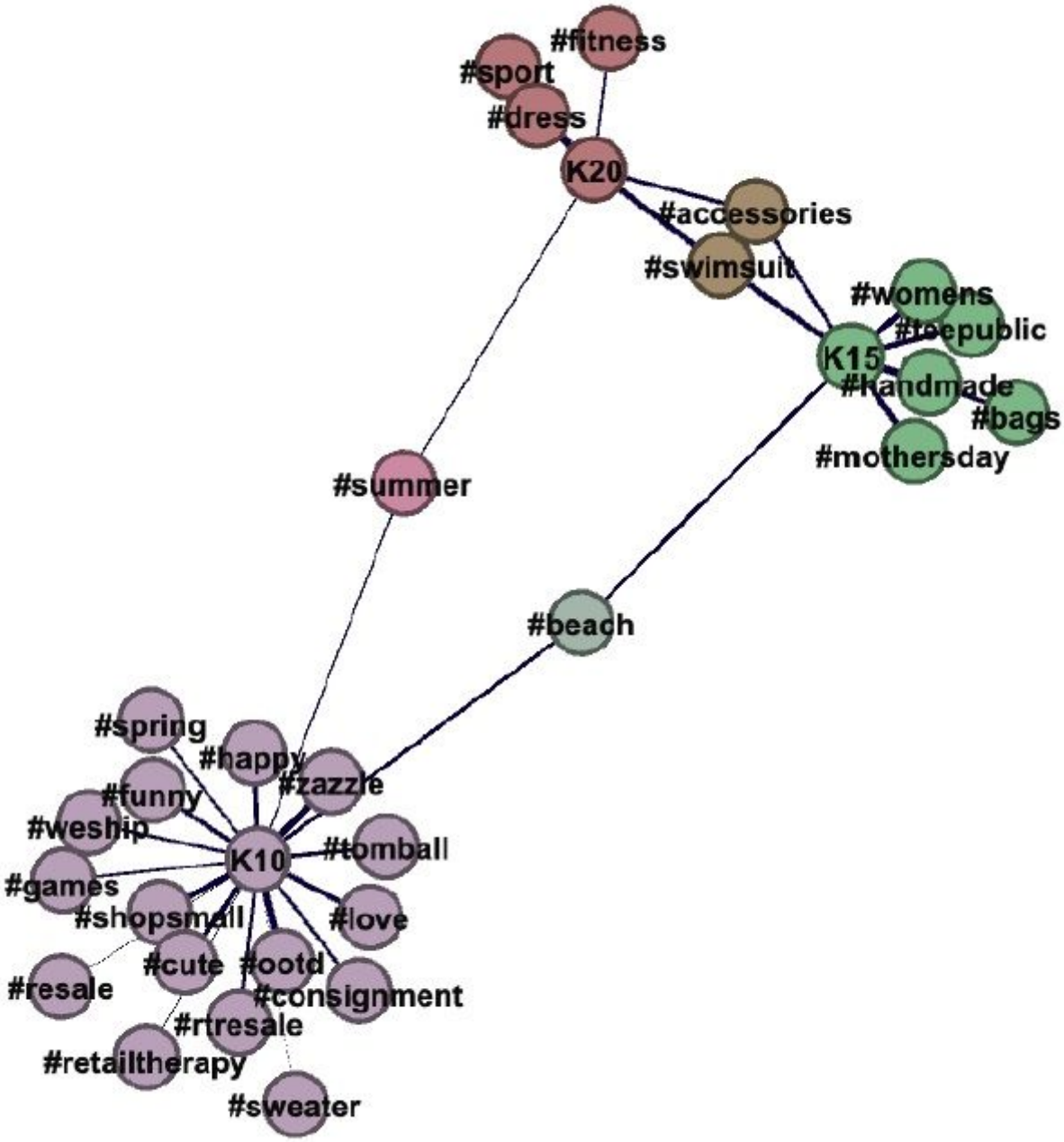

Figure 2

Resultant closely related topics for G = #shopping and t = #summer (k=10, k=15 and k=20)

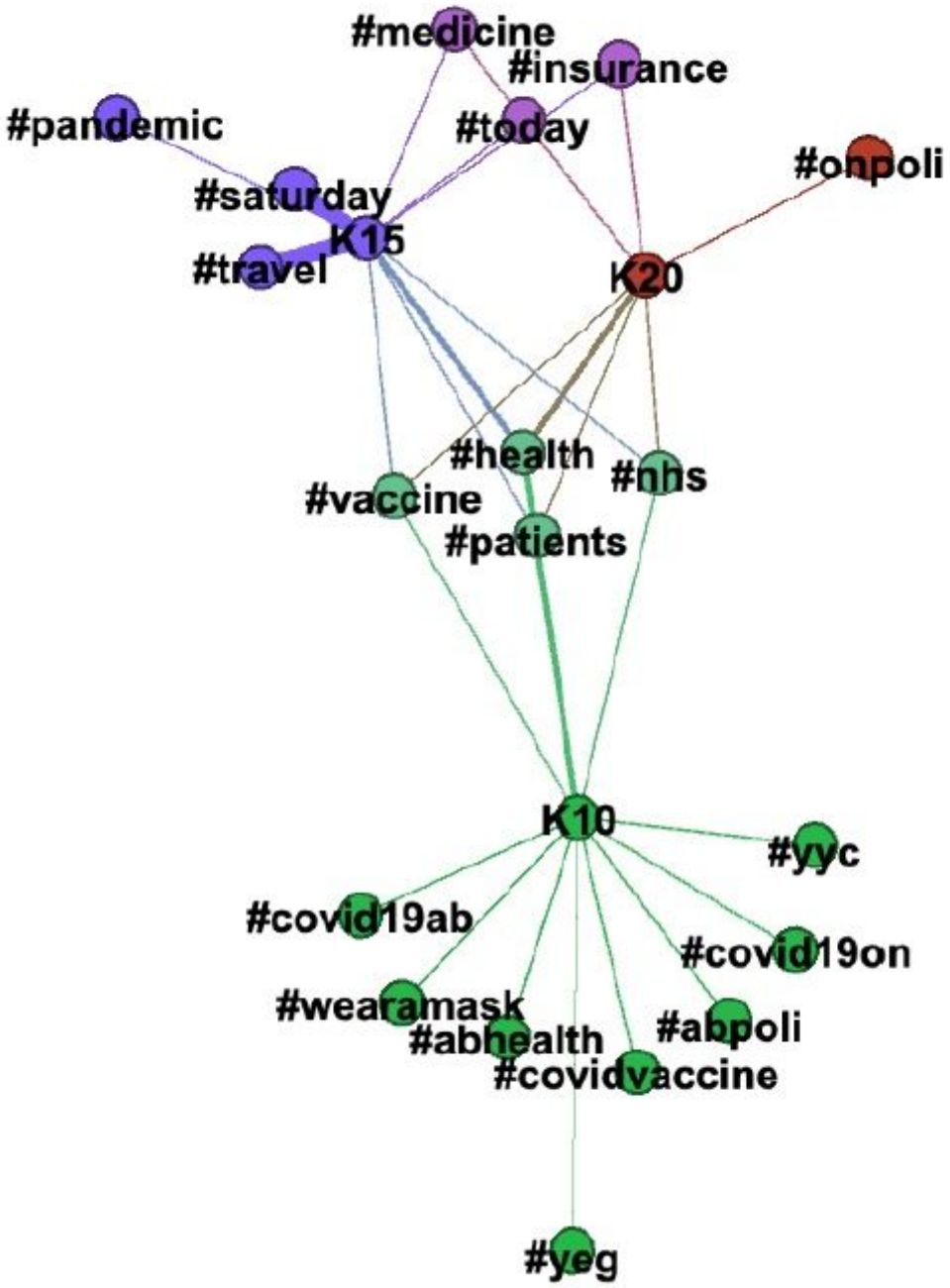

Figure 3

Resultant closely related topics for G = #covid19 and t = #vaccine (k=10, k=15 and k=20)

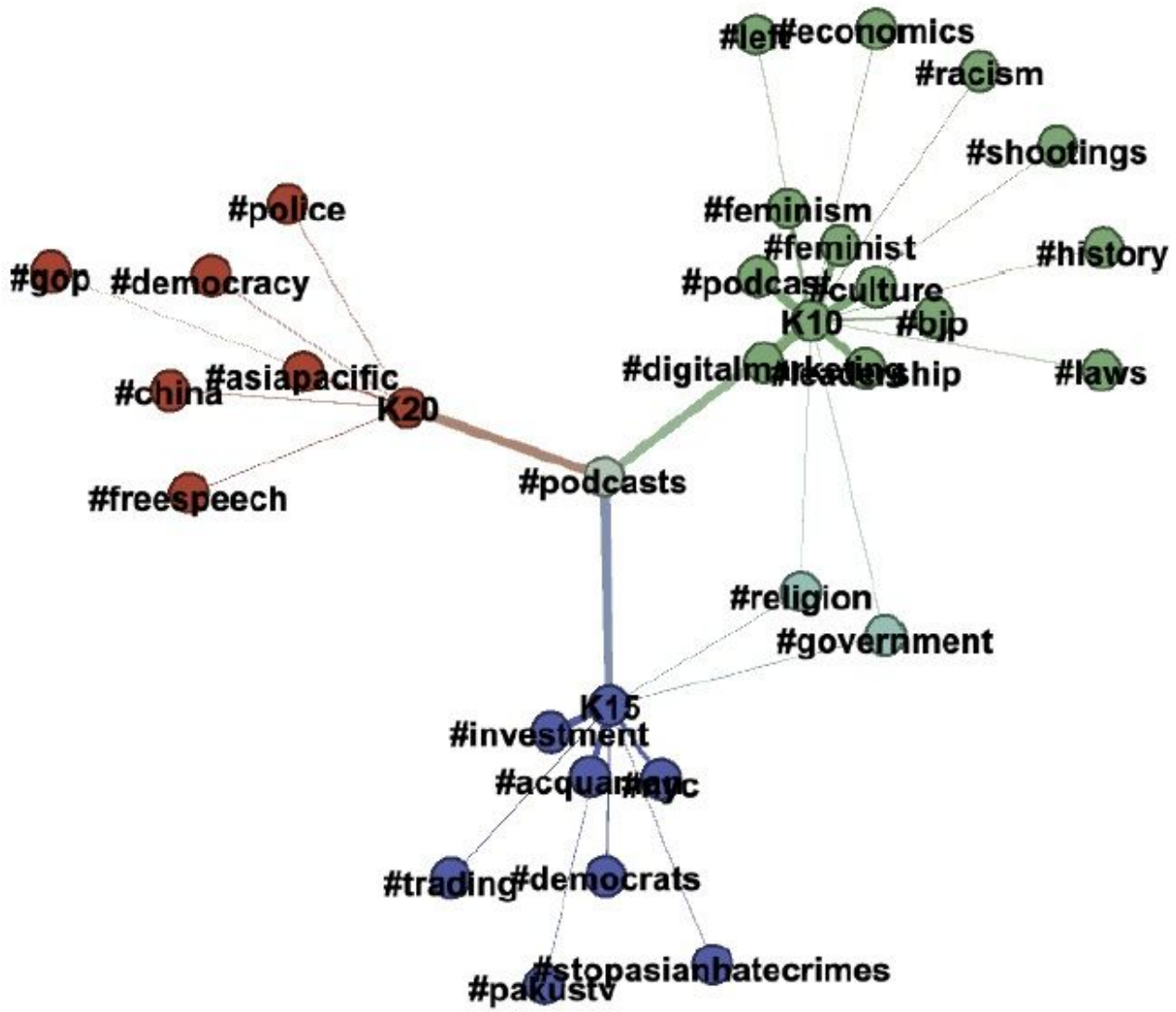

Figure 4

Resultant closely related topics for G = #politics and t = #issues (k=10, k=15 and k=20)

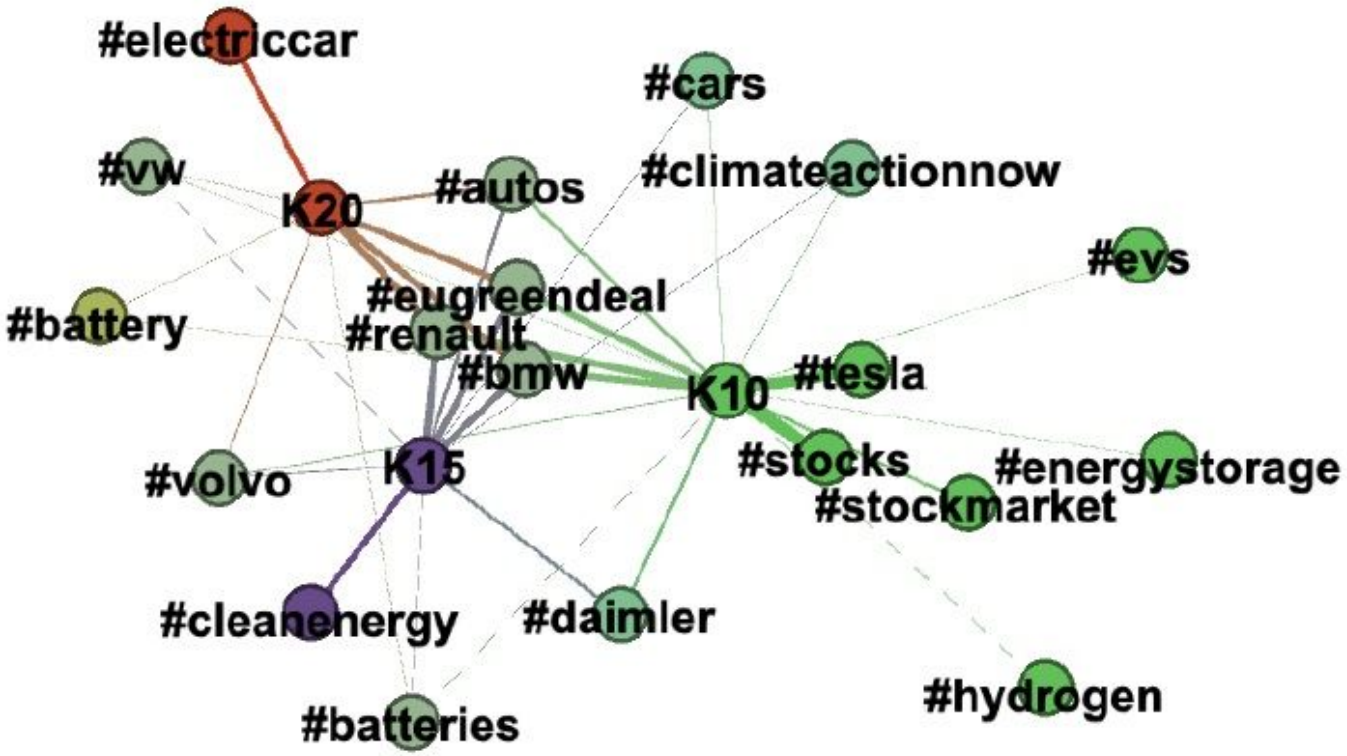

Figure 5

Resultant closely related topics for G = #electricvehicles and t = #tesla (k=10, k=15 and k=20)