Korean popular culture analytics in social media streaming: evidence from youtube channels in Thailand
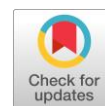
Wirapong Chansanam [a,1,*], Kulthida Tuamsuk [a,2], Kanyarat Kwiecien [a,3], Sam Oh [b,4]

[a] Department of Information Science, Faculty of Humanities and Social Sciences, Khon Kaen University, Khon Kaen, Thailand
[b] Faculty of Library and Information Science, Sungkyunkwan University, Seoul, South Korea
[1] wirach@kku.ac.th; [2] kultua@kku.ac.th; [3] kandad@kku.ac.th; [4] samoh@g.skku.edu
* corresponding author

ARTICLE INFO

ABSTRACT

This research aimed to study and analyze the influence and impact of Korean popular culture (K-pop) on Thai society. In this study, we used Social Network Analysis (SNA) to analyze streaming data obtained from a variety of YouTube channels belonging to YouTubers across the world, text analytics to analyze demographic characteristics, YouTuber's presentation techniques, as well as subscriber behavior, and multiple correlations analysis to analyze the relationship between factors affecting YouTube Channels in Thailand. The findings revealed that five Thai YouTube Channels were influencing Thai society. Furthermore, there were robust positive correlations between the number of dislikes and the number of comments (0.79), and the number of likes and comments (0.65). Additionally, there was a positive correlation between the number of views and the number of dislikes and one between likes and dislikes. Future research can supplement the present findings with other social media sources to yield an even more diverse and comprehensive analysis. These analytics can be applied to various situations, including corporate marketing strategies, political campaigns, or disease/symptom analysis in medicine. This research extends to social computing by revealing intelligent trends in social networks.

1. Introduction

A social network consists of a selection and interconnections of members [1]. In particular, culture and nature form the foundation of a social network; however, most relations in real-life social networks remain hidden. Such connections have become transparent, open, and exploitable in the web era with social networking platforms, such as Facebook, Twitter, Instagram, and YouTube [2]–[8]. Details of the network can also be collected and evaluated [9][10]. The network can be interpreted as a graph with nodes representing network actors and arrows suggesting relationships between interactors. Social Network Analysis (SNA) aims to reveal patterns within network systems and research actors to detect the effect of those patterns on individuals and organizations [11]. The focus of information scientists and experts in social networking has increasingly shifted towards enhancing people's motivation experiences through social networks for information purposes. Social networking provides opportunities to connect and share information to address problems and increase awareness, and SNA gives new insights to analyze the network structure [12]–[19].

SNA reflects on the framework of relations between various social players, such as individuals, associations, organizations, and countries, or human activity goods. SNA is also used to define the structure of social networks by analyzing the participants and explaining the network's structure based

on graph theory, algebra, statistical probability, and simulation [20][21]. SNA has been widely used in many research disciplines (e.g., economics, psychology, anthropology, biology, politics, business, and medicine) [22]–[24]. However, the issue as to whether SNA can be suited for analyzing Korean popular culture has yet to be addressed in depth. The present research employs SNA tools to examine patterns and assess user relationships in social network platforms [25]–[27].

This paper presents a set of social science SNA measurements and models and analyzes YouTube accounts concerning Korean popular culture (K-pop) networks in Thai society. First, we apply different network simulation techniques as a preliminary channel structure analysis, then we use descriptive SNA indices to analyze the overall network structure and main country location, and finally, we research the network dynamics. The paper aims to provide an outline of SNA procedures since the forms of relationships within the K-pop fandoms are too diverse, and many Thai K-pop fans have a specific understanding in the sense of the Korean Wave. More specifically, we aim to discuss the roles that influential SNA metrics play in interpreting the dynamics of the Korean Wave and patterns that cannot be obtained from 'classical' statistics and models. The reasoning behind each visualization and SNA index collection is clarified after its use.

2. Method

This study applies SNA to transform social media streaming information into knowledge. It provides insights into the critical role of social media in the distribution of alternative information. The steps of this study are: 1) Collection of data from K-pop YouTube channels between June 9 and July 22 2020; 2) Exploration of links and patterns generated in comments and replies through combined interactions on the K-pop YouTube channels; 3) Identification and Analysis of social roles and social network key players; 4) Analysis of comments on videos such as discussions, top words, pairs of words, and multiple correlations within the K-pop YouTube channels; and 5) Suggesting further action to make information distribution through social media streaming more effective.

2.1. Analysis of the relationship between all YouTube channels

We began by looking for YouTube channels that are important to Popular Korean culture. We decided to use the keyword "Korea", extracted searches from all over the world, implemented an iterative binary searching technique and ranked corresponding resources according to their search relevance. Following the investigation, we discovered that there were 1,241 network-correlated YouTube channels. We, therefore, used Gephi software for SNA. The tool used for extracting data is named YouTube Data Tools (https://tools.digitalmethods.net/netvizz/youtube/).

2.1.1. Prevalence and Patterns of Korean Popular Culture YouTube Channel

Gephi's accelerated "crawling" of the K-pop YouTube community (Data extraction) led to 5,356 channels being manipulated. The resulting data collection of 5,356 channels were "cleaned" by deleting messages not applicable to critical research mentions. There were 1,241 distinctive vertices and 5,061 edges in the network of famous channels within the K-pop YouTube community. The edges of the survey included original posts, comments, and references which were all guided. The representation of complex networks provides an understanding of the flow of form or information [28]. Gephi has been implemented from the outset to navigate complex networks quickly and intuitively. The structural design supports graphs with a structure or object that differs over time and provides a schedule portion containing a network segment. The machine tests all nodes and edges in the timeline slice's time span, which coincides with the view module, and updates it. Thus, it is possible to perform a complex network as film sequences. The dynamic module may receive network data from a graph file or external data sources. During execution, a data source will still transmit network data to the dynamic controller and then immediately display the effects in the displaying module. For example, to see the network construction over time, a web-crawler can be linked to Gephi. The architecture is interoperable and can be conveniently generated for collaboration with current applications, databases, or online servers by third parties [29]. Consequently, Table 1 includes an overall description of the case graphs.

Table 1. Overall graph metrics for the Network of K-pop YouTube Channels

Type of Graph	Directed
Overall Nodes	1,241
Overall edges	5,061
Degree of Average	4.078
Path length of Average	3.938634762140362
Epsilon	0.001
Probability	0.85
Weakly Connected Components Number	23
Strongly Connected Components Number	1,112
Iterations Number	100
Change of Sum	0.09619521979796743
Modularity	0.380
Graph density	0.003

The top five viewed Korea related YouTube Channels are: UCstEtN0pgOmCf02EdXsGChw (ABS-CBN Entertainment) 371,504,541 views, UC3ZkCd7XtUREnjjt3cyY_gg (WorkpointOfficial) 260,879,384 views, UCHkj014U2CQ2Nv0UZeYpE_A (JustinBieberVEVO) 371,504,541 views, UCp0hYYBW6IMayGgR-WeoCvQ (TheEllenShow) 206,248,898 views, UC-8Q-hLdECw QmaWNwXitYDw (KatyPerryVEVO) 194,667,225 views.

2.1.2. Performance from impacts and network analysis

This section focuses on the K-pop YouTube community's internal communication and social network scale. It also reports on the characteristics of each vertex, dependent on in-degree and out-degree, closeness, betweenness, and eigenvector centrality [30]. The SNA algorithms were developed as shown in Table 2.

2.1.3. In-Degree and Out-Degree Centrality Results

The in-degree value is the number of K-pop YouTube channels that replied to or mentioned other K-pop YouTube channels. The top three vertices had over 100 arrows pointing to them, based on the in-degree values created by the Gephi statistics. In this investigation, the top three most common networks listed from top to bottom were (1) UCs2t9DlvAuyTBkvk2djGZKw (Genierock) – an in-degree of 71; (2) UC3ZkCd7XtUREnjjt3cyY_gg (WorkpointOfficial) – an in-degree of 65; (3) UC8f7MkX4MFOOJ2SerXLInCA (one31) an in-degree of 60; and (4) UCjVb9mJZTeuPLN4hOjmvQnw (Bie The Ska) and UC4plRabXFGdAE6HP-tBQKdQ (HEARTROCKER) in-degree of 49. Genierock, WorkpointOfficial, and one31 thus seem to be the most prominent sources in this investigation. The rest of this social network's members perform various "in-between" roles.

Success is not a single indication of effect in a social media network. The important medium (out-degree centrality) was considered in this investigation. Second, there were only ten channels on the Korean YouTube Channel community that communicated specifically with UCMz oAF3aplb-2sF8DHiNzA (ซีรีย์เกาหลียอดฮิต). When the YouTube channels were rated out-of-degree, however, the top YouTube channel was UCh3HXrUrB6V0XtZbGWCFGlQ (เกาหลีเกะกะ까오 리 께까) – appearing as a citizen from the YouTube channel info. What this also means, however, is that this is a popular account, which is very outspoken and cites several others in the K-pop YouTube channel debate within the website. The out-degree of a YouTube channel measures the network's reaction to the channel. It measures the number of arrows pointing away from the channel or the number of channels it reacts to. When the authoring channel talks to others, it has the effect of putting them on the route or reaching out to them for a second time, even if they were already in the network. An account that is interested in its growth shows that interest to others.

Table 2. The list of Social Network Analysis algorithms

Algorithms	Functions

Degree centrality (in a certain relationship hierarchy)	Degree (depth, r=1, N)
	{
	  int D=0;//count node
	  while (N is linked with C){
	  D++;
	  if (C has links && r !=depth && C is not in S[]){
	  Degree (depth, r++, C);
	  }
	  }
	  S[] = (N, D);
	}

Closeness centrality	Closeness (depth, N)
	{
	  SID[nodes]=network(depth, N);
	  value=0;
	  for (int i = 1; i ≤ nodes; i++){
	  for (int j = 1; j ≤ nodes; j++){
	  value = value + Shortpath_num(SID[i], SID[j]);
	  }
	  S[i]=value/(nodes-1);
	  }
	  return S[];
	}

Betweenness centrality	Betweenness (depth, N)
	{
	  SID[nodes]=network(depth, N);
	  std = (nodes ^ 2−3 * nodes + 2)/2;
	  for (int i = 1; i ≤ nodes; i++){
	  value=0;
	  for (int j = 1; j ≤ nodes; j++){
	  for (int k = 1; k ≤ nodes; k++){
	  if (j !=k){
	  value = value + Shortpath_through(SID[j], SID[k], SID[i]);
	  }
	  }
	  }
	  S[i]=value/std;//standardized by (N²-3N + 2)/2
	  }
	  return S[];
	}

Shortest path analysis	Shortest (depth, r=1, N, Att)
	{
	  while (N is linked with C){
	  if (C ∩ Att && C is not in S[])
	  S[] = (C, r);
	  if (C has links && r !=depth)
	  Shortest (depth, r++, C, Att);
	  }
	  Sort (S[], r);
	  return S[];
	}

Data on Closeness Centrality. According to the methods in this article, closeness centrality is calculated by summing up the total distance to all nodes, and then all nodes are given a rank according to their shortest path distance. In this kind of centrality, we identify the more prominent persons with the ability to influence the network as a whole better. Thus, the centrality of closeness will help identify

good 'broadcasters' in a social network. Of the 1,241 Korean YouTube Channels culture, just 11.36 percent (141) of users scored a comparable score of 1, while 88.64 percent of the overall Korean YouTube Channels culture users scored 0. Therefore, it can be deduced in this inquiry into the K-pop YouTube community that network communication is greatly linked in a complex way.

Betweenness Depends on Centrality. Fig. 1 reflects the findings of the inquiry into the K-pop YouTube community's centrality of betweenness. This analysis demonstrates that YouTube channels in the K-pop industry play an essential role in bridging the network by following the shortest pathways and calculating how many times each vertex occurs. The Faster technique for betweenness centrality [31] was used to represent the graph-distance between all pairs of nodes for the Network Diameter, which comprised betweenness centrality, closeness centrality, and eccentricity. Information spreads on YouTube through relatively short paths. Then, those YouTube networks, on short paths, monitor the distribution of information via the social media network. Thus, YouTube channels with several short paths have high centrality of betweenness, and are considered powerful gatekeepers of information. In the case of K-pop YouTube channels, the YouTube channels with the highest betweenness centrality was that of UC5zGJZpxeZPFcds5gFcDE7Q (zbing z.) followed by UC9VmaM36k50ZLGkFL5Zh_pg (Aueyauey เอ๋ยเอ้ย) and UCQp9wDCeuLrRs-u7mC0QZVQ (Bstars Music). YouTube channels described in the topic above are ranked by the in-degree centrality. Thus, these three YouTubers can be counted as the most famous and the most prominent social network accounts in the K-Pop YouTube community.
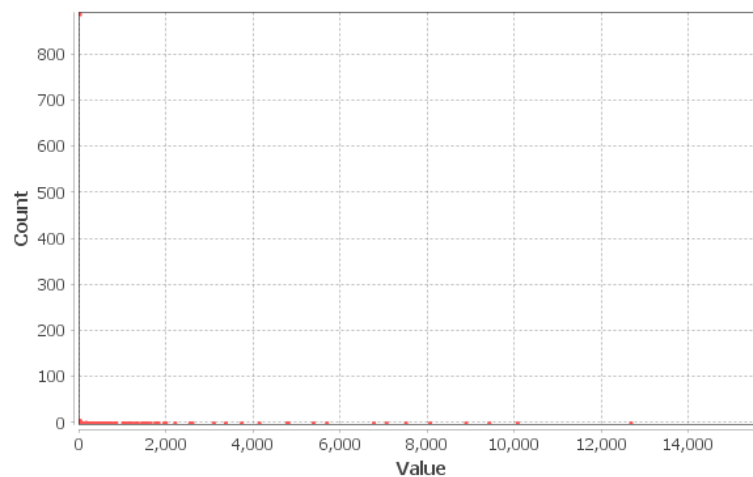

Fig. 1. Betweenness Centrality distribution

Evidence of the Centrality of Eigenvectors. Eigenvector centrality is known to be a form of the centrality of a "higher degree." A YouTube channel with fewer links may have a very high centrality of its vector with Eigenvector centrality. This means that it is more advantageous to connect to some vertices than a relation to others. However, those few connections need to be very well connected to make high variable value connections. In the investigation of K-pop YouTube channels, the highest scores for eigenvector centrality were that of the UCs2t9DlvAuyTBkvk2djGZKw (Genierock) followed by UCOmHUn--16B90oW2L6FRR3A (BLACKPINK) and UCjVb9mJZTeuPLN4hOjmvQnw (Bie The Ska), which provides adequate evidence that suggests certain Korean YouTube channels within the community is more helpful to other social network users than others.

2.1.4. Data Analytics and Visualizations

The socio-gram layout is shown in groups in Fig. 2. A given algorithm clusters the groups. The relative density of the network classifies these groups. These clusters tend to integrate vertical groups with high network capacity. These clusters may be merged. Additionally, this applies to network customers that are overly clustered and obsolete. These network members are also known as influencers for the network. These groups also promote the clustering of network users with a lower network capacity and ignore them as discrete instances that are not necessary for the simulation of clusters because they don't interact on the network with anyone. The Clauset, Newman, and Moore algorithms [32]

were used to show the relations between these vertices for this analysis and visualization. This algorithm uses modularity as a network property to set up an intercommunal network.

In the case of YouTube channels in K-pop culture, Gephi created 31 classes. The groups were arranged to show the isolates in a single group in separate boxes. Then Gephi calculated the clusters based on the criteria used in selecting the groups [33]. The resulting socio-gram (Fig. 2) shows the clusters in separate boxes with connections to particular clusters across a selection of colors. The isolates are located in different boxes at the figure's top and bottom right-hand corners. Based on its non-communication in the network, these isolates struggle to affect the overall visualization, which is also the reason why, in the figure, the relations are seen in a circular shape. Even care should be made of the coordination between the parties. From Fig. 2, with connections to several other nodes in the social network, the main clusters are clustered to the left.
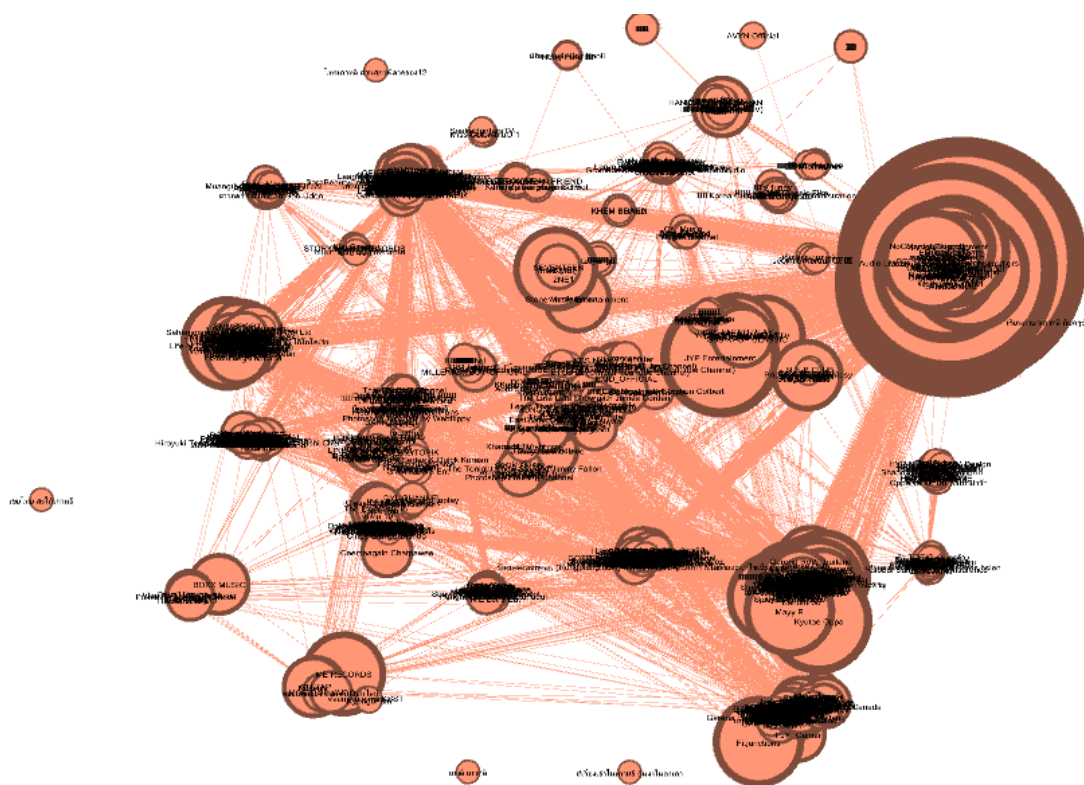

Fig. 2. Network of Large Groups of Clusters within the K-Pop Culture

It is possible to argue that the principal drawback of this analysis is that it falls short of a certain degree of minimal effects. In fact, this applies to the apparent lack of public contact in social media with the case of YouTube channels in the K-pop culture. This apparent lack of popularity on social media, especially on YouTube, may at the time have been affected both globally and locally by many other internet news events. At the same time, information overload is a reality that will not change shortly, particularly from social media. This will thus give further analysis the opportunity to examine social media users' dispersion, especially those explicitly influencing individuals, during essential cases.

2.2. A specific analysis of Korean popular culture

From the global community of K-pop YouTube channels as presented in the findings of phase 1, we selected a total of ten channels, eight of which are Thai-based, for further analysis in this section. The criteria for selecting these channels were: a) that only K-pop content was presented, b) that the channel had more than 200,000 followers or subscribers, c) that up-to-date news alerts were given. The collection and consensus results are shown in Table 3.

Table 3. Eight Korean Popular Culture YouTube Channels in Thailand

Channel ID	Label	View	Subscriber	Video
UCxoVgoUhfhFBeWeRKtFOs7w	เทพลีลา	84,255,961	99,000	303
UC7e3Krtw6AHasOuaaTAnSEw	Grace Maneerat	79,146,634	614,000	246
UCHQXqbz6NFExjXhnZQr677A	Mink Mink	58,711,530	431,000	760
UCE5NeNZxvOB3Xfk_s7Xdf8g	jaysbabyfood	35,586,714	264,000	235
UCiBqB5aji6Rm81LdpYHJ7kQ	ติ่งรีวิว รีวิวซีรีส์เกาหลี	25,017,207	216,000	168
UCGEX7Jt1kVJmBRDFoDzksVQ	AUM AT	2,667,784	36,700	190
UCuj8RdA_27pxe-hbC_sWAxQ	Pungnoeynim	423,872	16,000	15
UC2tpxtLyV8aojLb2tdOYYWQ	KhongPangKhongKwan	394,565	11,500	36

Then, we selected five Thai-based channels and sorted the remaining information according to the findings in Table 3 to further examine common issues related to K-pop and evaluate K-pop's impact on Thai society.

2.2.1. Data collection

Since our analysis indicates that the classification cannot adequately be distinguished from the subjective existence of particular questions, our approach relies on the discovery – and time-consuming variance of search results – for a series of chosen questions. We utilized the YouTube Data Tool 'Video List' module, which uses the 'search: list' endpoint of the YouTube API v3.4, to use the 'order' parameter on the Web platform and smartphone apps for user-focused search and hence our analysis explored the basic rating process. The capture module was then immediately called on for the top 300 results from June 9 to July 22, 2020, for 44 days simultaneously. The complexities arising from this unique configuration are further elaborated in this paper's subsequent sections. In the absence of a simple methodological equilibrium, the researchers must creatively handle the theoretical possibilities and limitations. We understand that location and personalization on YouTube will play an important role, but we keep the API data as close as possible to a 'baseline' point of view.

Our data collection method was used in the manual selection of the following five YouTube ID queries: [UCxoVgoUhfhFBeWeRKtFOs7w], [UC7e3Krtw6AHasOuaaTAnSEw], [UCHQXqbz6N FExjXhnZQr677A], [UCE5NeNZxvOB3Xfk_s7Xdf8g], and [UCiBqB5aji6Rm81LdpYHJ7kQ]. This collection reflects on current and contentious topics, of which it is possible to say that the search results are linked to politically motivated controversies and that they are discussed. This preference is supported by the practice of depicting conflicts in 'situations where actors differ, "and the development of social life can be seen through a high-contrast lens [34]. While extending our approach to less-heated YouTube parts such as interviews would be important, we were primarily searching for questions where we could fairly predict substantial improvement over time in response to recent incidents. Except for Gamergate, which had settled before we started gathering data, all words suggested ongoing disputes which have been perforated and deeply divisive by current events, which made an overview of the grade shift especially prominent. Our selection has also mirrored the three authors' current experience and allowed us to address the data with more trust. We understand that our ad hoc inquiries do not encourage us to make general statements about YouTube cultures but instead help us develop and test a technique that we intend to apply in future research systematically.

We also focused on Google Trends data, a database that offers statistics on the search volume for four Google sites over a predefined period to contextualize the rankings we also collected with a metric of public interest: site search, product search, news, and YouTube. If the given data is not true search figures but averaged to the maximum value for the period, the variance measurement is endorsed over time. In each of our case studies, we were able to examine search volume variations on YouTube.

For the ethical component of our research, we did not want to anonymize channel identities, either in our fundamental data or for publishing purposes. Contrary to YouTube, the use of real names is not mandated; most notably, the questions we selected did not extend to private spaces where privacy

standards pose practical issues but to networks with multiple subscribers addressing a broader audience. While YouTube maintains social networking sites elements, our queries highlighted a vast array of networks that subscribe to a broadcasting framework that is in several respects reminiscent of speech radio. The public interest in understanding this new media environment is essential.

Data is collected from the YouTube API for this study. A script is scheduled for running every day during the year 2020 and retrieves data from YouTube API about the popular videos of the day. The script will then process and archive the data obtained from YouTube in text files.

In this article, Thailand will examine trend video clips, which contain a total of 72,994 videos. YouTube usually adds 200 videos into the trend list every day to have 73,000 (365 x 200) videos. We have obtained 72,994, perhaps because the number of popular videos was marginally less than 200 on a couple of days. On YouTube, for several days, the same video could appear on the trend list. We, therefore, have data to examine on 72,994 trend videos. The 72,994 videos are also not special ones. We currently have 11,177 exclusive videos out of the 72,994 videos. In other words, 11,177 videos were featured in the trend list all year long. Some of them could have been on the registry for ten days, others 20 days, etc. We can see more detail on this later in the report.

2.2.2. Data analysis

Python and the great community of Python libraries, including Pandas, Matplotlib, NLTK, ImageAI, WordCloud, and more, performed the data analysis of the present study. The research was carried out in a Jupyter Notebook. Comments posted by followers on video clips in the five YouTube channels were reviewed. The results of the analysis of the five channels are shown below in Table 4.

Table 4. Basic statistical results

Statistic	CategoryID	ViewCount	LikeCount	DislikeCount	CommentCount
count	1442.00	1442.00	1437.00	1437.00	1023.00
Mean	22.99	189770.06	4035.49	125.92	370.17
std	1.29	245252.72	4991.22	351.03	675.10
min	1.00	0.00	5.00	0.00	0.00
25%	22.00	43863.50	995.00	17.00	104.00
50%	23.00	110382.50	2464.00	41.00	240.00
75%	24.00	236057.00	5009.00	116.00	439.00
max	26.00	2442916.00	54430.00	7629.00	16255.00

The mean average for the number of views is 110,382.50. This means that one-half of the trend videos hold less views than this figure, while the views for the other half exceed this figure. The average number of people who like a trending video is 4,035.49, while the average number of dislikes is 125.92. The average number of comments is 370.17, while the low is 240.00. We can see that the percentage of videos with less than 390 views is approximately 71 percent, and the number of videos with less than 5 million views is roughly 91 percent.

Currently, we can see that the bulk of pattern videos are 2000 or more with a video peak of 120 or less. Similarly, we can see that the percentage of videos with less than 2,000 likes is around 44.33%. We shall verify that by plotting a scatter plot between Viewscount and Likescount to visualize the relationship between these variables.

We see that Viewscount and Likescount (Fig. 3) are truly positively correlated: as one increases, the other increases too—mostly. We see that most trending videos have around 2000/7 = 286 comments since each division in the graph has seven histogram bins. Similarly, we can see that the rate of videos with less than 286 comments is around 57.48%.
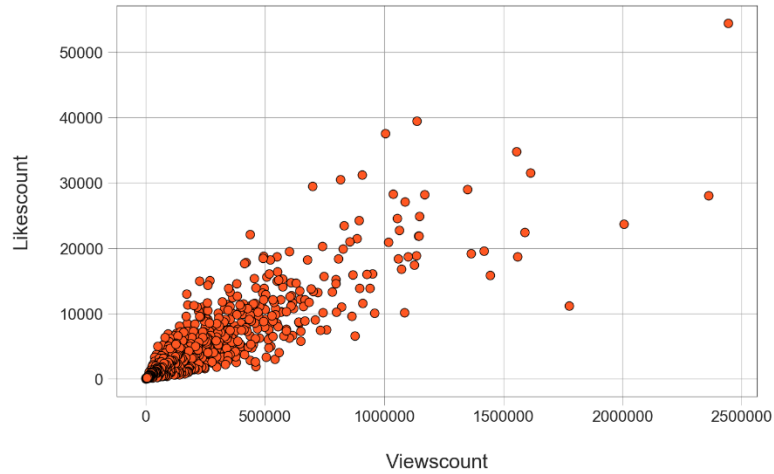
Fig. 3. The scatter plot between Viewscount and Likescount relationship

2.2.3. Video title lengths

We shall now add another column to our dataset to represent the length of each video title, then plot the histogram of title length to get an idea about the lengths of trending video titles. We see that title-length distribution resembles a standard distribution where most videos have title lengths of approximately 40 to 65 characters. Then, a dispersion plot is drawn between the title duration and the number of views to see the relation between these two variables. Looking at the dispersion graph, we may conclude that there is no relation between the title duration and the number of views. However, we note an odd trend: videos with 1,000,000 views and more have a title length from about 30 to 70 characters.

2.2.4. Analysis of follower comments in five YouTube channels

As shown in Table 5, the highest number of replies to comments was 199. The comment with the most replies was about a contest hosted by Busted!, which offered incentives in the form of prizes to engage people in the conversation.

Table 5. Comments reply analysis result

ID	CommentTextDisplay	CommentTotalReplyCount
24294	มาชิงของฝากจาก Busted! กันจ้าาา มีเข็มกลัด 2 ร...	199
9487	น้องไม่เหมาะเป็นยูทูปเบออ่า พูดไม่ค่อยน่าฟังเท...	180
38621	คือเราก็เป็นผู้หญิงนะ แต่กูก็รู้พร้อมๆกับพี่เข...	103
211	ใครไม่เคยไปเที่ยวต่างประเทศทางนี้ 🐸 🐸 🐸	83
20946	ติ่งbtsมาแล้วจ้า. ขอเสียงโน้ยยย	75

Table 6 illustrates that the comment that garnered the highest number of likes at 9,309 was about women.

Table 6. Text comments analysis result

ID	CommentTextDisplay	CommentLikeCount
38621	คือเราก็เป็นผู้หญิงนะ แต่กูก็รู้พร้อมๆกับพี่เข...	9,309
39724	ช่องนี้น่ากลัวขึ้นทุกวันละ5555	5,904
25793	ใครอยากจัดฟันไลค์	5,279
6390	ชอบความประชดพี่เกรซอะ 5555+😂😂😂😂	5,170
27498	น่ารักกก 555	5,019

Fig. 4 shows that, in general, followers of five K-pop YouTube channels comment on most video clips in a pleasing, amusing, and happy manner. If they are teenagers, commenters use words like 'cute', 'funny', or 'fun', as well as expressions that indicate laughter (e.g., 555 or lol) to show that something is amusing or pleasing to them. Moreover, if they are female, commenters use words like "elder" (i.e., the word 'พี่' to refer to someone older than them) and "woman" next to other positive words or descriptors like "beautiful" so as to express their positive attitudes.

Fig. 4. The word cloud of five K-pop YouTube channels comments on video clips

3. Results and Discussion

In the year 2020, a channel by the name of Mink Mink created the most popular videos on YouTube. Content related to product reviews, restaurants, accommodation, tourist attractions, as well as personal vlogs rather than corporate ones, receive the most popularity in the K-pop YouTube community. Furthermore, after Mink Mink, the 2nd and 3rd channels with the most popular videos were entertainment channels. While a video may only be on the trending list for a day or two, six videos were on the list for a total of 44 days. Thep Lee La received the most views, subscribers, likes, and comments. It was, nevertheless, the fourth most hated video. Even when some videos only received approximately 50,000 views, they were included on the trending list, as shown in Fig. 5.

Fig. 5. The graph of videos in five YouTube channels

In popular video titles, words like "official," "video," "2020," "vs.," "trailer," "music," "game," "new," "highlights," "first," and "challenge" were among the most prevalent. This perhaps suggests that including the current year in the title of a YouTube video will increase its chances of trending. In popular

video titles, the use of emojis were most common. Most popular videos have names that are between 36 and 64 characters long, with a minimum of three and a maximum of 100. Even when there were just seven comments on a video, it still became popular. The most hated YouTube Rewind videos of 2019 and 2020 were the first and second most disliked videos, respectively. Videos take an average of 1.5 days to reach the top of the trending list for the first time. While 5,000 characters is the maximum permitted, most popular videos have descriptions that are between 500 and 1500 characters long. Many popular videos include social network links in their captions. Tags are used in almost all popular videos. On average, there are 21 tags. An object-detection algorithm was applied to the thumbnails of popular movies, and it was discovered that a human is the most prevalent item in the thumbnails. As a result, one might choose to include a personal photo in their next video.

As shown in Fig. 6, videos that fall into the entertainment category (approximately 710 videos) were the trendiest videos among other categories, followed by people & blogs with some 700 videos and satire with some 200 videos, and so forth. Fig. 7 demonstrates that significantly fewer circulating videos are released on Saturdays and Sundays than on other days of the week. Except for Monday, the number of releases on other days of the week is not significantly different from Saturday's. On Monday, only 140 trend videos were released.
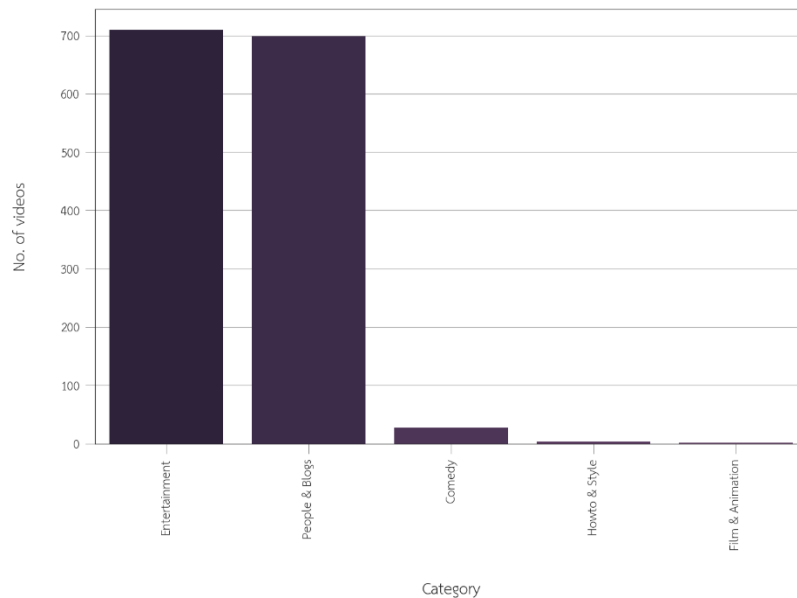
Fig. 6. The categories of videos in five YouTube channels
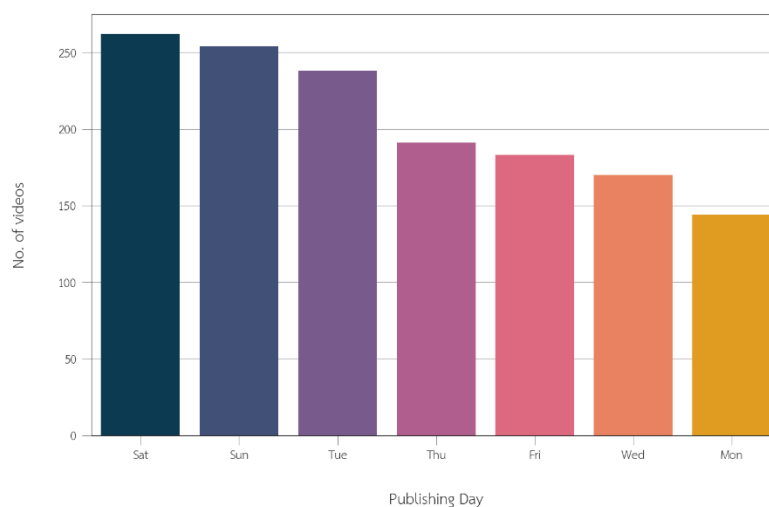
Fig. 7. The release date of videos

The graph shown in Fig. 8 displays the number of trending videos released each day from 12 am to 11 pm, beginning at 0 (00:00 hrs.). We can see that popular videos are mostly published during the period between 9 am to 1 pm. We also observe that the least popular videos are published between 12 am and 8 am and between 2 pm and 11 am. Some possible explanations behind this finding may well be due to YouTube's trending video algorithm or it may be due to the fact that far more videos are released between 2 pm and 7 pm. The two above graphs do not suggest Saturday's videos are more likely to turn into a trend because Saturday may be the day when videos were more frequently posted than on other days. To verify those statements, we also need to know how many videos (not just trending ones) have been released every day of the week in the year 2020. The same holds true for hours of the day.
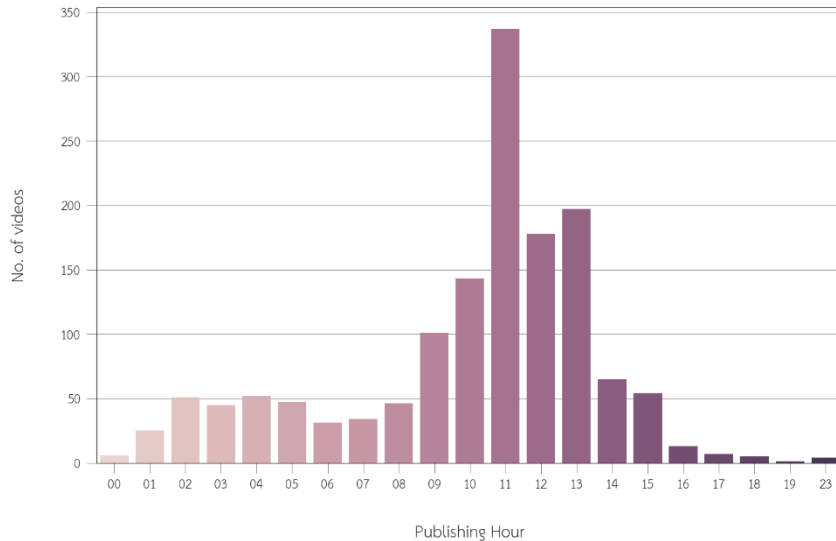
Fig. 8. The time period of publishing videos

Fig. 9 shows that YouTubers prefer to include their channel's names in the title and descriptions of their videos followed by the content of the video clip they want to feature or highlight. In descending order are the words 'Mink' (348 occurrences), 'เทพลีลา' (257 occurrences), 'jaysbabyfood' (229 occurrences), 'Maneerat' (200 occurrences), 'Grace' (195 occurrences), and 'ติ่งรีวิว' (140 occurrences), which all represent the name of the channel.

Fig. 9. The Word Cloud of YouTubers' preferred words to add to their videos and descriptions

We shall now compare the video stats mentioned in the previous sections. Views with Likes is directly correlated to view count, and the length of the title is in line with the number of views. A heat map visualization has been created to show the relationships between different measurements. Fig. 10 shows a correlation map, and the correlation matrix shows that views and attitudes are strongly correlated. Lighter shades show a higher positive correlation in this heat map, and darker shades suggest a weaker positive correlation. The highest positive correlation possible is 1, and the highest negative correlation (the lowest positive correlation) is -1.
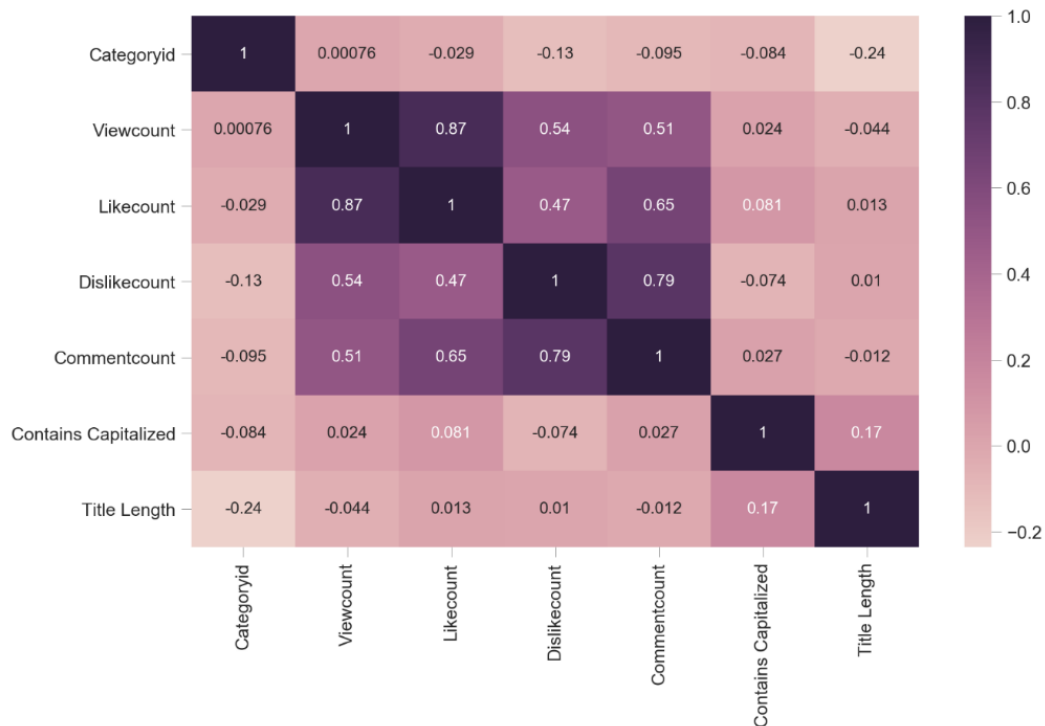
Fig. 10. The YouTube channels correlation matrix

We can see that the correlation (0.87) between views and likes is strongly optimistic. This means that the two begin to rise concurrently in trending YouTube channels. A strong positive correlation can also be shown (0.79) between Disklikecount and Commentcount, as well as between Likecount and Commentcount (0.65). In addition, there is a positive correlation between Viewscount and Dislikescount, and one between Likescount and Dislikescount.

4. Conclusion

When we looked at ways to explain YouTube's "relevance" algorithm, we discovered that simple variables like screen size, comment counts, and search results had few clear connections. There is no reason to believe that a more detailed multivariate approach would result in better relations, since (a) the single YouTube variable used to define a factor directly — time to view — is not accessible from the API, and (b) query-dependent statistic variables like 'videos users who are looking for a word' cannot be replicated from the outside. While it would be worth exploring the tremendous functional information gathered by content producers, our highly straightforward approach provided various fascinating findings considering that apparent causal factors remain unchanged. In light of the Thai Government's growing interest in social impact work and the region's economic growth, we were inspired to map the field's academic landscape and analytically examine its development. We utilized SNA, data science, and social science tools to differentiate the development and essence of the K-pop industry in terms of social media platform, sharing, comments, replies, actions, respective collaboration features, and analysis focus areas. The SNA approach helps disclose possible knowledge focused mostly on title video, comment reply, and YouTube channels' presentation, but this does not provide a comprehensive understanding of the complex patterns of growth and interaction in the rapidly expanding entertainment sector. In future studies, we expect to add more data sources and variables, such as Facebook, Instagram, Twitter, and qualitative and quantitative methodologies, to provide a more detailed study of Korean Popular Culture.

Declarations

Author contribution. All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.

Funding statement. This research was funded by the Faculty of Humanities and Social Sciences, Khon Kaen University, Thailand.

Conflict of interest. The authors declare no conflict of interest.

Additional information. No additional information is available for this paper.

References

[1] Z. Xu, Y. Li, and L. Hao, "An empirical examination of UTAUT model and social network analysis," *Libr. Hi Tech*, vol. ahead-of-p, no. ahead-of-print, Dec. 2019, doi: 10.1108/LHT-11-2018-0175.

[2] A. S. Aribowo, H. Basiron, N. S. Herman, and S. Khomsah, "An Evaluation of Preprocessing Steps and Tree-based Ensemble Machine Learning for Analysing Sentiment on Indonesian YouTube Comments," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 5, pp. 7078–7086, Oct. 2020, doi: 10.30534/ijatcse/2020/29952020.

[3] H. H and M. Rafi, "BlogNewsRank: Finding and Ranking Frequent News Topics Using Social Media Factors," *JOIV Int. J. Informatics Vis.*, vol. 2, no. 3, p. 166, May 2018, doi: 10.30630/joiv.2.3.134.

[4] J. Savigny and A. Purwarianti, "Emotion classification on youtube comments using word embedding," in *2017 International Conference on Advanced Informatics, Concepts, Theory, and Applications (ICAICTA)*, 2017, pp. 1–5, doi: 10.1109/ICAICTA.2017.8090986.

[5] R. Novendri, A. S. Callista, D. N. Pratama, and C. E. Puspita, "Sentiment Analysis of YouTube Movie Trailer Comments Using Naïve Bayes," *Bull. Comput. Sci. Electr. Eng.*, vol. 1, no. 1, pp. 26–32, Jun. 2020, doi: 10.25008/bcsee.v1i1.5.

[6] A. S. Aribowo, H. Basiron, N. F. A. Yusof, and S. Khomsah, "Cross-domain sentiment analysis model on Indonesian YouTube comment," *Int. J. Adv. Intell. Informatics*, vol. 7, no. 1, p. 12, Mar. 2021, doi: 10.26555/ijain.v7i1.554.

[7] F. I. Tanesab, I. Sembiring, and H. D. Purnomo, "Sentiment Analysis Model Based On Youtube Comment Using Support Vector Machine," *Int. J. Comput. Sci. Softw. Eng.*, vol. 6, no. 8, pp. 180–185, 2017, available at : http://ijcsse.org/published/volume6/issue8/p2-V6I8.pdf.

[8] A. K. Mohamad, "Employ Twitter Data to Perform Sentiment Analysis in the Malay Language," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 2, pp. 1404–1412, Apr. 2020, doi: 10.30534/ijatcse/2020/76922020.

[9] I. M. Nadhiroh, R. Hardiyati, M. Amelia, and T. Handayani, "Mathematics and statistics related studies in Indonesia using co-authorship network analysis," *Int. J. Adv. Intell. Informatics*, vol. 4, no. 2, p. 142, Jul. 2018, doi: 10.26555/ijain.v4i2.120.

[10] C. Murendo, M. Wollni, A. De Brauw, and N. Mugabi, "Social Network Effects on Mobile Money Adoption in Uganda," *J. Dev. Stud.*, vol. 54, no. 2, pp. 327–342, Feb. 2018, doi: 10.1080/00220388.2017.1296569.

[11] A. Martínez, Y. Dimitriadis, B. Rubia, E. Gómez, and P. de la Fuente, "Combining qualitative evaluation and social network analysis for the study of classroom social interactions," *Comput. Educ.*, vol. 41, no. 4, pp. 353–368, Dec. 2003, doi: 10.1016/j.compedu.2003.06.001.

[12] K. J. Patel and H. J. Patel, "Adoption of internet banking services in Gujarat," *Int. J. Bank Mark.*, vol. 36, no. 1, pp. 147–169, Feb. 2018, doi: 10.1108/IJBM-08-2016-0104.

[13] G. Yi, N. M. M. Zainuddin, and N. A. B. A. Bakar, "Conceptual Model on Internet Banking Acceptance in China with Social Network Influence," *JOIV Int. J. Informatics Vis.*, vol. 5, no. 2, May 2021, doi: 10.30630/joiv.5.2.403.

[14] M. Clark, "Converting purchase commitments into purchase fulfillments: An examination of salesperson characteristics and influence tactics," *Ind. Mark. Manag.*, vol. 85, pp. 97–109, Feb. 2020, doi: 10.1016/j.indmarman.2019.09.002.

[15] R. Safeena, A. Kammani, and H. Date, "Exploratory Study of Internet Banking Technology Adoption," 2018, pp. 333–355, doi: 10.4018/978-1-5225-5201-7.ch015.

[16] D. I. Calibo and J. D. Niguidula, "Metadata Extraction Analysis: A Review of Video Data in Effect to Social Media Compression," *JOIV Int. J. Informatics Vis.*, vol. 3, no. 1, Jan. 2019, doi: 10.30630/joiv.3.1.216.

[17] X. Hu, X. Chen, and R. M. Davison, "Social Support, Source Credibility, Social Influence, and Impulsive Purchase Behavior in Social Commerce," *Int. J. Electron. Commer.*, vol. 23, no. 3, pp. 297–327, Jul. 2019, doi: 10.1080/10864415.2019.1619905.

[18] J. H. J. Min, H. J. J. Chang, T.-M. C. Jai, and M. Ziegler, "The effects of celebrity-brand congruence and publicity on consumer attitudes and buying behavior," *Fash. Text.*, vol. 6, no. 1, p. 10, Dec. 2019, doi: 10.1186/s40691-018-0159-8.

[19] S. Kim, J. Kandampully, and A. Bilgihan, "The influence of eWOM communications: An application of online social network framework," *Comput. Human Behav.*, vol. 80, pp. 243–254, Mar. 2018, doi: 10.1016/j.chb.2017.11.015.

[20] J. Galaskiewicz and S. Wasserman, "Social Network Analysis," *Sociol. Methods Res.*, vol. 22, no. 1, pp. 3–22, Aug. 1993, doi: 10.1177/0049124193022001001.

[21] S. Wasserman and K. Faust, *Social Network Analysis*, 1994, doi: 10.1017/CBO9780511815478.

[22] D. Chawla and H. Joshi, "The Moderating Effect of Demographic Variables on Mobile Banking Adoption: An Empirical Investigation," *Glob. Bus. Rev.*, vol. 19, no. 3_suppl, pp. S90–S113, Jun. 2018, doi: 10.1177/0972150918757883.

[23] S. K. Abbas, H. A. Hassan, J. Asif, H. M. Junaid, and F. Zainab, "What are the key determinants of mobile banking Adoption in Pakistan?," *Int. J. Sci. Eng. Res.*, vol. 9, no. 2, pp. 841–848, Feb. 2018, doi: 10.14299/ijser.2018.02.012.

[24] M. U. Hassan, A. Iqbal, and Z. Iqbal, "Factors affecting the adoption of internet banking in Pakistan: an integration of technology acceptance model and theory of planned behaviour," *Int. J. Bus. Inf. Syst.*, vol. 28, no. 3, p. 342, 2018, doi: 10.1504/IJBIS.2018.092530.

[25] L. B. Shyamasundar and P. Jhansi Rani, "A Multiple-Layer Machine Learning Architecture for Improved Accuracy in Sentiment Analysis," *Comput. J.*, vol. 63, no. 3, pp. 395–409, Mar. 2020, doi: 10.1093/comjnl/bxz038.

[26] F. H. Khan, U. Qamar, and S. Bashir, "Enhanced cross-domain sentiment classification utilizing a multi-source transfer learning approach," *Soft Comput.*, vol. 23, no. 14, pp. 5431–5442, Jul. 2019, doi: 10.1007/s00500-018-3187-9.

[27] A. A. Alalwan, A. M. Baabdullah, N. P. Rana, K. Tamilmani, and Y. K. Dwivedi, "Examining adoption of mobile internet in Saudi Arabia: Extending TAM with perceived enjoyment, innovativeness and trust," *Technol. Soc.*, vol. 55, pp. 100–110, Nov. 2018, doi: 10.1016/j.techsoc.2018.06.007.

[28] D. L. Moody, "Theoretical and practical issues in evaluating the quality of conceptual models: current state and future directions," *Data Knowl. Eng.*, vol. 55, no. 3, pp. 243–276, Dec. 2005, doi: 10.1016/j.datak.2004.12.005.

[29] D. L. Hansen *et al.*, "Do You Know the Way to SNA?: A Process Model for Analyzing and Visualizing Social Media Network Data," in *2012 International Conference on Social Informatics*, 2012, pp. 304–313, doi: 10.1109/SocialInformatics.2012.26.

[30] L. C. Freeman, "Centrality in social networks conceptual clarification," *Soc. Networks*, vol. 1, no. 3, pp. 215–239, Jan. 1978, doi: 10.1016/0378-8733(78)90021-7.

[31] U. Brandes, "A faster algorithm for betweenness centrality*," *J. Math. Sociol.*, vol. 25, no. 2, pp. 163–177, Jun. 2001, doi: h10.1080/0022250X.2001.9990249.

[32] A. Clauset, M. E. J. Newman, and C. Moore, "Finding community structure in very large networks," *Phys. Rev. E*, vol. 70, no. 6, p. 066111, Dec. 2004, doi: 10.1103/PhysRevE.70.066111.

[33] C. Udanor, S. Aneke, and B. O. Ogbuokiri, "Determining social media impact on the politics of developing countries using social network analytics," *Program*, vol. 50, no. 4, pp. 481–507, Sep. 2016, doi: 10.1108/PROG-02-2016-0011.

[34] T. Venturini, "Diving in magma: how to explore controversies with actor-network theory," *Public Underst. Sci.*, vol. 19, no. 3, pp. 258–273, May 2010, doi: 10.1177/0963662509102694.