In Quan-Haase A & Sloan L (eds.) *The SAGE Handbook of Social Media Research Methods*, 2nd ed. London: SAGE, pp. 487-502.
Available at: https://uk.sagepub.com/en-gb/eur/the-sage-handbook-of-social-media-research-methods/book272098

Chapter 34: NodeXL - Twitter social media network insights in just a few clicks

Wasim Ahmed[1], Harald Meier[2], Marc Smith[2]

[1]Stirling University Business School, [2] Social Media Research Foundation

Dr. Wasim Ahmed is a Senior Lecturer in Digital Business. His research and teaching interests include digital business, digital sport, and digital health. He has previously worked at the University of Sheffield, Newcastle University, and Northumbria University.

Mr. Harald Meier is a geographer studying the landscape of social media. He leads the development of the NodeXL Pro and INSIGHTS reports for Twitter.

Dr. Marc Smith is a sociologist and Director of the Social Media Research Foundation. His research focuses on the mapping of social media networks to identify the structure and dynamics of computer-mediated collective action.

Abstract

NodeXL enables the collection, analysis, visualization, and reporting on collections of annotated connections. NodeXL network data sets can be imported from a range of data sources and processed into insightful reports that can highlight the overall shape of a network, the main divisions or clusters within it, and the leading voices at the center of each cluster. NodeXL output can be used to inform analysis of complex relationships, often drawn from social media platforms. By removing the need to master programming skills to perform social media network analysis, NodeXL is intended to expand access to the essential data needed for critical empirical studies of computer mediated collective action. This chapter provides an introduction on how to retrieve data using NodeXL and a guide to incorporate findings in academic research and practice. We illustrate these findings with a study of the tweets related to medical condition discussions and the Russell Group universities in the UK. NodeXL seeks to simplify the mechanical process of data collection, analysis, visualization and reporting so that researchers can focus on the meaning and implications of images and data rather than on the process of creating the image and data.

Keywords:

Social Media, Social Networks, SNA, Network Analysis, Data Visualization, Collective Action, Computational Social Science, Digital Sociology, Big Data, Influencers

Introduction

There are a number of popular automated and manual textual approaches to analyzing social media data as well as various visualization packages (Ahmed, 2017). This chapter outlines a popular and widely cited academic tool which is a Microsoft Excel plugin known as the *Network Overview, Discovery and Exploration for Excel* and also abbreviated as *NodeXL*.

NodeXL is designed to enable the collection, analysis, and generation of insightful reports about patterns found in collections of annotated connections. Networks are present in a wide range of natural and social phenomena. We can think of social media platforms as networks with users as nodes, the links between them as edges, and the messages or images exchanged are the annotations on these connections. The World Wide Web or WWW can also be thought of as a large network where pages are nodes and the Uniform Resource Locators (URLs) that link from one page to another are the edges. The Internet itself is also a network where nodes are computers and edges are physical connections between these devices.

Social media is a common focus for social network analysis. Social media, by its nature, is always composed of a social network that emerges as people interact with one another. As people reply, mention, like, favorite, and forward messages they are creating connections among users. These connections, in aggregate, form complex network structures.

Until recently, collecting and analyzing network data often required advanced software development skills. Built into the context of the familiar Excel spreadsheet application, NodeXL has made social network analysis accessible to all and as easy as making a pie chart. NodeXL Basic and NodeXL Pro are add-ins for Microsoft® Excel® (2013, 2016, 2019) that support social network and content analysis (see Fenton and Parry, Chapter X, this volume). NodeXL Basic is available freely and openly to all. It is positioned as a browser for files created with NodeXL Pro which offers advanced features for professional social network and content analysis. By using social data importers, it is possible to easily import data from social media platforms such as Twitter, YouTube, Flickr, and Wikipedia.

A recent review of the uses of NodeXL within academic research found that it had been utilized across a wide variety of research areas and topics such as to study social media content related to natural disasters, public health, political movements, and gun control among others (Ahmed and Lugovic, 2019). However, the review also noted that NodeXL had been utilized outside the area of social media research, for instance, to study blackboard discussion boards, internal security, and email-networks among others. This is because it is possible to import any data which has network relationships into NodeXL for analysis. Network data can be drawn from pre-digital materials as well, for example representing the connections among people mentioned in ancient Greek texts (Cline, 2012; Huner and Suárez, Chapter X, this volume).

Social network analysis is based on a set of interrelated concepts (Scott, 1990; Wasserman and Faust, 1994). Networks can be understood to have four levels or resolutions in which they can be viewed: 1) the individual connection or edge, 2) the individual vertex, 3) groups of vertices, and 4) the network or graph as a whole. Data can be reported at each of these resolutions: the edge, the vertex, the group, and the graph. Studying network groups can help to uncover the frequency and popularity of discussions and the opinion leaders within these discussions. Networks are composed of "nodes" or "vertices" that represent the people or entities in a population. These vertices can connect to one another, forming what are called "Edges". Each "edge" represents the connection between two vertices. Each edge often contains the text of a message or other annotation. Collections of edges form networks that have structures and patterns that can be measured and divided into clusters or regions that are often called “groups". Each vertex has properties like its “degree”, the number of connections it has, and its “centrality”, a measure of the uniqueness of those connections.

We demonstrate these levels of analysis via a study of tweets related to diabetes and to a number of Russell Group Universities in the United Kingdom. A key goal of the chapter is to illustrate the steps needed for data to be imported, cleaned, and analyzed. This chapter demonstrates the ways NodeXL simplifies the mechanical process of data collection, analysis, visualization and reporting so that researchers can focus on the meaning and implication of images and data rather than on the process of creating the graphs and data.

NodeXL Installation

NodeXL Basic and Pro versions can be downloaded from the NodeXL Graph Gallery web site via the following link ((NodeXL Graph Gallery, n.d.) and by completing a short form. NodeXL Basic allows users to browse and explore NodeXL social media network graphs; whereas NodeXL Pro (license required) also provides access to social media network data streams, advanced network metrics, time series, and text and sentiment analysis. Readers are encouraged to download and install NodeXL Pro to follow along with the following examples. After installation, a new menu tab appears in Microsoft Excel entitled ‘NodeXL Pro’ (Figure 1). While NodeXL can handle data from multiple social media platforms, this chapter focuses on NodeXL’s Twitter capability as a way of demonstrating how the software works.

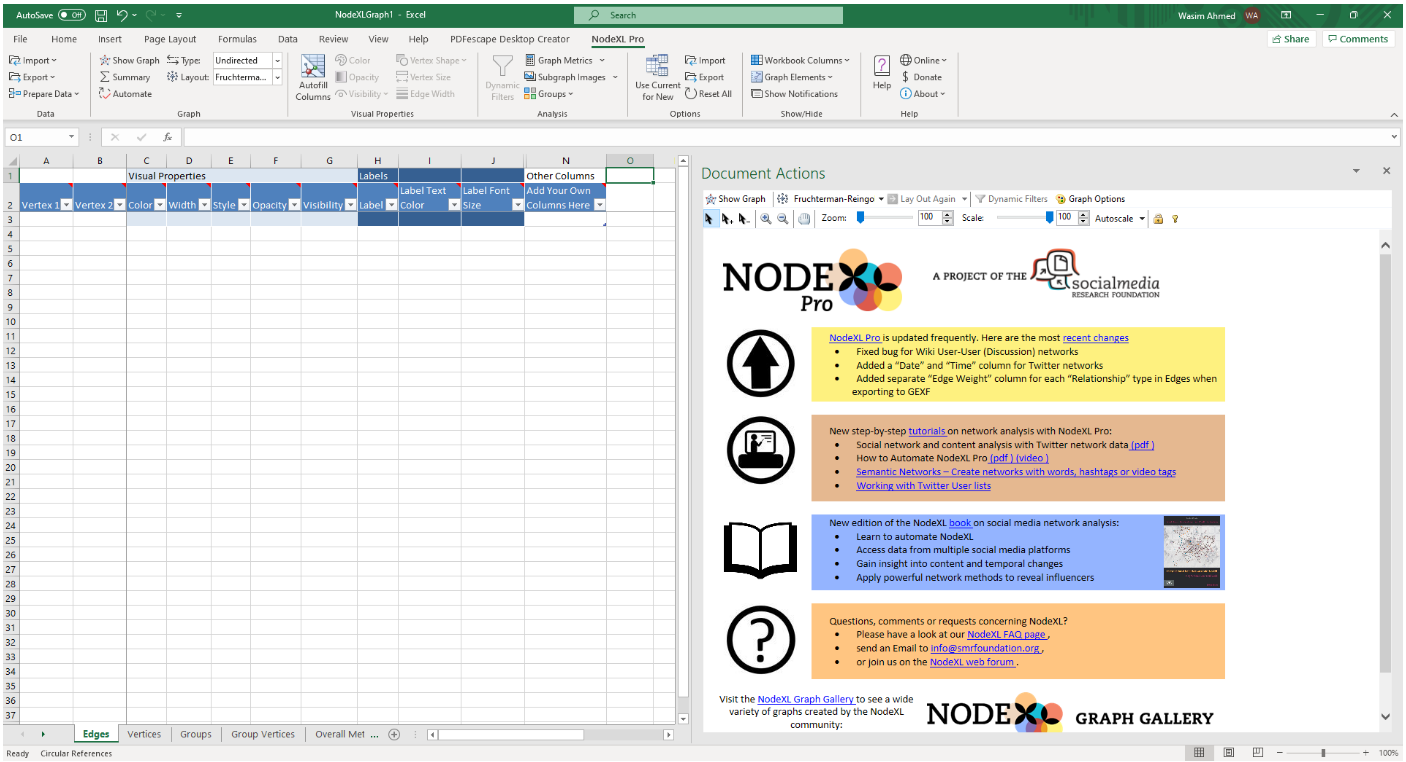


Figure 1. Overview of NodeXL Pro within Microsoft Excel

Importing data via the NodeXL Pro Twitter Search Importer

One of the main features of NodeXL Pro is the ability to easily import data from a variety of external social media network data sources. In Figure 2 we show the NodeXL interface showing how data can be imported into NodeXL Pro manually, from existing network datasets (e.g., social media network data) and/or via the use of third-party importers. Social media can be downloaded by searching for specific words or phrases. For instance, if we search for the word 'diabetes' Twitter will retrieve and return a set of tweets which contain this word. It is also possible to download data by using hashtags such as '#diabetes' or by searching for a user's account name (Twitter "handle") such as '@DiabetesUK'. Users can access data using a combination of keywords, hashtags and Twitter accounts to retrieve data such as '#diabetes' AND '@DiabetesUK', collecting tweets with both terms. It is also possible to exclude certain keywords that may be spammed or to download only tweets in a certain language. It is strongly recommended that readers design their search queries with help of Twitter's rich set of standard search operators (Search Operators, n.d.).

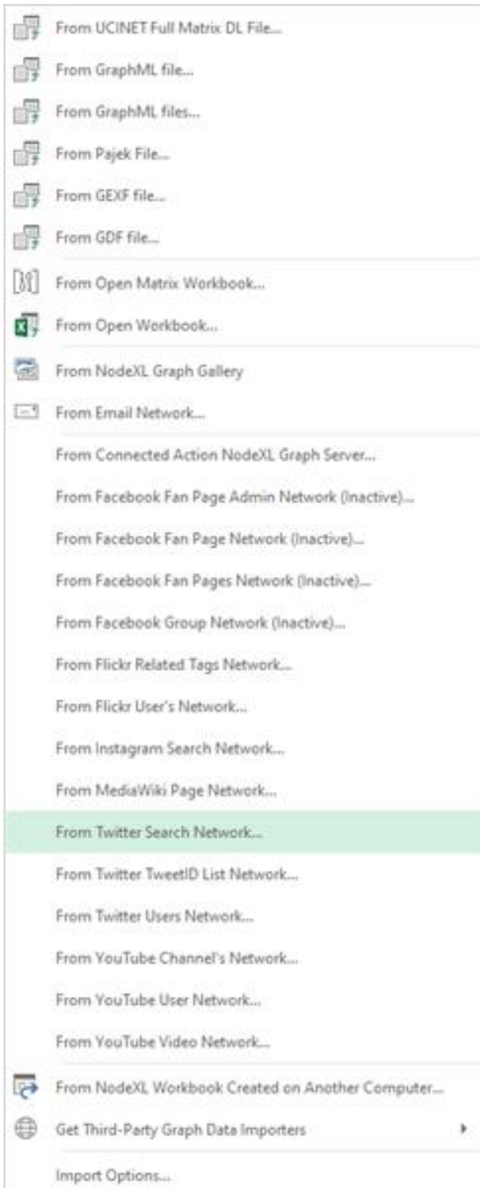


Figure 2. NodeXL import Options

As shown in Figure 2, network data can be imported from a range of network data sources and file formats.

Importing data via NodeXL Pro Twitter Search Importer

NodeXL enables easy import of network data from Twitter (Figure 3). The *Twitter Search Network* importer allows researchers to search Twitter and retrieve data related to the keyword(s) chosen. When retrieving data via the 'Search' Twitter Application Processing Interface (API), the maximum accessible amount of data is 18,000 tweets and retweets per search, approximately 7-10 days prior to the search date. Moreover, the data volume is constrained for trending topics that have a large volume of messages. Most Twitter researchers use the Search API, but Twitter also provides a costly commercial 'Firehose' API service which permits access to 100% of tweets, including historical tweets. This service is often priced beyond the reach of many research budgets.

Recently, Twitter has expanded access for academic researchers via its "Academic Twitter" program ((Twitter API for Academic Research, n.d.). This new service allows academics to apply and potentially be granted much larger volumes of data and full access to the historical archive of all tweets. This is a major development for social media scholarship, and we encourage researchers to apply for this new level of access. If granted, users of the Academic Twitter API can transfer the Tweet IDs provided by that service and import them into NodeXL via the NodeXL Twitter Tweet ID List Importer. We also recommend examining Twitter's documentation for more detailed information (Search operators, n.d.) about authoring queries using advanced query operators. In the NodeXL search box, the default keyword is *NodeXL;* we have replaced it with *Diabetes* (Figure 3). Once the data are downloaded, a pop-up dialogue box may note that it will be faster to import by turning off text-wrapping (click *yes*). Alternative Twitter importers in NodeXL also allow users to retrieve data created by specific users (Import from Twitter Users Network...)

or by providing the Twitter Tweet ID number of a list of specific tweets (Import from TweetID List network).

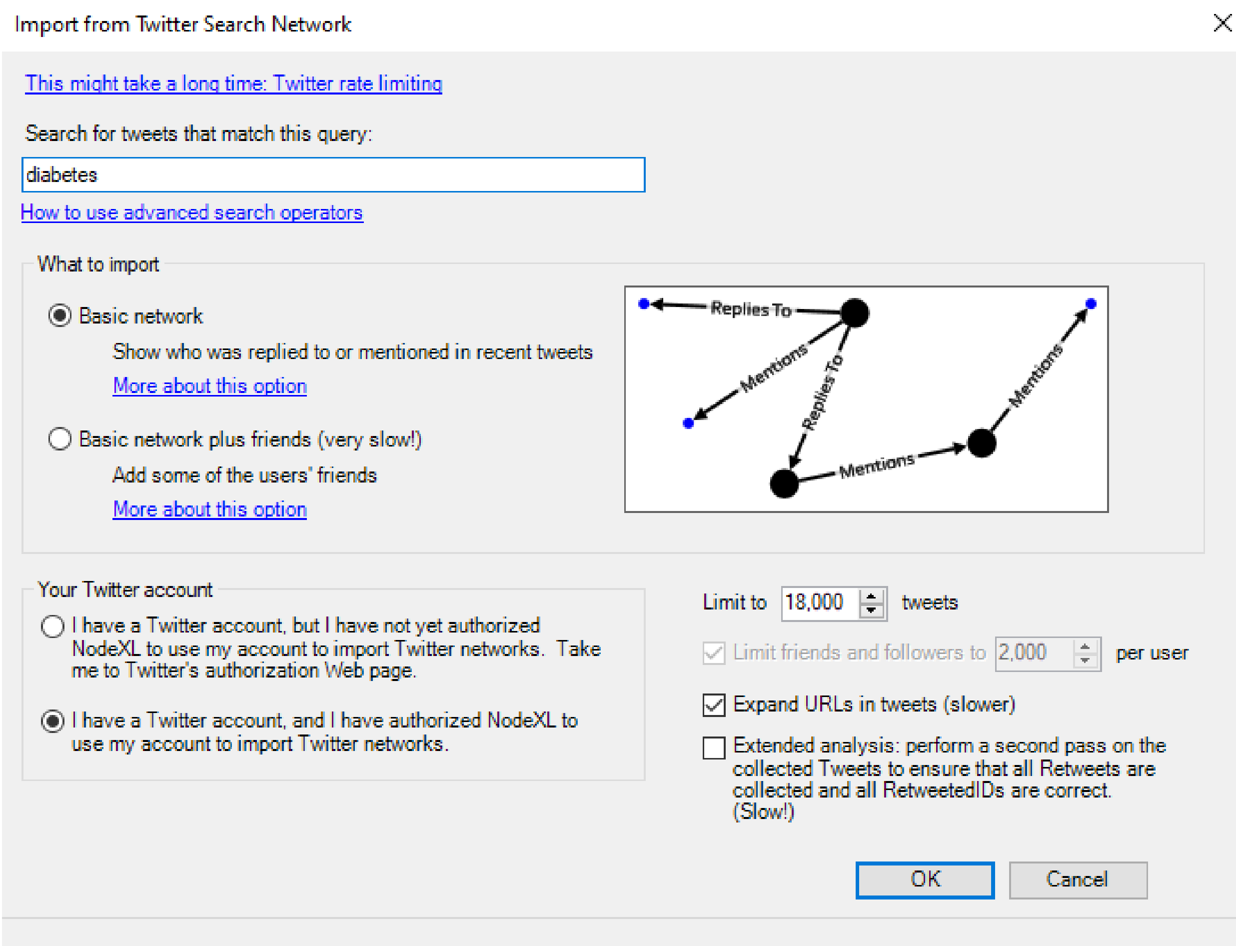


Figure 3. Importing data from the Twitter search API

Raw Data

When Twitter determines that it has delivered enough data, NodeXL then imports and displays the results in a series of worksheets in the Excel Workbook. Figure 4 shows the worksheet with imported data from Twitter and columns.

NodeXL can manage different numbers of edges based on the available computer system resources. The following are rough estimates:

- 4 GB of RAM: SMALL networks only less than a few thousand edges
- 8 GB of RAM: MEDIUM networks of less than 10-15 thousand edges
- 16GB of RAM: LARGE networks of less than 80-100 thousand edges
- 32 GB of RAM: VERY LARGE networks of less than 200 thousand edges

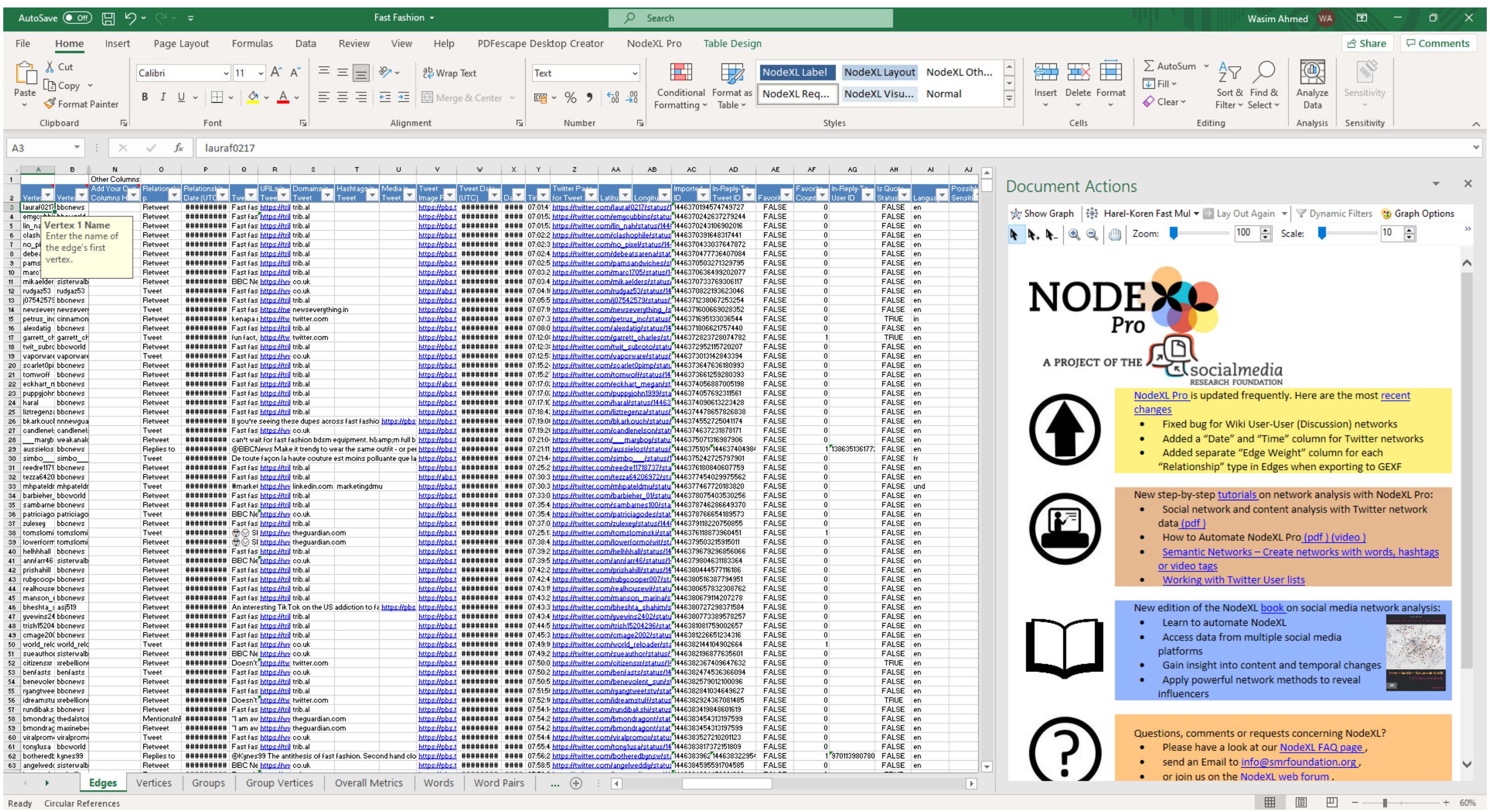


Figure 4. NodeXL Edges worksheet

Automated Network and Content Analysis

Automated network and content analysis of Twitter data is a key feature of NodeXL. NodeXL allows users to apply a “recipe” to their data that instructs the application in the steps needed to analyze and visualize the data. These “recipes” are technically referred to as “settings options files”. Researchers may download and import appropriate NodeXL Pro settings options files in order to automate the analysis of their data sets (Step 1, 2 and 3 in Figure 5). NodeXL settings options files for Twitter data contain all the configurations needed to automate tasks for in-depth social network and content analysis. A variety of NodeXL options files can be downloaded for free from the Social Media Research Foundation website (NodeXL Pro Task Automation, n.d.). We suggest that readers use the standard Twitter settings options file, which is optimized to analyze datasets composed of Tweets. Once this step is complete, the NodeXL graph pane will show a social network graph with images of the Twitter users as vertices in the network.

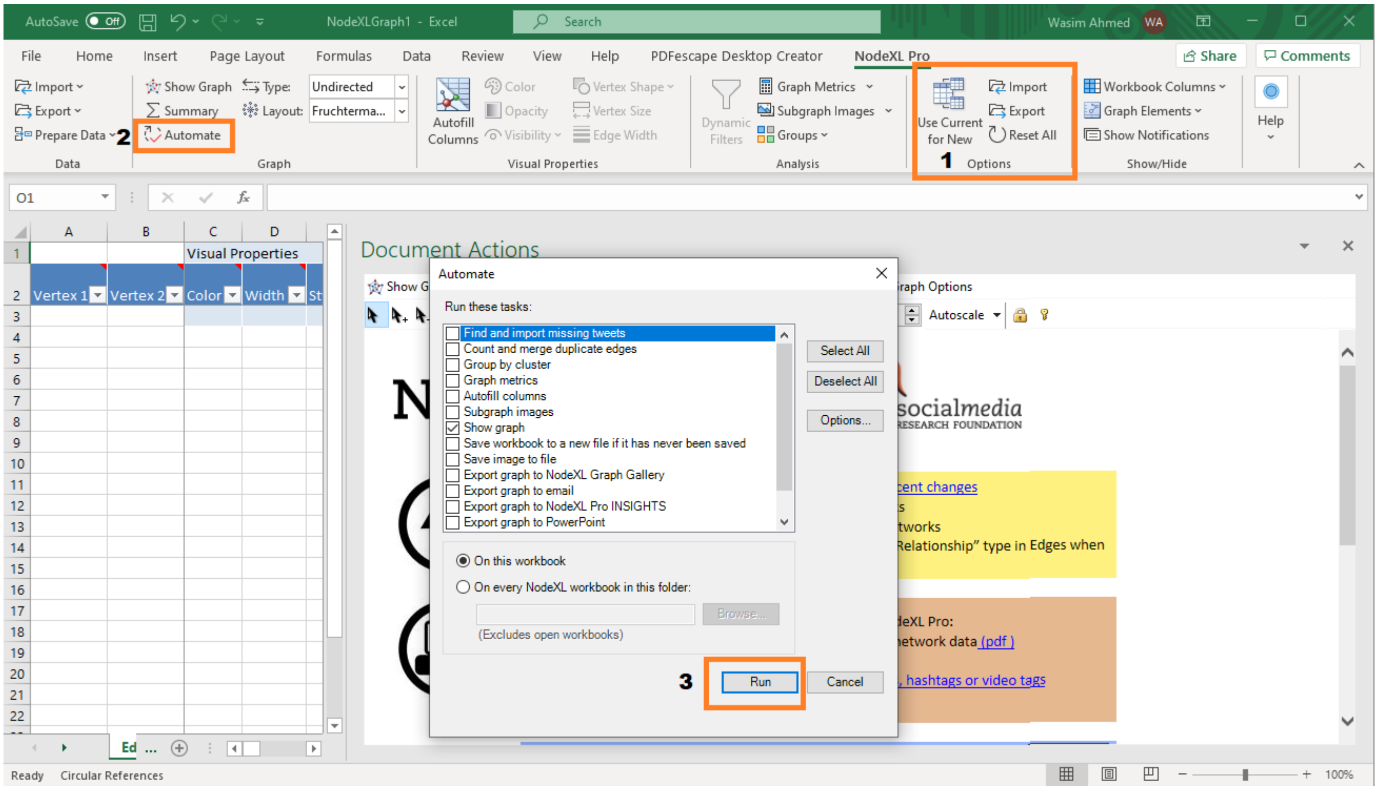


Figure 5. Automated Network and Content Analysis

NodeXL provides tools for the analysis of the contents of the annotations often found in networks created by the exchange of messages. For example, tweets are composed of short texts that can be analyzed for word frequency and specific word usage. Text data analysis is an automated step that can be included in the automated NodeXL analysis of Twitter data. NodeXL analyzes text at multiple levels, starting with the individual message where it counts and categorizes the words used in the text. Text from all the messages from a single person or user account is also collected and analyzed. Similarly, each group of people or user accounts (as defined by the groups identified from the patterns of network connections) has a set of messages that can be analyzed. The largest resolution is the set of all messages collected in the data set. Researchers may also categorize words by creating up to three vocabulary lists. Each list the researcher provides can be labeled. By default, NodeXL provides lists of positive, negative, and angry American English words (Liu, 2015). However, users are encouraged to modify or replace these lists to tailor these vocabulary lists to their research domains.

The NodeXL Workbook

When all steps of the data automation process are completed, you will see the network graph visualization in the "Document Actions" graph pane, and a number of tabs will appear along the bottom of the NodeXL workbook. Here is a short overview of the available information in the worksheets within the NodeXL workbook:

- **Edges**: A table with all tweets and edges (connections) between vertices (users) in the dataset along with metadata about the tweet such as tweet ID, tweet date, language, device used, number of retweets and favorites.
- **Vertices**: A table with all Twitter users in the dataset along with network metrics like Degree and Betweenness Centrality and metadata about the users such as self-description and location.
- **Overall Metrics**: Summary of major network metrics that describe the size, shape, density, and structure of the network. Identifies the data source, time of data collection and the time span of the data set.
- **Words and Word pairs**: Tables with all words and word pairs with overall count and count by group.
- **Network Top Items:** Top Influencers, Top URLs, Top Domains, Top Hashtags, Top Words, Top Word Pairs by overall count and group count.
- **Time Series:** A pivot table that visualizes the tweet activity in the dataset over time.

Moreover, NodeXL provides export features to easily share the NodeXL workbook. This is an important feature for researchers because many journals such as *PLOS One* require the raw data to be publicly available as part of their publication standards. NodeXL can generate PowerPoint presentations, email reports or publish a summary of the network via the website NodeXL Graph Gallery where multiple data sets uploaded by many users can be reviewed or downloaded for further analysis (NodeXL Graph Gallery, n.d.). Figure 6 provides a screenshot of the NodeXL graph gallery. NodeXL also can export to the GEXF and GraphML network file formats. These formats are used by other network analysis tools, like Gephi (Gephi - The Open Graph Viz Platform, n.d.), which can create distinctive large-scale visualizations of networks collected by NodeXL. Recently, the new NodeXL Pro INSIGHTS product extends these report options by building on the Microsoft Power BI data analysis cloud platform. NodeXL Pro INSIGHTS provides data visualization discovery and presentation features that build on top of the data collection, analysis, visualization, and reporting features in NodeXL Pro.

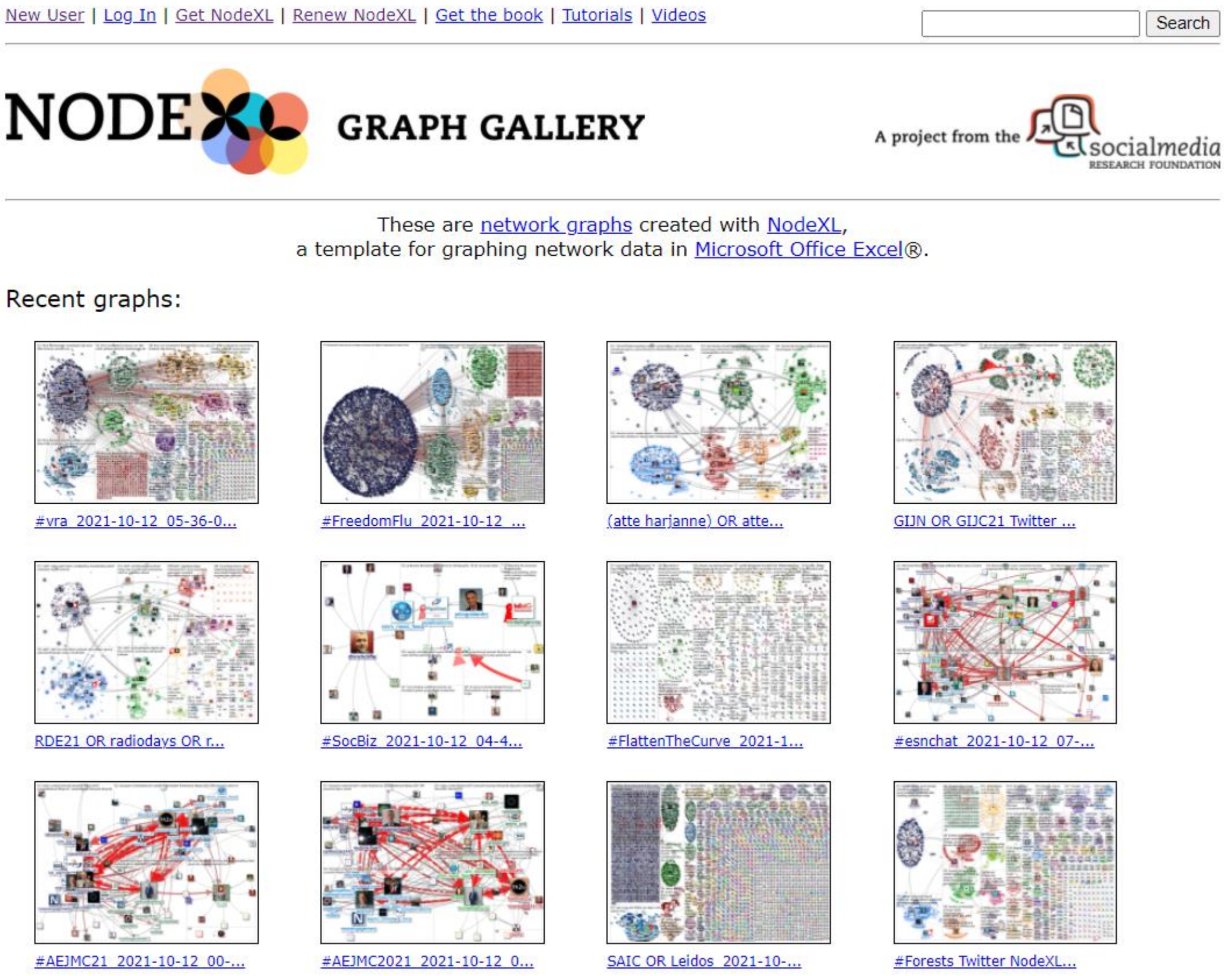


Figure 6. NodeXL Graph Gallery

Social Network Graphs

On Twitter, connections are formed when user accounts create messages that encode connections with other user accounts. Replying to a user creates a connection from that user to the author of the initial message. "Mentioning" a user (@username) is another common way to create a connection. Analyzed in aggregate, these connections form clusters that can be identified with various community detection algorithms. Within each cluster it is often the case that only a few strategically located individuals are in positions that indicate influence and power. Twitter networks typically follow one of six types of shapes, and guidance exists on how to interpret these graphs (Smith, Rainie, Shneiderman, and Himelboim, 2014) which will be outlined in the next section.

One result of social media network analysis is a visualization of the social network graph (Wasserman and Faust, 1994). These images illustrate webs of interactions and provide visual indicators of different social patterns. Figure 7 shows the social network graph generated for our Twitter search of *diabetes*. The top hashtags in tweets related to *Diabetes* include *#insulin,* and *#health*. The most influential Twitter users in this discussion network include official diabetes-related accounts such as *@DiabetesUK* (a leading British charity for people living with diabetes in the UK) as well as a number of individual patient advocates.

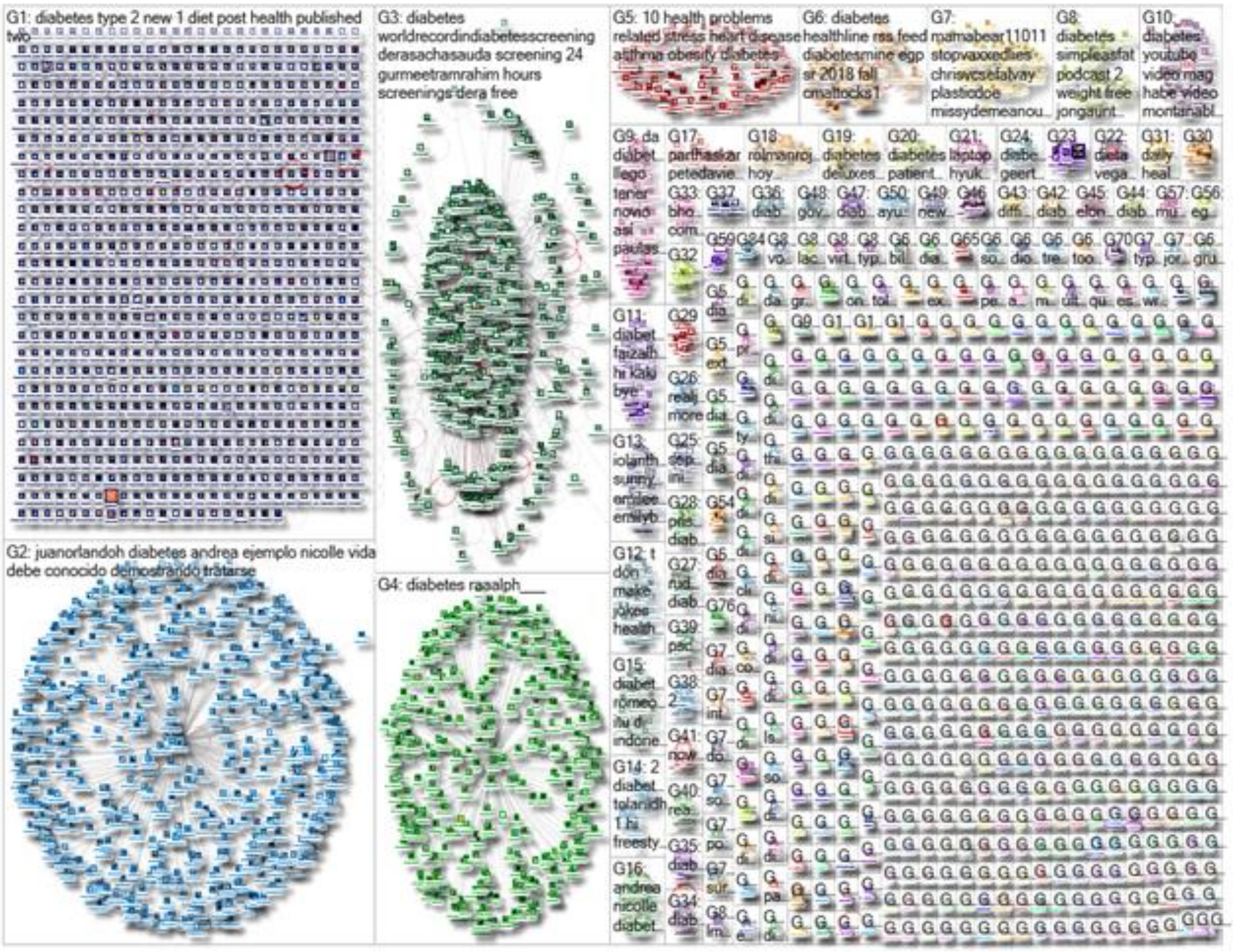


Figure 7 – Social Network Graph of Diabetes

Six Types of Twitter Networks

Previous research has identified six types of network shapes that Twitter topics tend to emerge from the interactions of many users (Smith, Rainie, Shneiderman, and Himelboim, 2014). Figure 8 provides a simplified overview of these different. The figure shows that different topics on social media can have contrasting network patterns. Online discussions of sports can be a useful example. In a ***polarized crowd,*** discussions among one set of users may focus on the "home" sports team as well as on rival teams. In a network visualization, this pattern would appear as two distinct clusters where users from one cluster are rarely connected to users in the other cluster.

In a unified crowd, users talk to one another with a high amount of overlap among themselves. In contrast, ***brand clusters*** are composed of many people who lack connection to one another and are not mentioned by any other users either. Like confetti, these isolated users are indicators that a topic has attracted a "periphery" of lightly engaged users. While these users, individually, are not influential, in aggregate the proportion of the population in a network that are isolates is an indication of the public or "brand" awareness of the topic. Many brands attract a high percentage of disconnected "isolates".

In a ***community cluster,*** distinct groups of users talk about a common topic with minimal connections across groups. These topics are often the "news of the day" and attract content from a large number of broadcast users. ***Broadcast networks*** form around influential users who create content that is retweeted by a large number of other users. Visually, broadcast networks form "wagon-wheel" shaped structures with a hub (the influencer) surrounded by spokes (the people who retweet the influencer). The distinctive feature of this structure is the lack of connections among the spokes (other than their common connection to the hub). We can think of ***support***

networks as the mirror image of the broadcast network, it is centered on accounts which reply to a large number of accounts. Often these networks are centered around customer support accounts for large businesses and services, like banks, airlines, and internet providers which may reply to a large number of Twitter users.

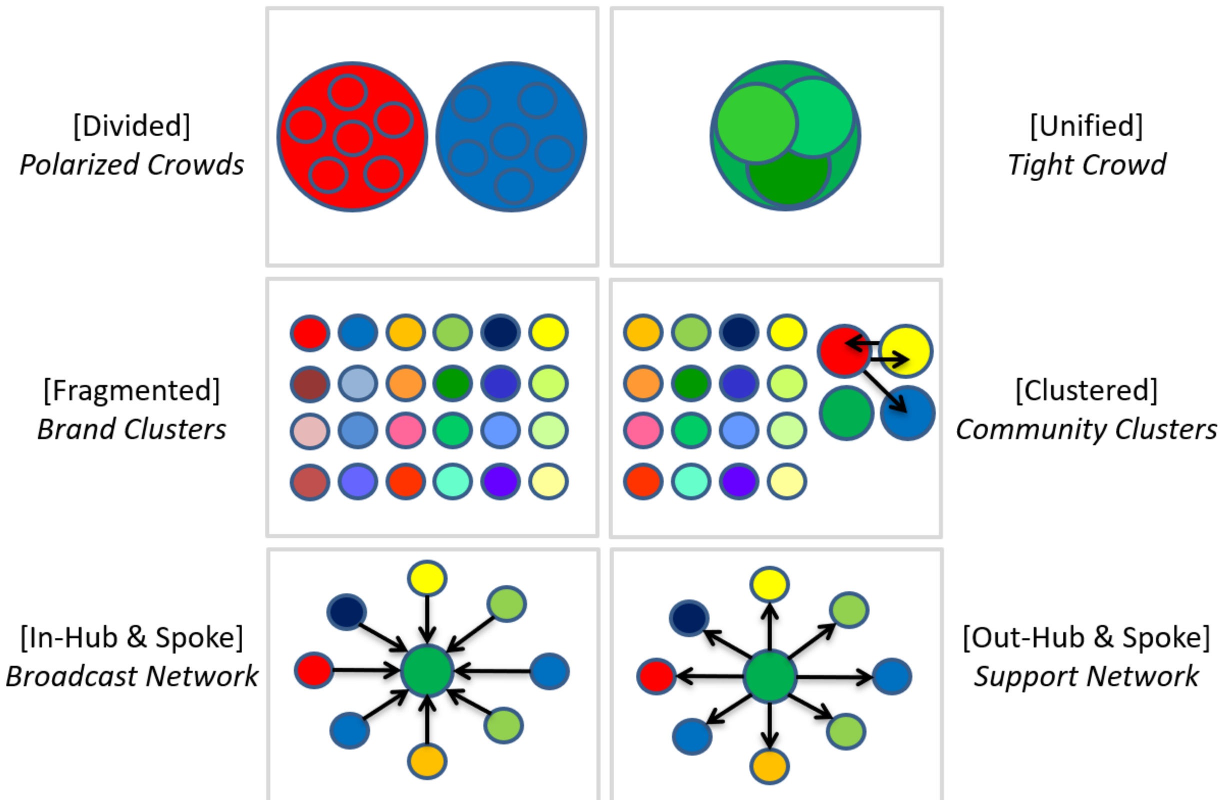


Figure 8- Simplified View of 6-Types of Twitter networks

Analysis of social media networks can reveal the overall structure of the collection of connections formed within message archives while highlighting key people positioned strategically in often rare locations within the network. These influential people or user accounts are able to command the attention of many more people than the average participant (who often attracts little to no attention at all).

NodeXL packages this analysis in an easy-to-use data workflow that can often be performed almost entirely automatically. Inspired by the evolution of cameras that now make photography easy, NodeXL simplifies the mechanical process of data collection, analysis, visualization and reporting to enable researchers to focus on the meaning and implications of images and data rather than on the process of creating them.

One of the key advantages of NodeXL is its ability to import and view raw Twitter activity data, allowing analysts to further interpret and interrogate the data set. Social media data from conferences and other social media activities can be analyzed in this way. For example, a recent study analyzed tweet content from a large European cardiology conference (Hudson and Mackenzie, 2019).

The following sample analysis will demonstrate the entire research process by focusing on a sample topic and research question. This example will illustrate how NodeXL could be utilized in order to address a research question.

Analysis of Tweets from UK based *Russell Group* research Universities

Social media platforms contain multiple topics of discussion and permit a wide range of academic research to be conducted. Research on social media has explored topics such as identifying vaccination influencers (Sanawi, Samani, and Taibi, 2017), political uses of social media (Batool, Ahmed, Mahmood, and Saeed, 2021), climate change (Joseph, Fred, Philip, William, and Gerald, 2016), and political issues such as Brexit (Chung and Kim, 2019).

Social media network maps and reports reveal the various ways topics and issues of research and social interest are collectively structured. While each individual chooses what to tweet or who to tweet to, the behavior of thousands of individuals form emergent patterns that reflect the different kinds of generative social processes at work.

The United Kingdom higher education sector is a good example. Twitter is a public facing platform that can be used to transmit the brand image of an institution to the outside world. It is also a place where the public can flock to share their thoughts and opinions when a certain university is under heightened interest. This information can be used to identify influential users in the swarm of content in order to engage with leading voices and strengthen connections that can help spread messages further.

The higher education market has become increasingly competitive, and universities are known to monitor their place in national and international league rankings, perceptions of the quality of their research and teaching, and their international image. In recent years, social media reach and influence has become an important aspect in promoting the brand of a university and many institutions will use social media to actively recruit students from around the world.

Our sample analysis takes data samples from 2017 for three popular and recognized Russell Group Universities, the University of Glasgow (Figure 9), and Imperial College London (Figure 10) and generates a network visualization of each from the same week and then performs a two-way comparison of the network shapes and structures as shown in Figure 11.

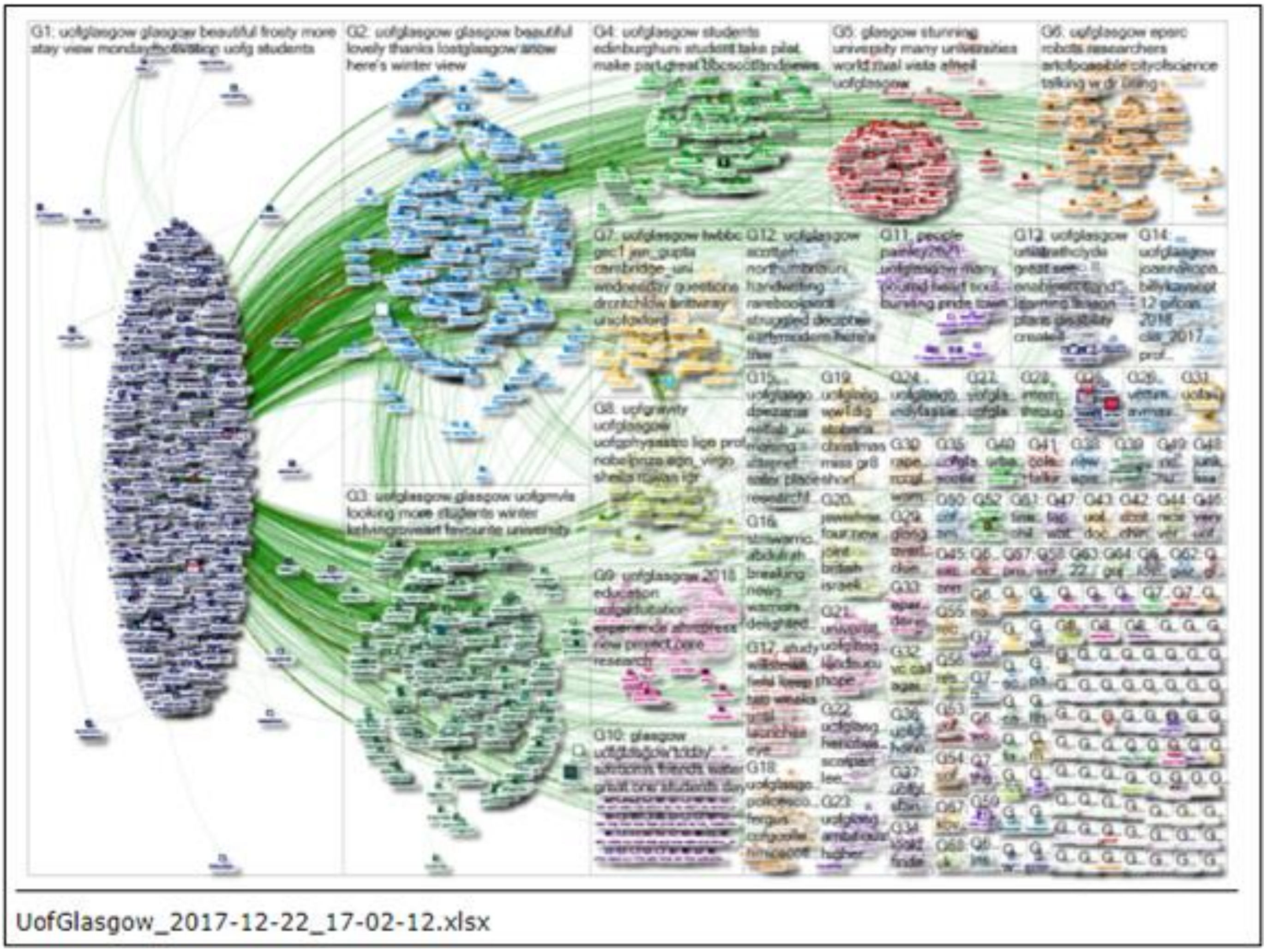


Figure 9 - University of Glasgow

The graph above demonstrates that there are a number of different groups of users conversing about different topics which are related to the University of Glasgow. The most frequently occurring hashtags are: *Glasgow, mondaymotivation, uofgwinter, Christmas, worldchangers, snowday, uksnow, teamuofg, uofgchristmas and artofpossible.*

The tweets in the network were tweeted over the 13-day, 23-hour, 34-minute period from Thursday, 07 December 2017 at 01:16 UTC to Thursday, 21 December 2017 at 00:50 UTC. There were 3,100 users in the network.

The network shape is dominated by the "broadcast" structure of the nine largest groups. Each features a clear "hub and spoke" or "broadcast" pattern that is generated when an influential user tweets a message that is then retweeted or replied to by a large group of otherwise disconnected users.

The users at the center of these hub and spoke patterns are the most influential contributors in this content collection. The key accounts, ranked by betweenness centrality in this network consist of:

- **@uofglasgow** - This corresponds to the official Twitter account of the University of Glasgow
- **@afneil** - This corresponds to Wes Streeting. The Labour MP for Ilford North. Shadow Exchequer Secretary to the Treasury
- ID anonymised - An account belonging to a Lecturer (as the account no longer exists it was not possible to obtain consent to use the Twitter handle)
- **@Paisley2021** - An account representing the town of Paisley.
- **@rarebookscot** - This account corresponds to the 'Rare Books in Scotland' organisation which is a forum for people working with rare books in Scottish libraries.

The tenth largest group in this network is the “brand” cluster which is composed of users who have no connection to any other user. This fragmented population is an indication of the amount of brand awareness for this University. The isolate users mentioned the University but did not reply-to or retweet it. The relatively small size of the brand cluster in this network indicates a lower proportion of public awareness of this institution in contrast to the other two universities.

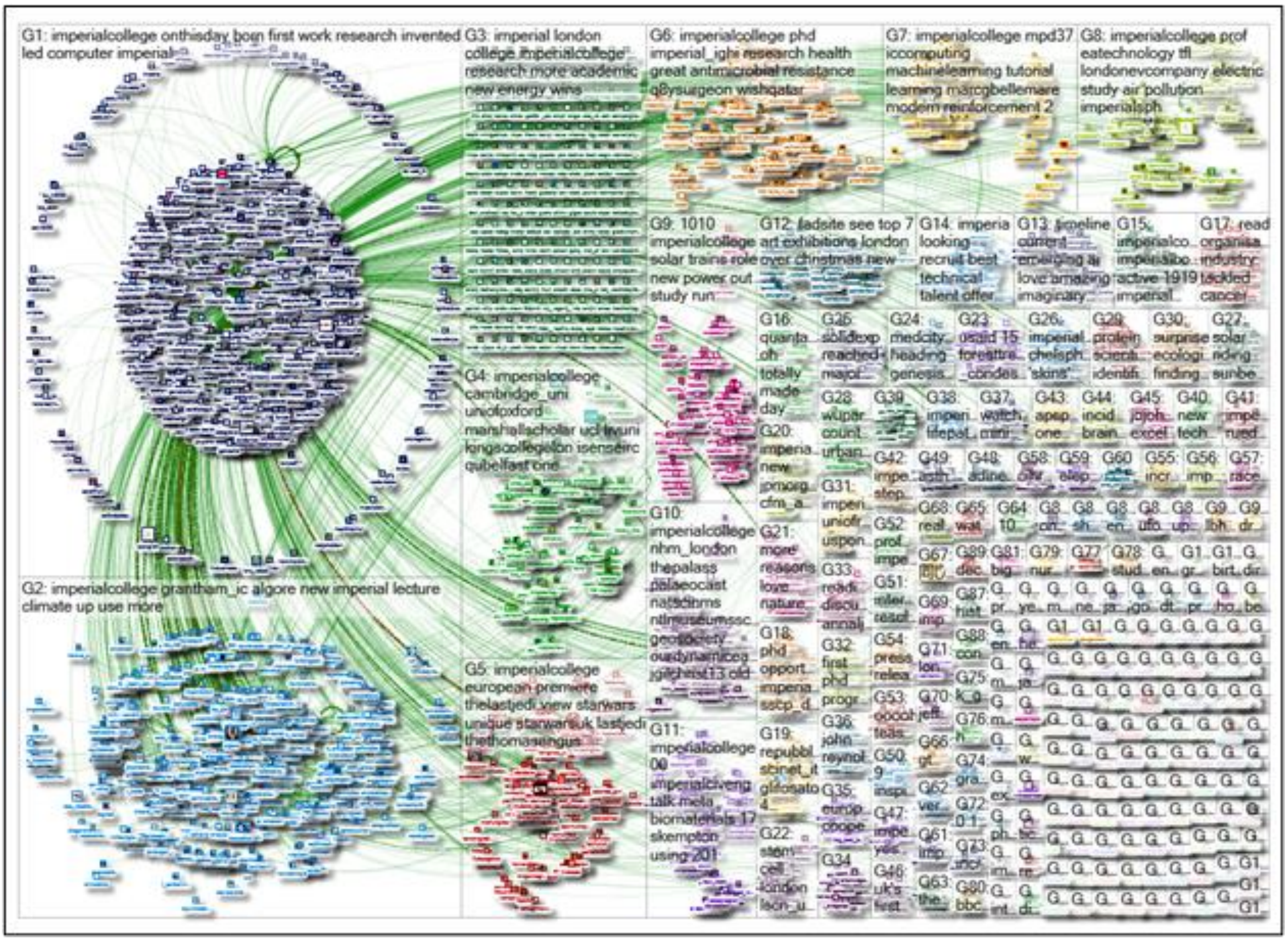


Figure 10 - Imperial College London

The network graph of Imperial College London demonstrates that there are a number of different groups of users sharing content related to the University. The most frequently occurring hashtags are *onthisday, machinelearning, thelastjedi, imperialcollege, grantham10, ai , lastjedi, malaria, earthscience, and london*.

The most frequently shared URL was titled ‘UK's first PhD programme to tackle antimicrobial resistance offers studentships’, and the second most shared URL was an Eventbrite page titled ‘Ethics in AI’

In contrast with the University of Glasgow’s network, the network for the Imperial College London demonstrates the benefits of Twitter as a means to promote events among the general public. It is interesting to note the sizeable isolates group (users who are not connected to each other or to others within the network) who share various news articles and/or events from Imperial College London. This indicates that Imperial College London may have a broader appeal outside their own network.

The influential users in the Imperial College London Twitter network consist of:

- **@imperialcollege** - this corresponds to the official Twitter account of Imperial College London
- **@grantham_ic** - this organizational based account corresponds to Imperial College London’s hub for climate & environment, leading on world-class research, training & innovation towards a sustainable, resilient, zero-carbon society.
- ID anonymized - this account belonged to a citizen (it was not possible to obtain consent to use the handle as the account had closed down).
- **@imperialmed** - this official organizational Twitter account belonged to the Department of Medicine at Imperial College London.

- **@Fisher85M** - this user was a technology influencer.

The tweets in this network were tweeted over the 13-day, 23-hour, 31-minute period from Thursday, 07 December 2017 at 01:15 UTC to Thursday, 21 December 2017 at 00:46 UTC. There were 3,093 users in the network.

NodeXL can modify the display of network graphs with hundreds of customization features. In the table below we have simplified each network graph by removing group and edge labels, and by replacing the vertex shape with a disk.

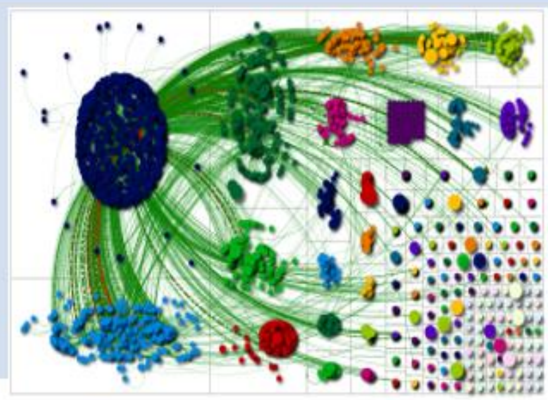
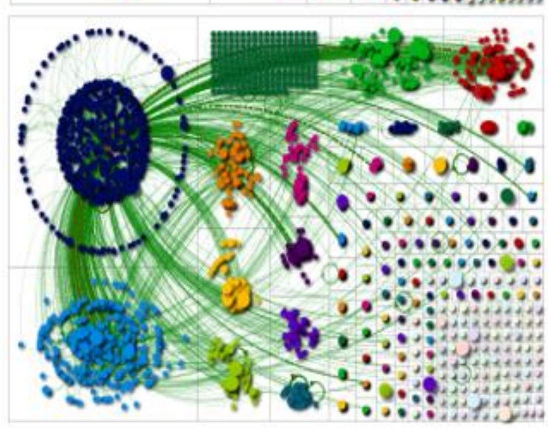
Rank	Institution	Network Structure
1	University of Glasgow	
2	Imperial College London	

Figure 11. Network Comparison of Network Shapes and Structures.

With figures 8 and 9 in mind (network shapes and structures), we note that a common feature of these network graphs are the multiple prominent broadcast networks which indicate that the content generated has attracted an audience but not a community. These network graphs showcase the ways topics cluster into sub-groups, rather than a single monolithic discussion, social media features a large number of smaller discussions.

It is also possible to examine and compare further Russell Group Universities as shown in figure 12 below.

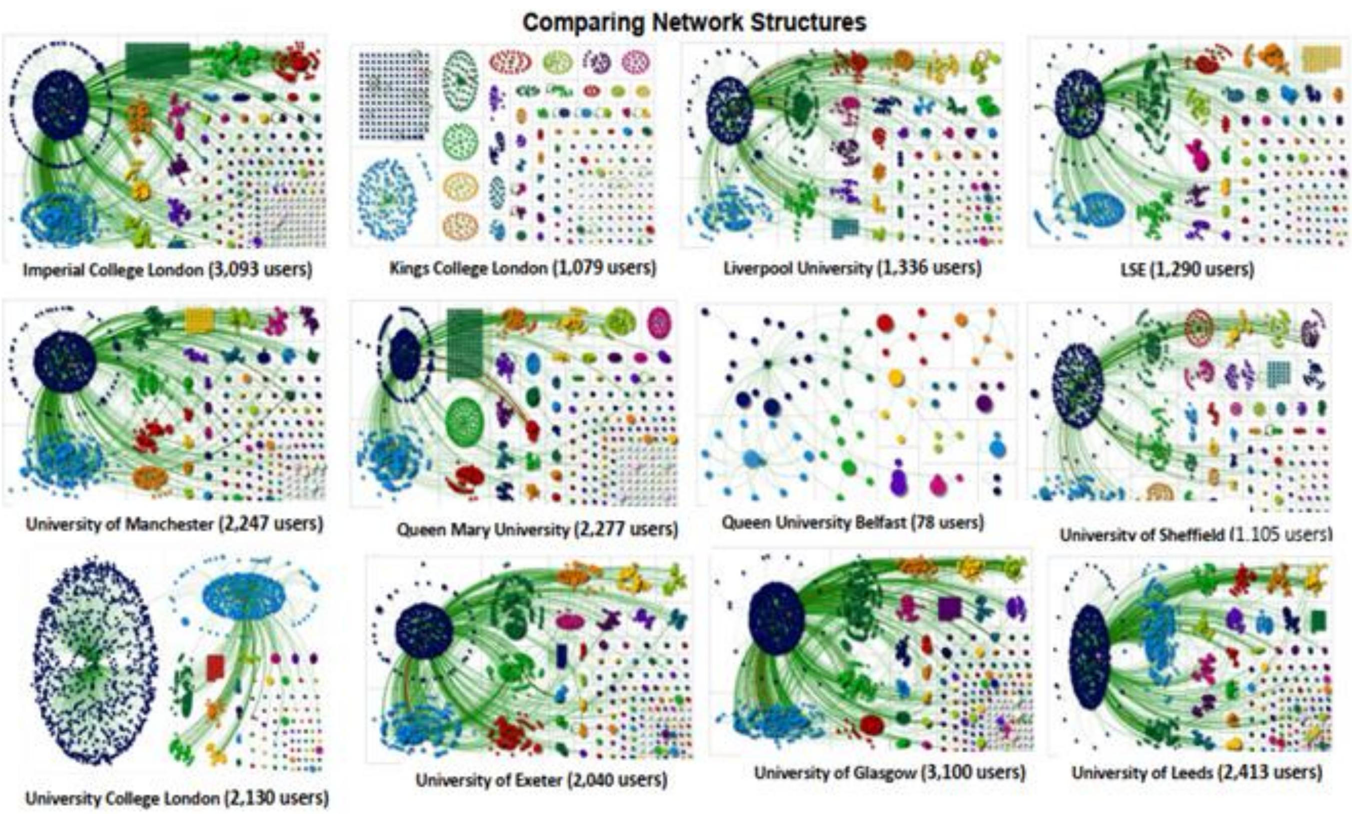


Figure 12- Overview of 12 Russell Group Networks

Taking Action: From Theory to Practice

How can universities leverage these insights? Social network graphs have many actionable features designed to optimize social media network engagement and message amplification. Social media and communications managers can use social media network maps to gain a broader view of their conversation topics, making it possible to recognize their overall shape and the key leaders within them. Similarly, they can monitor the topics related to competing universities to maintain situational awareness in their markets.

Each network shape may or may not be the one that is best for every purpose. A series of social media network maps made over time can guide and track the evolution of a topic network from one network structure into another, more desirable, network structure. For example, brands often want to foster communities, divided groups sometimes want to form unified groups, and broadcasters often want to become brands. Social media network analysis can provide measures and guidance to clients who seek to manage the transformation of a social media space from one pattern of connection to another (as shown in Figure 14).

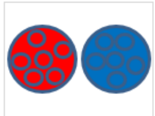
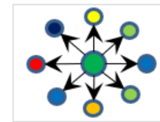
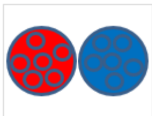
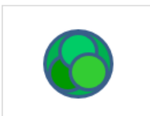
	[Divided] *Polarized Crowds* 	[Unified] *Tight Crowd* 	[Fragmented] *Brand Clusters* 	[Clustered] *Communities* 	[In-Hub & Spoke] *Broadcast Network* 	[Out-Hub & Spoke] *Support Network* 
		[Low probability] Find bridge users. Encourage shared material.	[Low probability] Get message out to disconnected communities.	[Possible transition] Draw in new participants.	[Possible transition] Regularly create content.	[Possible transition] Reply to multiple users.
	[Undesirable transition] Remove bridges, highlight divisions.		[Low probability] Get message out to disconnected communities.	[High probability] Draw in new participants.	[Possible transition] Regularly create content.	[Possible transition] Reply to multiple users.
	[Undesirable transition] Increase density of connections in two groups.	[Low probability] Dramatically increase density of connections.		[High probability] Increase retention, build connections.	[Possible transition] Regularly create content.	[Possible transition] Reply to multiple users.
	[Undesirable transition] Increase density of connections in two groups.	[Low probability] Dramatically increase density of connections.	[Undesirable transition] Increase population, reduce connections.		[Possible transition] Regularly create content.	[Possible transition] Reply to multiple users.
	[Undesirable transition] Increase density of connections in two groups.	[Low probability] Dramatically increase density of connections.	[Low probability] Get message out to disconnected communities.	[Possible transition] Increase retention, build connections.		[High probability] Increase reply rate, reply to multiple users.
	[Undesirable transition] Increase density of connections in two groups.	[Low probability] Dramatically increase density of connections.	[Possible transition] Get message out to disconnected communities.	[High probability] Increase retention, build connections.	[High probability] Increase publication of new content and regularly create content.	

Figure 13. Taking Action using NodeXL reports

Social media managers are confronted with questions such as:

- How do I increase visibility of selected messages?
- How do I raise the number of followers, likes and retweets?
- How can I become a top influencer of a discussion?
- How do my messages become viral?
- How can I gain actionable insight?

From a network point-of-view this translates to:

- How do I build network reach?
- What divisions or groups are present?
- Who are the most influential people in the discussion?
- What exactly are they talking about?
- This leads to the ultimate question: "To whom should I say what?"

Network science can help to answer all of these questions. This approach is different from other social media analysis techniques that simply aggregate all content related to a particular topic into a single collection of messages. Instead of treating a large group of users in a uniform way, social network analysis reveals that social media crowds form a variety of groups with very different shapes and structures and a very few people occupy positions of influence and authority within these groups. The combination of social network and content analysis provides necessary information to support influencer and content marketing strategies in social media.

Reducing the need to read thousands of tweets and messages on a range of topics is a key benefit of social media network maps and reports like the ones created with NodeXL. NodeXL reports summarize vast amounts of data and provide valuable sets of scientific metrics which can be used for measuring and monitoring a range of topics of interest. Using this approach, social media managers can take a step back and see discussion networks as a whole. These overviews also allow users to drill down and zoom into the details of social media networks, for instance, displaying the messages associated with the most influential users. Scholars using these data sets can document the evolution of topics and changes in leadership in a variety of online discussions.

Discussion

It is understood that financial and commercial marketplaces require practices like accounting and auditing in order to manage inevitable issues with fraud and deception and manipulation. Markets that resist these forms of corruption often require reporting and auditing of accounting to ensure that outside investors can get an accurate sense of the health and prospects of a business. In a marketplace of ideas there is a need for an analogous accounting system. Tools like NodeXL can be thought of as accounting software for social media. They allow for each message, person and group to be credited or debited with connections and influence based on the empirical mapping and analysis of connections and content over time. During the COVID-19 Pandemic, for example, academic research made use of social network analysis using NodeXL to identify users who were influential in spreading disinformation (Ahmed, Vidal-Alaball, Downing, and Seguí 2020).

Specific challenges and concerns may arise when conducting social media research. These include cost of access to data, spam, ethical, legal and privacy issues (see Burkell, Regan, and Steeves, Chapter X, this volume; Lutz, Chapter X, this volume). As there is a 7-10 day and 18,000 tweet limit on downloading Twitter data (via the free Search API) forward planning is required to study longer or larger campaigns if using this API. However, with the advent of the new Academic Track API, it is now possible to retrieve historical tweets retrospectively. Furthermore, a pitfall of social media data analysis is that certain tools may not provide complete and transparent access to data (Quan-Haase, Hollingshead, and Blank, Chapter X, this volume). Recent studies show NodeXL successfully downloads the great majority of the tweets when compared with Twitter search results (for example, for one large international conference NodeXL identified 97.5% of tweets displayed by the web search interface) (Cevik, Ong and Mackenzie, 2019). NodeXL has been shown to collect more tweets than other social network analysis tools (Søreide et al., 2019).

The Twitter API, however, can perform unpredictably, omitting prolific tweeters at peak tweeting times, so it is worth downloading extracts more than once, and sharing preliminary results with the relevant Twitter community (e.g., conference tweeters) to identify obvious gaps in data. These issues are changing with the expanded availability of the “Academic Twitter” service which removes many of the volume and time constraints imposed on the public Twitter API.

The results of Twitter searches depend upon the quality of the search terms used. Poorly crafted search queries may not return expected content or may collect unwanted and unrelated content. In some cases more than one term or hashtag is used to identify the same topic or event, for example tweets about a conference may use a mix of four digit or two digit years in the hashtag. Inconsistent use of hashtags and other query terms may require researchers to refine and modify the queries they use to collect data sets. Unlike many popular internet search engines, Twitter does not perform “fuzzy” searches where similar terms are found along with the search term. As a result, the search term “haemophilia” would not identify tweets using the US spelling, while “haemophilia” OR hemophilia” would capture both English language spellings of the term. Extending the search term to include the various permutations is possible, but there is an upper character limit on the search term and more complex terms may provide less complete results. Ironing out inconsistencies and gaps in search terms is important early on in the research data collection process, and researchers may need to adapt query terms over time when studying longer durations as topics shift.

Conclusion

Network maps can be combined with text analysis to reveal an overview of social media networks while providing detailed insights into leaders, topics, and segmentation. Using NodeXL enables wide access to this level of analysis. Researchers using NodeXL should have an easier time collecting data in a timely way that can then be richly processed into an image and a report that can be quickly used to document a claim or observation about the collective activity in an online discussion. With expanded access to timely insights these tools may enable better social self-control and policy planning regarding social media platforms and services.

References

Ahmed, W., & Lugovic, S. (2019). Social media analytics: analysis and visualisation of news diffusion using NodeXL. *Online Information Review*.

Ahmed, W., Vidal-Alaball, J., Downing, J., & Seguí, F. L. (2020). COVID-19 and the 5G conspiracy theory: social network analysis of Twitter data. *Journal of Medical Internet Research*, *22*(5), e19458.

Ahmed, W. "Using Twitter as a data source: an overview of social media research tools (updated for 2017)." *Impact of Social Sciences Blog* (2017).

Batool, S. H., Ahmed, W., Mahmood, K., & Saeed, H. (2021). Twitter dialogue: an analysis of Pakistani politicians' information sharing. *Information Discovery and Delivery*.

Cevik, M., Ong, D. S., & Mackenzie, G. (2019). How scientists and physicians use Twitter during a medical congress. *Clinical Microbiology and Infection*.

Chung, S. W., & Kim, Y. (2019). The truth behind the Brexit vote: Clearing away illusion after two years of confusion. *Sustainability*, *11*(19), 5201.

Cline, D. H. (2012). Six degrees of Alexander: social network analysis as a tool for ancient history. *Journal of Sociology*, *98*(6), 1259-1319.

Developer.twitter.com. n.d. *Search operators*. [online] Available at: <https://developer.twitter.com/en/docs/tweets/rules-and-filtering/overview/standard-operators> [Accessed 12 October 2021].

Developer.twitter.com. n.d. *Twitter API for Academic Research*. [online] Available at: <https://developer.twitter.com/en/products/twitter-api/academic-research> [Accessed 12 October 2021].

Gephi.org. n.d. *Gephi - The Open Graph Viz Platform*. [online] Available at: <http://gephi.org> [Accessed 12 October 2021].

Hudson, S. & G. Mackenzie (2019). 'Not your daughter's Facebook': Twitter use at the European Society of Cardiology Conference 2018. *BMJ Hearts.* Volume 105, Issue 2

Joseph, O., Fred, M. I., Philip, N., William, O., & Gerald, M. (2016). Applying social network analysis, centrality measures in identification of climate change adaptation opinion leaders. *International Journal of Agricultural Research, Innovation and Technology*, *6*(1), 1-7.

Liu, B. (2015). Sentiment analysis: Mining opinions, sentiments, and emotions. Cambridge University Press.

Nodexlgraphgallery.org. n.d. *NodeXL Graph Gallery*. [online] Available at: <https://nodexlgraphgallery.org/Pages/Default.aspx> [Accessed 12 October 2021].

Sanawi, J. B., Samani, M. C., & Taibi, M. (2017). # vaccination: Identifying influencers in the vaccination discussion on twitter through social network visualisation. *International Journal of Business and Society*, *18*(S4), 718-726.

Scott, J. (1990). Social networks. Sage Publications.

Smith, M. A., Rainie, L., Shneiderman, B., & Himelboim, I. (2014). Mapping Twitter topic networks: From polarized crowds to community clusters. *Pew Research Center*, *20*, 1-56.

Social Media Research Foundation. n.d. *NodeXL Pro Task Automation*. [online] Available at: <https://www.smrfoundation.org/nodexl/automation/> [Accessed 12 October 2021].

Søreide, K., Mackenzie, G., Polom, K., Lorenzon, L., Mohan, H., & Mayol, J. (2019). Tweeting the meeting: Quantitative and qualitative twitter activity during the 38th ESSO conference. *European Journal of Surgical Oncology*, *45*(2), 284-289.

Wasserman, S., & Faust, K. (1994). Social Network Analysis: Methods and Applications. Cambridge University Press.