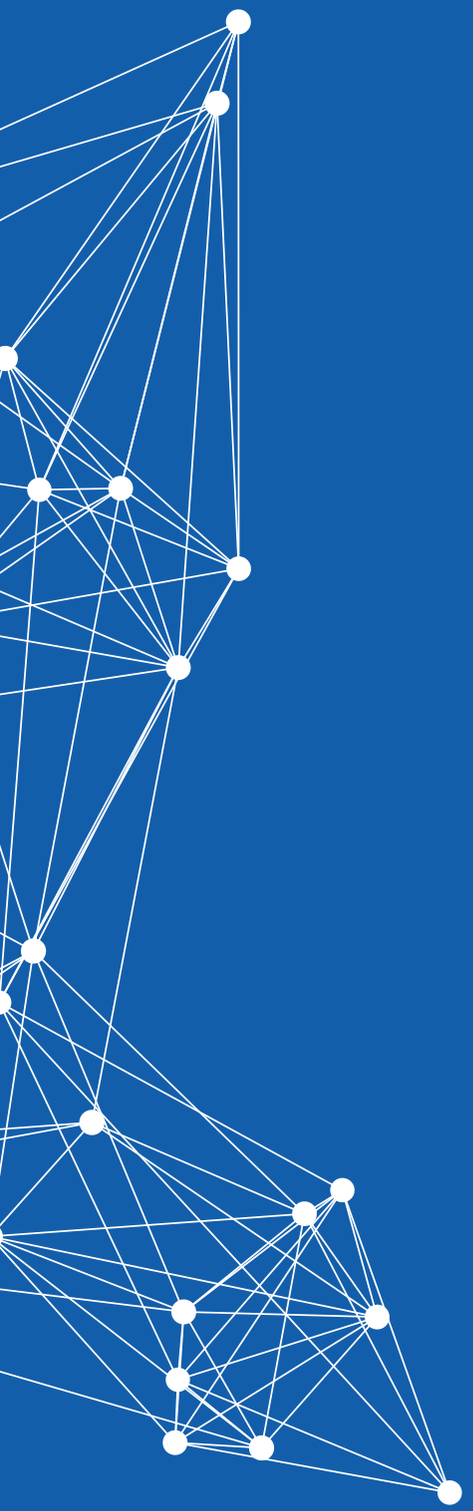
2019



Digital Rights Advocates Social Network Analysis

HANDBOOK

Prepared by Richard Ngamita

Internews

This report was funded by Internews

ABOUT THE AUTHOR

Richard Ngamita is a data consultant and technologist with more than 10 years of experience and training in data analysis, analytics, research methods and applied machine learning. He has worked at Facebook, Google and Medic, and served as a consultant in several fields including health, agriculture, and refugee issues with organizations such as the UN High Commission for Refugees, the Bill & Melinda Gates Foundation, and Conservation International. He has also worked as a consultant contributing to documentation & trainings for organizations like Ushahidi, CIPESA, Outbox, Refunite, Vital signs, Strongminds, Brac, and Intrahealth. Richard is a member of the "R Forward" open source community and contributes to the R programming language (R foundation) and recently founded KampalR user group.

TABLE OF CONTENTS

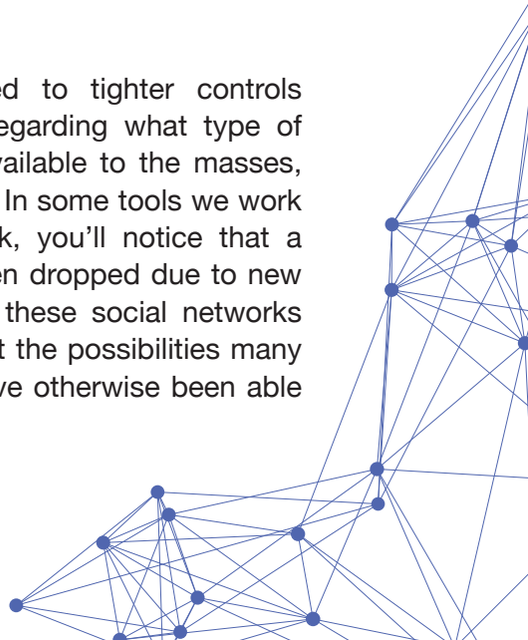
BACKGROUND

Social networks can be found wherever people are connected—within gatherings, communities, groups, organisations, between partners, and within markets. Put simply, a social network is made up of a number of actors who are connected by some type of relationship. Social network analysis is employed to understand the social structure that exists amongst entities in an organisation. The size, diversity, and ubiquity of a social network act in combination, thus necessitating an understanding of these networks in a systematic manner. Social network analysis (SNA) is the process of mapping these relationships as well as analysing the network's structure and the influence of different actors. This handbook is designed to provide a step-by-step guide to SNA's application for the Internews Digital Rights Advocates (DRAs). It has been developed in close collaboration with the KampalR (a group of data enthusiasts) and Outbox Hub data teams (Uganda, Nepal) and has received valuable contributions from several other research units. The approach draws upon social network theory, discussion-based tools (e.g., Voyant Tools), and other graphical software applications, such as Netvizz and Gephi.

Online social networks offer a considerable amount of information for studying human behaviour and social interaction. Twitter, Facebook, and Reddit are the most common sources used by researchers due to their data accessibility. Moreover, for advocates, social activists, and political scientists, social network data are extremely interesting due to the social and political nature of their conversations. Social network data are still a new source compared to other data types (e.g., surveys and interviews), and thus they might present a barrier for some advocates, researchers, and students. The case studies selected in this handbook will clarify initial questions digital advocates might pose. We share an introduction to SNA by providing its backgrounds and describing the most important terminologies. We then answer the question regarding where data from three different sources may be obtained, and what advantages and challenges each data type poses, specifically with regard to social network research. Next, we also address the outcomes or insights we can infer from the data to complete the initial questions.

Digital rights advocacy is on the rise in all parts of the world. Most activists working in-country on digital rights issues—that is, those related to online privacy, security, expression, and other human rights—regularly engage with complex issues of law and policy that have significant consequences for the social, political, and economic lives of their nations. This practice has birthed several activists who work with various data sources, including social networks. As we explore the use of social network data in this handbook, it is important that we note that, when working with open data or social network data, user privacy is a major concern. We learned the importance of this variable due to the 2018 Facebook–Cambridge Analytica data scandal, where it was revealed that Cambridge Analytica had harvested the personal data of millions of users' Facebook profiles without their consent for political advertising purposes. This incident has been described as a watershed moment in the public understanding of personal data that precipitated a massive fall in Facebook's stock price.

This incident has led to tighter controls on social networks regarding what type of data can be made available to the masses, including researchers. In some tools we work with in this handbook, you'll notice that a few features have been dropped due to new limitations set up by these social networks that significantly affect the possibilities many researchers would have otherwise been able to tackle.



WHY AND WHEN SHOULD DIGITAL RIGHTS ADVOCATES CONDUCT A SNA?

Reasons for performing an SNA:

I. To determine which actors are involved in a network; SNA breaks down the different actors across the network in a perfectly helpful and visual way such that advocates can easily see/identify them within the mentioned network.

II. To determine how actors are linked; the actors and linkages are displayed within the network, facilitating well-formed interaction across the selected data.

III. To determine how influential each actor is; the arrow sizes are indicators of influence that help identify which actors have more influence than others. Colour coding these actors and arrows helps identify additional relations across the network.

IV. To identify actors' motivations; the various actors mapped to others easily imply their motivations across the networks and feed into the cross-relations for more efficient and interactive explanations.

V. To determine how the network is structured; according to the structures, we can identify how the relations behave and how busy the network is by identifying the strong and weak points and opportunities for influence or improvement to make sense of the network as a whole.

From information into insights:

I. A more thoroughly informed project design may be created by SNA insights because they provide new information that would not be otherwise provided by alternative data sources.

II. Understanding the program audience helps drive more engagement and support for the programs.

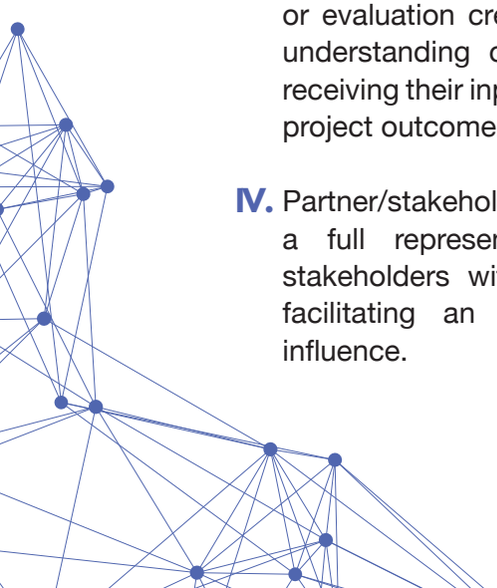
III. Partner input in the project design and/or evaluation creates the opportunity for understanding other partner views and receiving their inputs for a more favourable project outcome.

IV. Partner/stakeholder mapping provides a full representation of partners or stakeholders within the programs, thus facilitating an understanding of their influence.

V. Insights feed into and develop new advocacy strategies based on data and allows for data-driven decisions across the program's strategy process.

VI. Informed program transitions support the provision of new directions for programs transitioning towards creating more successful inter-program environments.

VII. Understanding programming and creating additional, sensitive touchpoints inclusive of gender, race, sexual orientation, and political background (among other variables) facilitate inclusive programming that considers important points that may have otherwise been ignored.



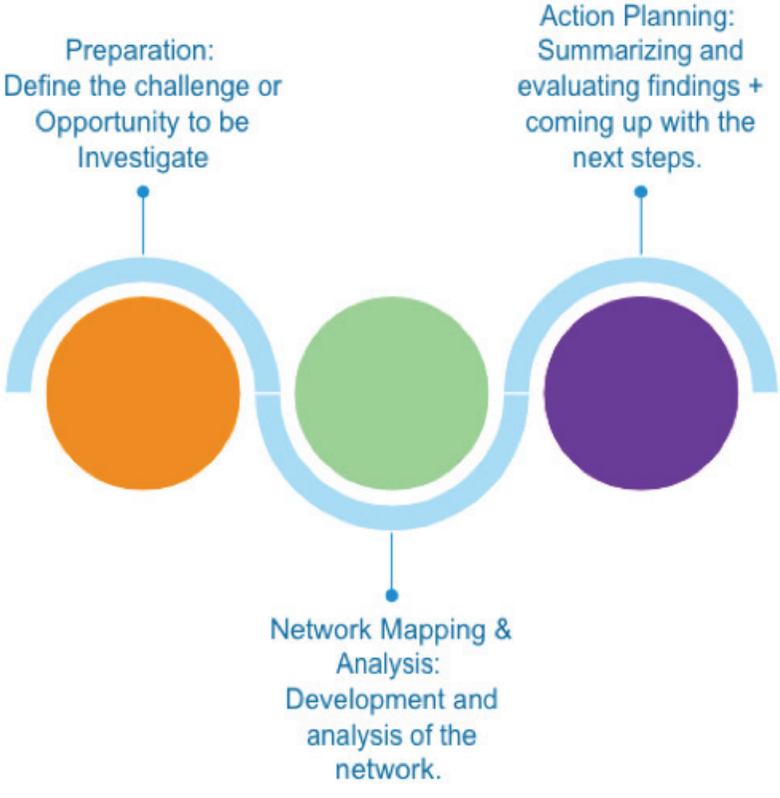
PROCESS

The SNA (Social Network Analysis) Process.

Social networks exist wherever people exist; they are inescapable, powerful, and often invisible structures that change over time and can either block or enable change. SNA is an intuitive and highly flexible process and as such can be readily adapted to the particular contextual circumstances and analysis requirements that we shall observe in the different case studies discussed in this handbook. SNA can increase the effectiveness and appropriateness of DRAs' work through a deeper understanding of the contextual dynamics in which they work.

The SNA process can be broken down into three discrete steps:

1. **PREPARATION,**
2. **NETWORK MAPPING & ANALYSIS, AND**
3. **ACTION PLANNING**



STEP 1

Preparation

In the preparation stage, we define the challenge or opportunity to be investigated. For example, under the case study regarding an understanding of LGBT content's online censorship, we would like to determine the major online players or organisations that work closely with LGBT content online such that we can start engaging with the specific organisations who have produced other initiatives or begin learning from past initiatives and strategies these organisations have previously applied to tackle similar issues.

We must clearly discuss this issue and spend more time formulating a solution, as SNA is most useful when a clear and focused question guides the analysis. Finally, at this stage, we identify which data sources most efficiently answer the questions that have been posed.

The time, number of people required, and technology available to complete the data collection and analysis will depend on the familiarity of those undertaking the approach. As a rough rule of thumb, following the open-source methods for collecting data—that is, those methods set out in this handbook—should take around 20–45 minutes to complete and should serve as a strong starting point for the advocates.

Decide what data you will use; SNA can be applied to any data that highlight relationships between, for instance, items, individuals, objects, and events. When looking at LGBT organisations, the approach works best with data that can capture pro-LGBT as well as anti-LGBT links since a wealth of useful information is contained in social links. As such, social network data may be particularly relevant and easily accessible for initiating the research.

STEP 2

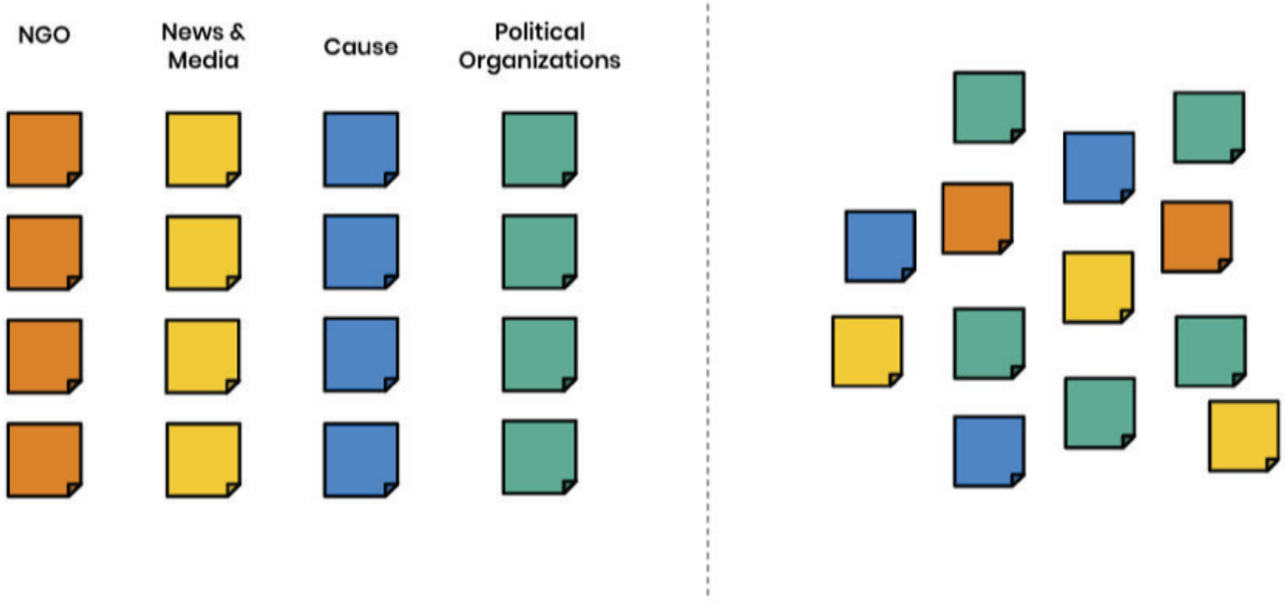
Network Mapping and Analysis

The initial steps herein include the development and analysis of the network map, which can be considered or worked through as concurrent processes. Identifying the most appropriate data source and respective data analysis will be performed while the network map is being developed and will likely inform iterative changes to the map. Under the aforementioned case study, we chose the Facebook data source to investigate the question: 'Which organisations can influence the censorship of LGBT content from looking into Facebook?'

This is a very important step and time-consuming SNA component because it involves several activities, including cleaning the data, categorising and positioning actors, mapping relationships, examining influence, analysing the network, and creating different network scenarios. At this stage, we shall employ the various network analysis software tools (e.g., Netvizz, Gephi) to manipulate the data into different forms as we gather insights to help answer our questions.

1. Select, Categorize, then position actors

To explain this step, we shall refer to our second case study, in which we wish to determine which organisations influence the censorship of LGBT content from Facebook. The first task is for advocates to list all the actors (i.e., individuals, groups, or organisations) whom they consider could have—or already have—some level of influence upon the issue identified. Advocates should answer the question: who can influence…(the issue identified)? Once the list is exhausted, group the actors into categories and assign each category a post-it note colour. Advocates should determine and define the most appropriate categories based on the actors they have identified according to function, organisation type, or political group, among other classifications. For the specific task under the use case, we employed the Gephi software to help categorise the various actors (Facebook pages) into the different categories (NGO, cause, person, government, news and media, etc.).



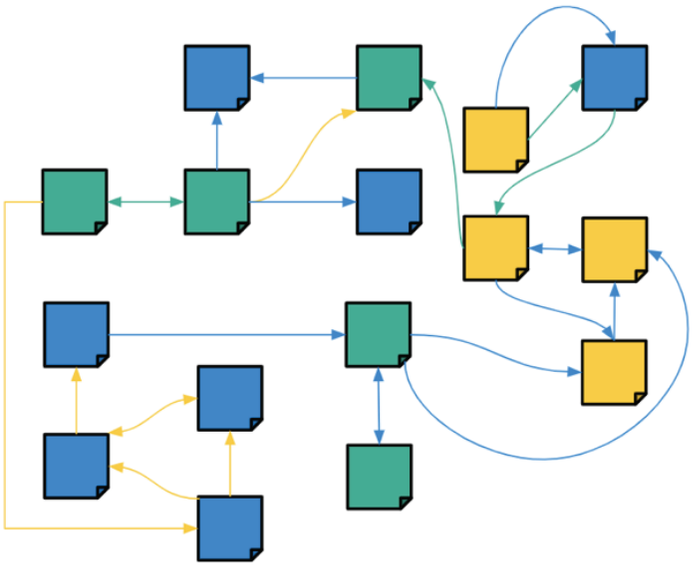
2. Relationship Mapping.

In continuation with the same LGBT case study, many different types of relationships or connections can be mapped on a social network related to the specific actors involved. Connections can be formal (e.g., hierarchy, reporting lines), informal (e.g., friendship, conflict), resourceful (e.g., political, legal, financial, in-kind support, corruption), or informational (e.g., consultation, advice exchange, or data flow).

We should reach no more than four to five—although preferably fewer— types of connections, as shown in the image to the right, such that the map does not become excessively complicated. It is important to be specific regarding the type of connection the advocates are mapping. If the connection's definition is vague or differently understood, then the network map may become confused. It is therefore important to check which advocates understand the connection according to the term defining the relationship type. Each connection type should be allocated a line colour that is indicated in a legend on the flip chart. Connections can be weak (dotted line), moderate (single line), or strong (thick line). As we shall discuss later, the Gephi software tool helps us map all these relations when working with a software tool.

Once the connection types for investigation have been agreed upon, advocates can start to identify the connections that exist between actors by asking the question: how are these actors linked? As connections are identified, advocates draw lines between actors. Connections can be drawn as a one-way line (e.g., hierarchy, reporting line) with one arrow to indicate the relationship's direction or as a two-way line with arrows at either end (e.g., friendship, conflict).

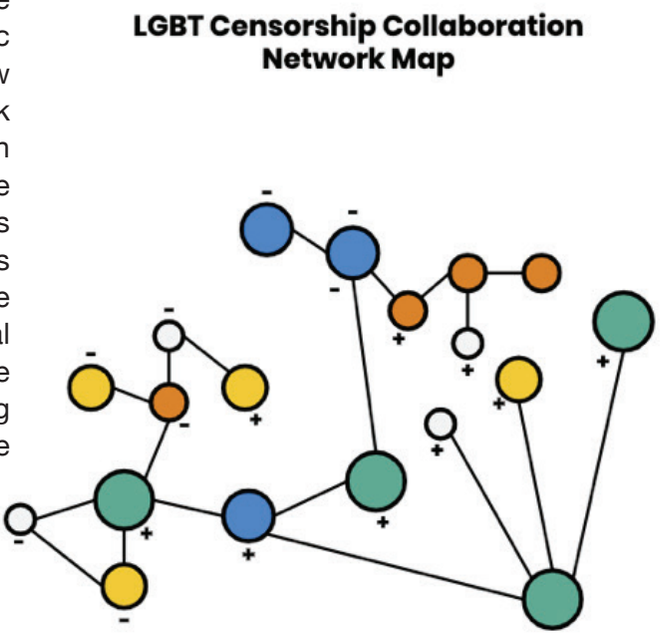
LGBT Censorship Collaboration Network Map



3. Examine Influence.

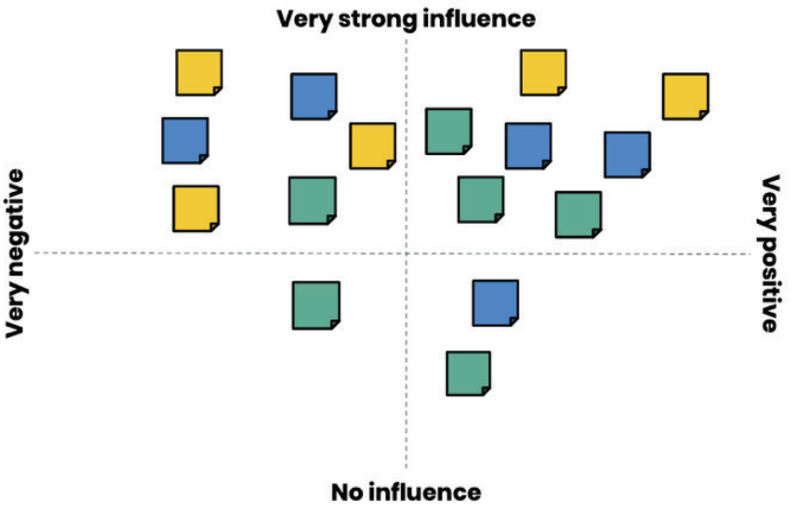
To continue with the LGBT use case, the influence each actor has over each specific question is assessed by determining how common factors vary (e.g., Facebook followers, 'likes', posts). How influential is an actor in relation to the issue? Their influence is ranked according to the number of counters they are assigned; for instance, if a page has 2k followers and another has 1k, then we determine the former page more influential than the latter—the higher up the tower, the more influential the profile. The following scale is recommended to manually sketch the network:

0 – 1 = no/low influence,
2 – 3 = moderate influence,
4 – 5 = high influence.



Advocates should also note that each actor's influence has a positive, negative, indifferent, or mixed effect upon the specific issue. They should accordingly identify these effects using specific tools or card note with + (for positive), - (for negative), or +/- (for mixed/indifferent) to aid the interpretation of the findings upon the analysis's completion.

Advocates can choose to develop an influence and positivity grid, which may help them assess who they would like to work with (those who are positive and influential), where the risks lie (those who are both negative and influential), and how they can start building coalitions of support or strategies to mitigate the negative influence of specific actors. Once the grid is developed, advocates are encouraged to develop strategies for engagement with each actor.



4. Analyse the network

Advocates should analyse the relationships between actors, making sure they consider their location within the network and the overall network structure. The following questions may help analyse the network structure:

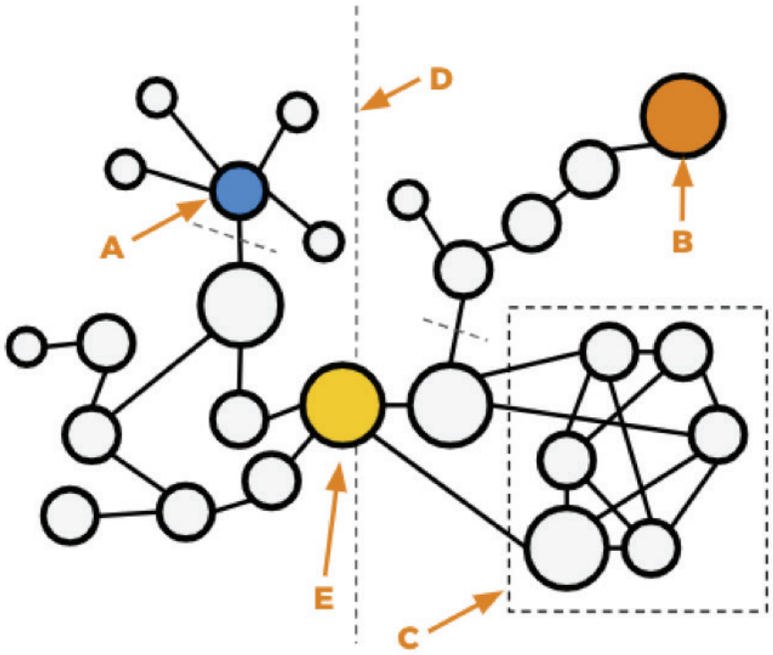
A) Do any actors have a high number of connections?

B) Do any actors appear peripheral to the network?

C) How centralised or interconnected is the network?

D) Are there any fault lines between or separate parts of the network?

E) Do any actors link together significant parts of the network?



When analysing their networks, advocates should look out for several common issues, of which we dive into a few below.

Dependency: The network may be highly dependent upon a single actor or source, which can create bottlenecks and sustainability concerns.

Dysfunctional/conflicting relationships: Some key broken relationships might impede the entire network. New actors or interventions can also introduce conflict for resources or control.

Marginalisation: Certain actors or groups of people may be excluded or marginalised within the network, perhaps owing to location, budget, age, gender, ethnicity, status, income, or other factors. An analysis of the reasons behind the network structure may help uncover the reasons for and how to best overcome marginalisation.

Disincentives for change: Certain actors may have disincentives to support the proposed change and may try to actively oppose it. Advocates should pay particular attention herein regarding how the intervention would change the resource flows or the level of each actor's influence.

'Like me' relationships: Advocates may notice that actors (people/pages/organisations) who share certain attributes, such as political affiliation, nationality, gender, age, education, ethnicity, religion, or status, tend to have many 'like me' relationships and fewer relationships with people who are different from themselves. This is a common pattern in many networks, and considering how it affects the specific issue may be worthwhile in determining how to overcome it.

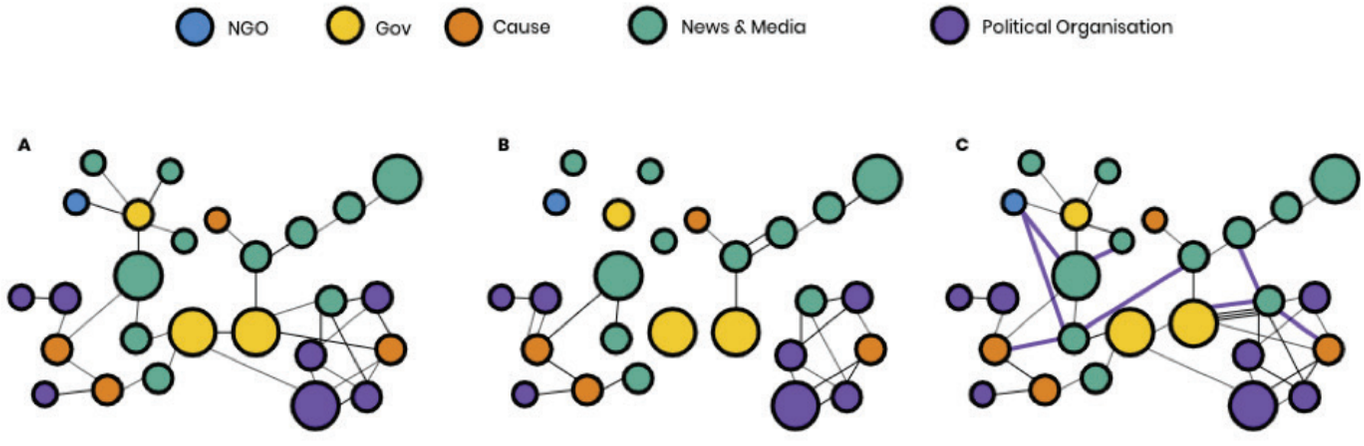
Structural challenges: Structural risks may include an overly centralised network or a structural split within the network.

5. Network scenarios

SNA changes frequently, as the data sources from most social networks are always different dependent on the time/day/month, etc.; for example, a hashtag trending today, such as #KeepItOn, will achieve different results in a month's time. It is therefore useful that advocates consider analysing what the network looks like now versus how it might change in the future. Advocates may wish to consider how different scenarios might affect the network, such as:

- **What would the ideal network look like and how could this be brought about?**
- **What would happen to the network if one actor were to be removed?**
- **What would happen to the network if followers were reduced from a specific actor?**

It is therefore helpful to consider in which relationships one should invest according to the best- and worst-case scenarios across the network.



STEP 3

Action planning

Analysing and validating findings is quite an important part of the action planning step. SNA can only reveal what the data reveal and does not provide all the context or details surrounding the data. The big picture may be incomplete or misleading in some places, while certain activities may be more visible than others, thus skewing the big picture; for example, pro-LGBT organisations may be more visible in the data than anti-LGBT organisations although both activities/ actors may well be present. For this reason, it is important to validate findings against operational experience.

After we have summarised and evaluated the findings, they must be presented and put into action— hence the need for action planning. The developed plans will need to be tailored to the specific issues identified during the SNA.

Examples of the types of questions to ask include:

- Do the findings match what is known?
- Does anything seem unusual?
- Can any unusual results be explained by issues with the data?

Validating the data in this way not only helps assure the findings' quality, but can also introduce
interesting aspects about the data for further exploration.

However, advocates may wish to further conduct a basic SWOT analysis of the network in order to help inform and prioritise their next actions. For example:

Strengths: a number of influential actors are positive towards the issue;

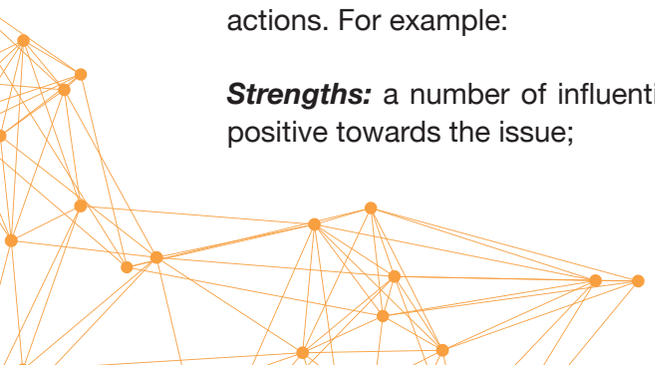
Weaknesses: a number of negative actors block the actions of those who support the issue. There is limited interconnectedness between parts of the network, the issue is dependent upon one actor or specific funding from one source, and sustainability is questionable.

Opportunities: There exists significant, untapped positive support with the potential for forming coalitions or formal partnerships.
Threats: The network is highly dependent upon a few key actors. If their influence was to wane or relationships were to break, the network would be severely disrupted.

Strategies

Once the specific opportunities and risks related to the network have been identified, DRAs and other participants should seek to develop strategies and practical action plans; for example:

- They may see the opportunity to change or extend program activities to capitalise on an opportunity or help realise existing objectives.
- They may consider what incentives might be introduced to manage the negative impact certain actors have on an intervention. They may want to assign specific responsibility to a team or organisation to manage a problematic relationship.
- They may act as facilitators to improve dysfunctional relationships or raise the influence of those positive about the intervention.
- They may consider changes to partnership arrangements or the focus of those partners with whom they engage and consider how best to challenge inequality and marginalisation.
- They may consider exit strategies that build network relationships.

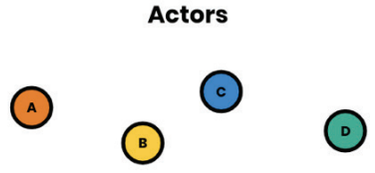


TERMINOLOGIES IN SNA

We must define some terminologies in order to use a consistent language when talking about social networks within this handbook.

1. ACTOR:

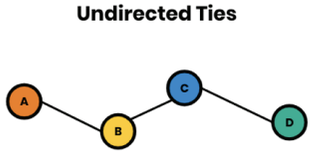
also called a node or a vertex, refers to an individual that can form relationships with other individuals—in this case, an individual or group of individuals we have chosen to study.

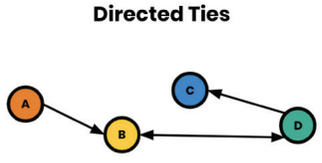


2. TIE:

also called a relation or edge, describes a particular, well-specified relationship between two actors, such as 'is part of the same political party', 'likes conservative views pages', or 'likes' or 'trades with'. Ties can be undirected (e.g., 'is part of the same political party'), in which case the relationship holds the same meaning for both actors.

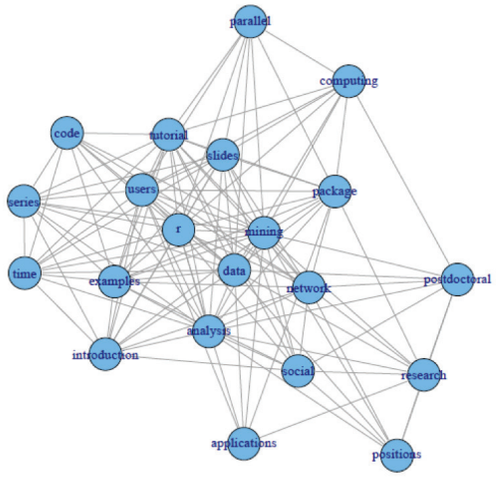
Ties can also be directed (e.g., 'reports up to') and either one directional or bidirectional.





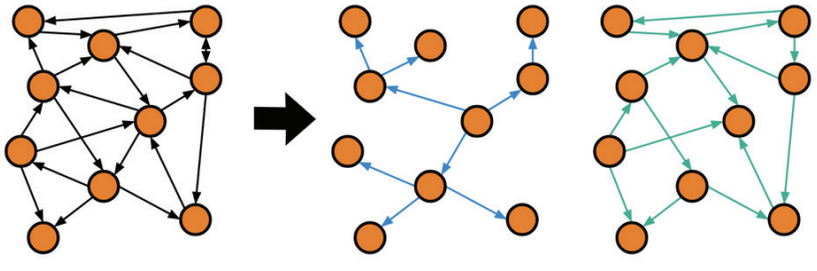
3. NETWORK:

also called a graph, particularly refers to a collection of actors and the ties between them. The figure below depicts a set of undirected relationships between text related to a blog post about data mining and analysis.



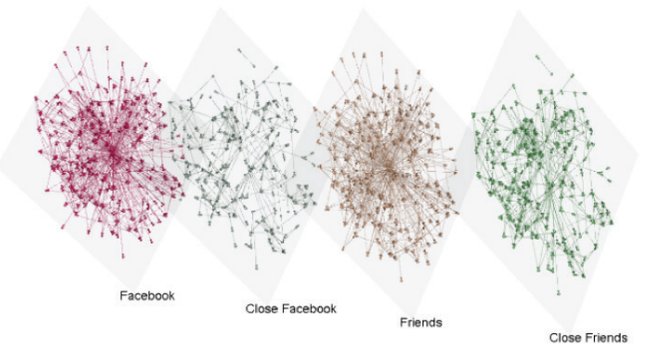
4. MULTIPLEX NETWORKS:

networks where more than one kind of tie is present; for example, if we were to collect information about several relationship types between political party leaders (e.g., 'comes from part of the country', 'supporters', 'is friends with', 'is close friends with', 'works for'), then we



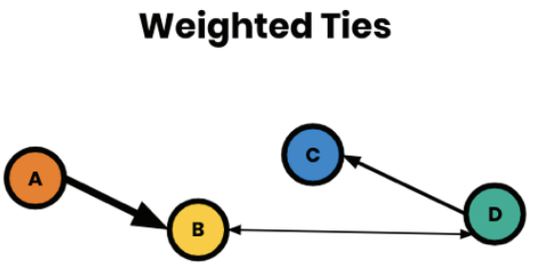
essentially end up with a network containing multiple tie types between actors.

In a practical Facebook network, a multiplex network would look like the image to the right.



5. WEIGHTED TIES:

just as networks can contain multiple kinds of edges between actors, they can also contain relationships of varying strength; for example, A might like B a whole lot, but B and C only like each other moderately.

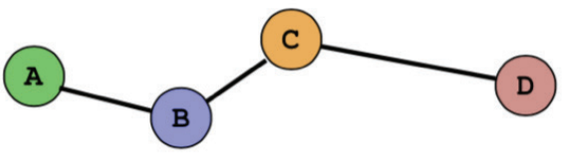


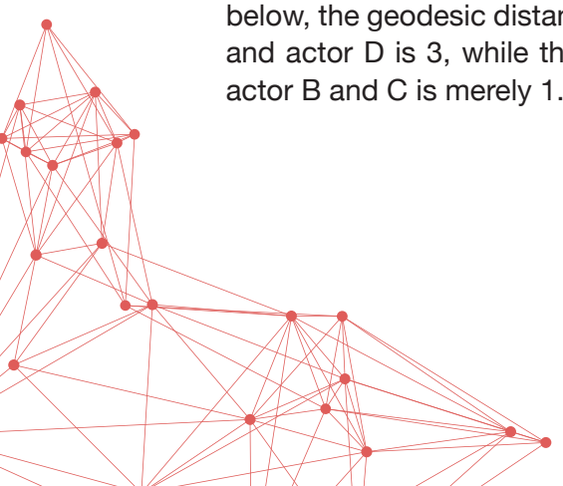
6. GROUP:

A group in a network is merely a subset of the actors who share some characteristics. If we were to look at a parliament network, one group might be made up of all the actors involved in the same political party. The definition of groups as a commonality on some salient trait allows us to examine a number of network hypotheses and define useful measures that (A) are conditional upon knowing actors' group membership; for example, we might want to test (B) a hypothesis concerning the number of friendship ties between politicians in a country who are part of different political parties versus those in the same political party.

7. GEODESIC DISTANCE:

the least number of connections (ties) that must be traversed to get between any two nodes; for instance, in the network depicted below, the geodesic distance between actor A and actor D is 3, while the distance between actor B and C is merely 1.





SOCIAL NETWORK DATA

Two main kinds of social network data exist: edge lists and sociomatrices. Each of these data formats has its own advantages and weaknesses, mainly concerning a trade-off between ease of entering, data storage, and ease of use for data analysis. For this specific handbook, we won't get into the details because they are pre-sorted for us when we download pre-formatted data sets from the well-known social networks (e.g., Twitter, Facebook).

1.

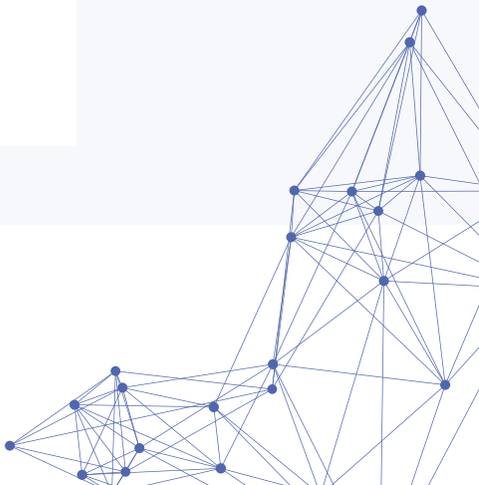
SOCIOMATRIX:

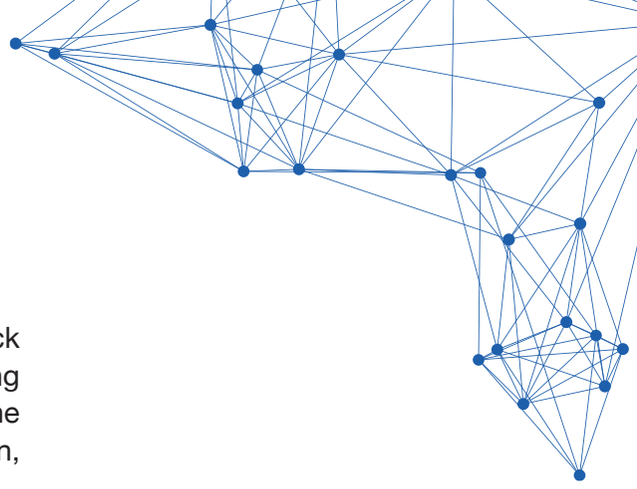
(also known as an adjacency matrix): a way of representing directed or undirected ties between actors using a numerical matrix, with one column and one row for each actor. In general, the matrix's diagonal elements (e.g., second row, second column) are always equal, thus signalling that actors do not tie to themselves. This is the data we shall be employing for most of our case studies within this handbook.

2.

EDGELIST:

exclusively captures information about existing ties and thus must be supplemented with knowledge of the total number of actors in the network (even if they do not have any ties). This form of data entry is most efficient for storing information about data that are collected by hand because they are very efficient to store and relatively easy to enter; however, one must be careful to use a common naming system and keep track of any nodes that do not have any ties to them.





PROJECT CASE STUDIES

SNA has the main ability to help practitioners and advocates track and understand relationships at a variety of levels when working on their development programs. If they want to understand the extent and types of social relationships they are operating within, then conducting an SNA can provide the necessary insights.

We all know relationships matter across our day-to-day activities– including those with our colleagues, friends, and family. We also know that in development programs such as digital rights advocacy, relationships inside an organisation or between organisations can mean the difference between smooth flying and turbulence when implementing a program. Thus, how do we ensure we are able to leverage relationships in our work rather than become buffeted by them? In the next pages, we dive into three practical scenarios or use cases that have been selected from projects on which the Internews DRAs are currently working. These case study analyses should guide advocates on how to get started with SNA in their programs.

We walk advocates through the three chosen case studies: counter internet shutdowns, censorship of LGBT content online, and advocacy for the Digital Rights and Freedom Bill.

	Topic	Data sources	Outputs	Tools
1	Counter internet shutdowns	• Twitter	• Sentiment analysis • Topic clusters	• Gephi • Tags
2	Censorship of LGBT content online	• Facebook	• SNA map • Group and page actors	• Gephi • Netvizz
3	Advocating for the Digital Rights and Freedom Bill	• PDF or HTML Digital Rights and Freedom Bill drafts	• Word clouds • Text patterns	• Voyant Tools

CASE STUDY #1

Counter internet shutdowns

Internet shutdowns occur when governments order companies to terminate or degrade public access to digital communications tools, such as Twitter, SMS, or Facebook. Calculating the broad and grave impacts that shutdowns inflict upon human rights remains beyond the capacity of most observers although sustains considerable challenges for citizens.

In recent years, States have adopted myriad laws to regulate content online by placing increasing pressure on private actors to censor content that they deem illegal or simply 'harmful'. Much of the world's online content is now regulated by the community standards and algorithms of a handful of Internet companies whose operations and processes lack transparency. Internet users' right to free expression is easily subject to abuse in this regulatory environment, the complexity of which is compounded by the fact that the internet is a public space built on decentralised private infrastructure.

Several activists across the globe have curated hashtags to follow or collaborate on issues with which they are concerned. For example, the hashtag #KeepItOn has been used for internet shutdown discussions and advocacy. Just what is an internet shutdown, and why need advocates care about the hashtag #KeepItOn? Herein, we seek the answers to these questions, the facts behind the campaign, and ways digital advocates can make an impact.

We shall look into how this data may be found from historical and trending topics related to internet shutdowns and identify countries from Tweets that have been affected by such shutdowns. For this case study, Twitter is our primary data source.

Twitter historical data importation

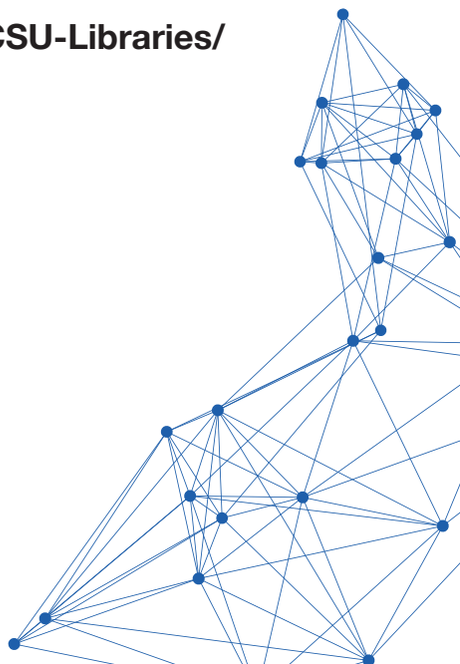
As a first step, we shall look into mining old historical data (Tweets) that are related to the #KeepItOn hashtag to help us answer the questions posed above. For this case study, we employ an open-source tool called TAGS.

TAGS is a free Google Sheet template that allows users to set up and run an automated collection of search results from Twitter.

Tool:
https://tags.hawksey.info/

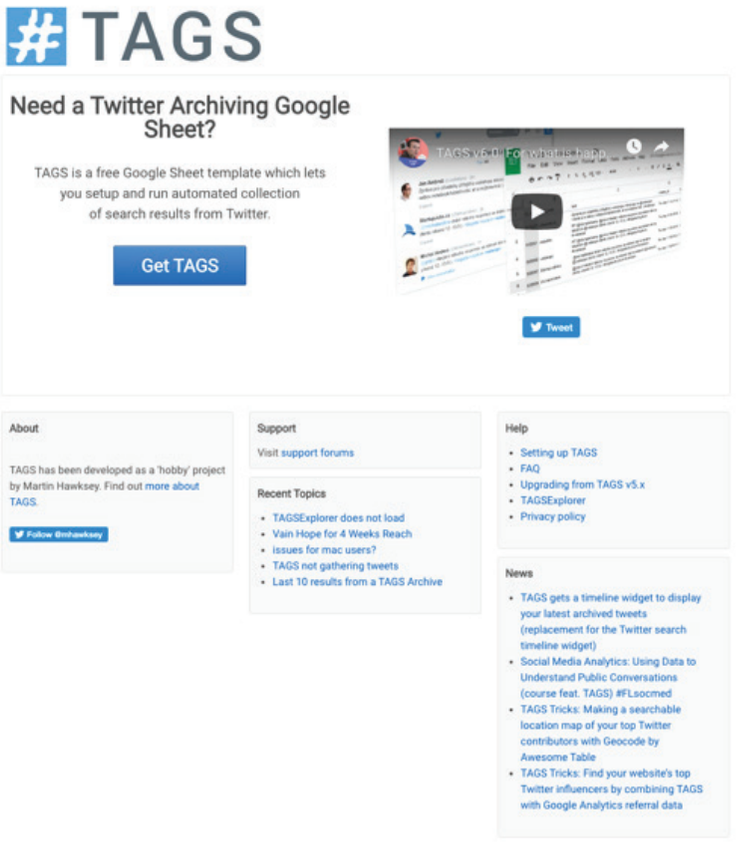
Twitter APIs:
https://gwu-libraries.github.io/sfm-ui/posts/2017-09-14-twitter-data

App:
https://github.com/NCSU-Libraries/twitter-shiny-app



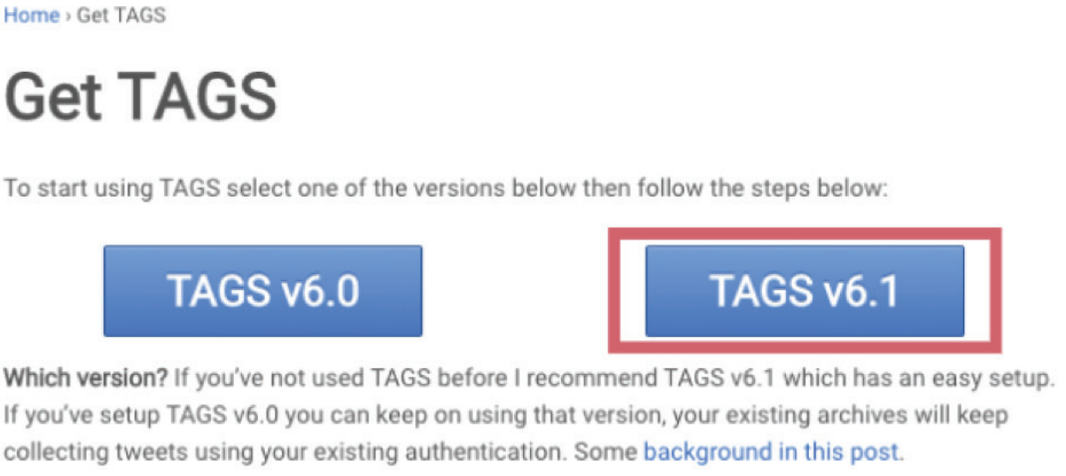
STEP 1:

go to https://tags.hawksey.info

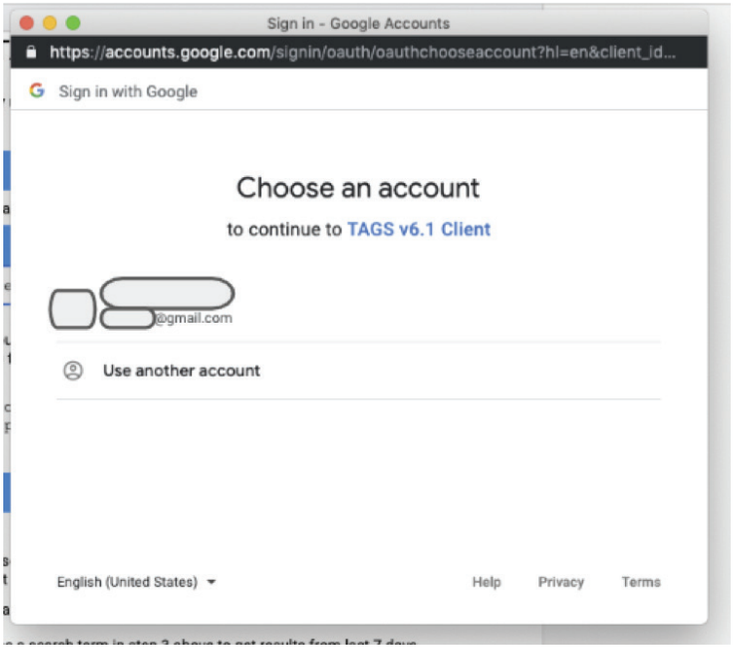


STEP 2:
Select TAGS v6.1

Here, we select the TAGS v6.1 button, which is easier to set up with Twitter. We do assume here that the advocates already own a Twitter account. After clicking on the v6.1 button, you'll be prompted to log into your Gmail account such that a copy of the spreadsheet may be stored in your Google Drive.



You will need to login with your respective Gmail or Google account.

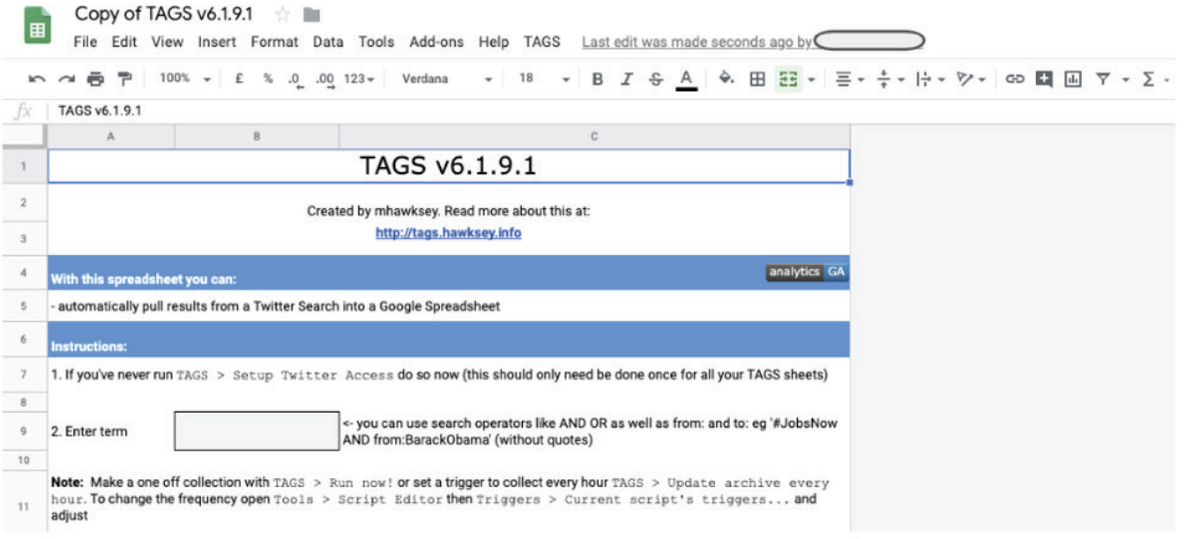


STEP 3:

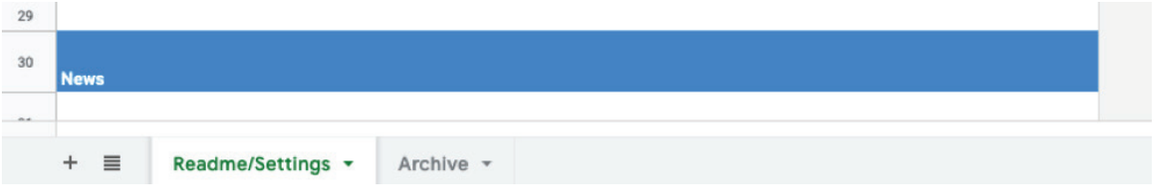
Click 'Make a copy'

STEP 4:

A copy should appear in your Google Drive, as shown below

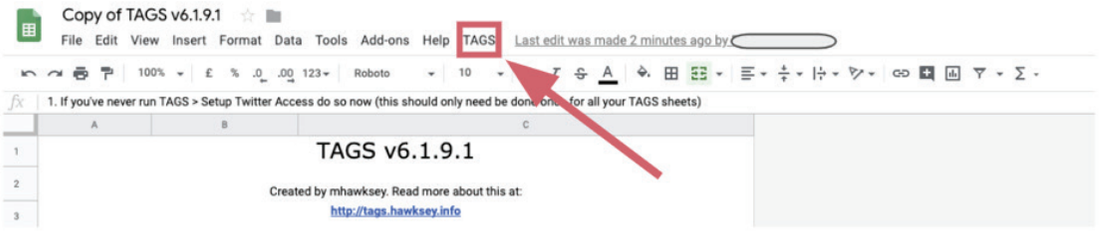


You should currently be positioned at the 'Readme/Settings' tab, as shown below.



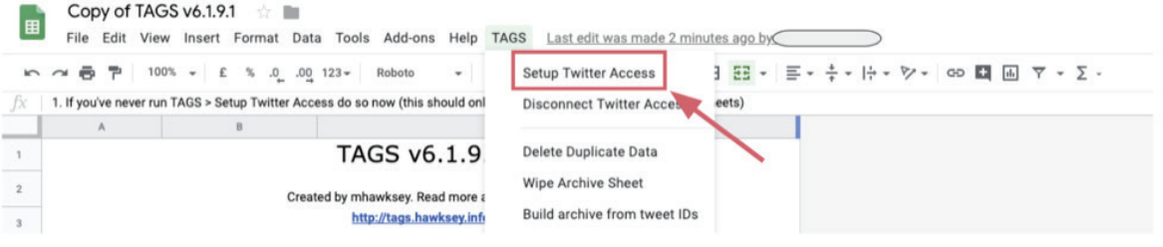
STEP 5:

Click on the 'TAGS' menu item

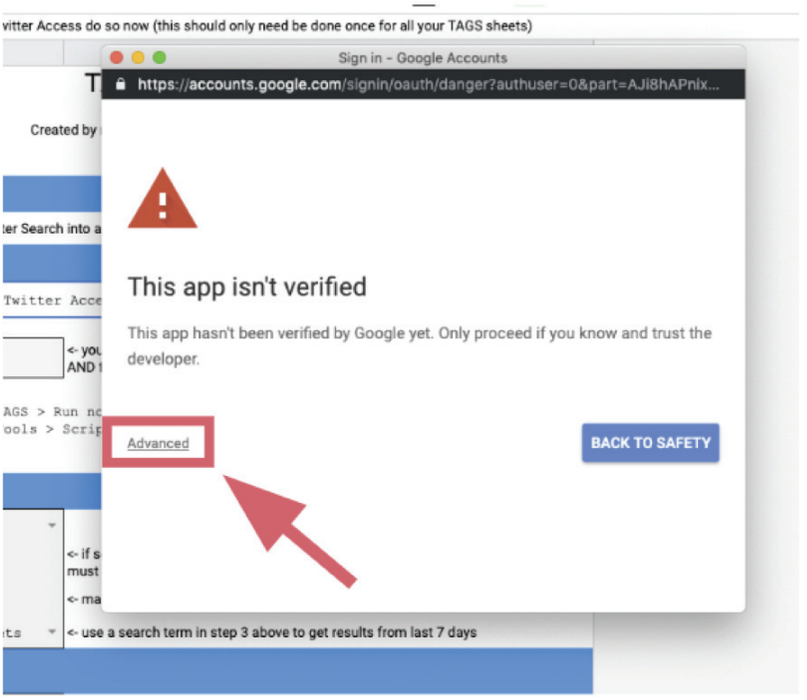


STEP 6:

Setting up Twitter



You'll receive the message box shown below. Go ahead and select 'Advanced'.

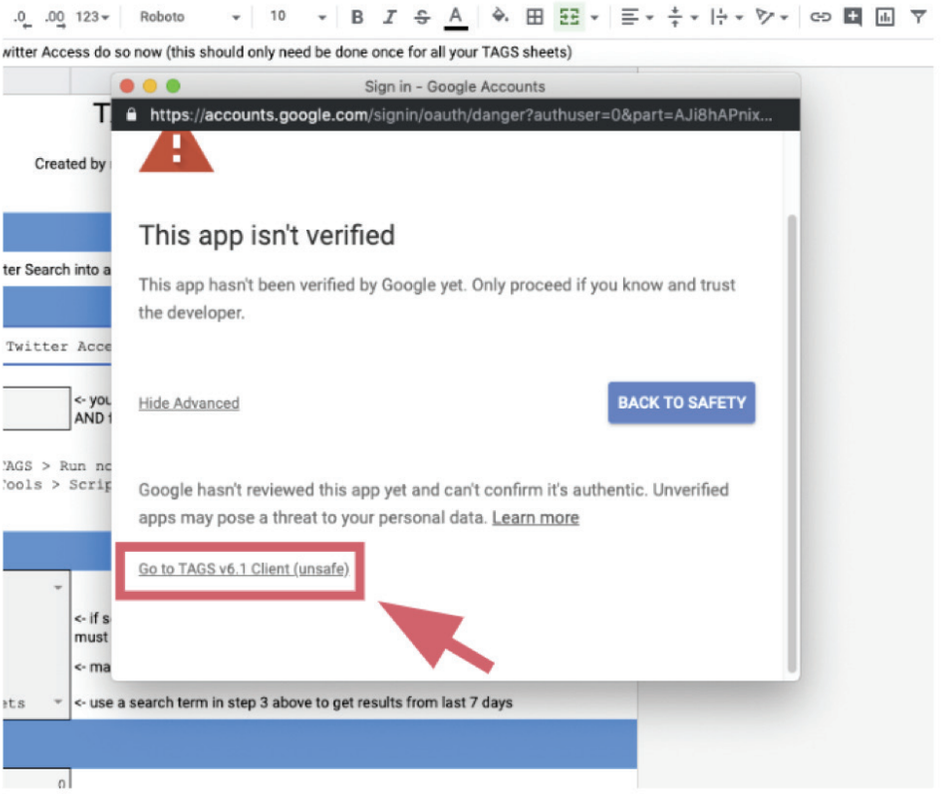


Note: 'This app isn't verified' notice

Google recently made changes to the authorisation process. Currently, TAGS is not a verified application, and the way it is currently distributed makes verification impossible. You can continue using TAGS by clicking 'Advanced' on the 'This app isn't verified' screen and following the prompted instructions.

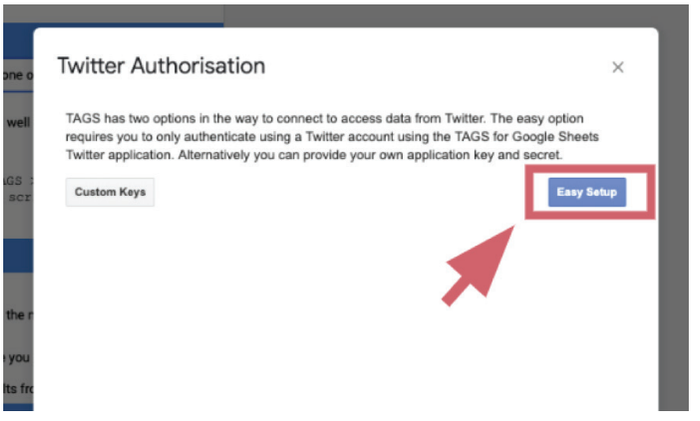
STEP 7:

Click 'Go to TAGs v6.1 Client (unsafe)'



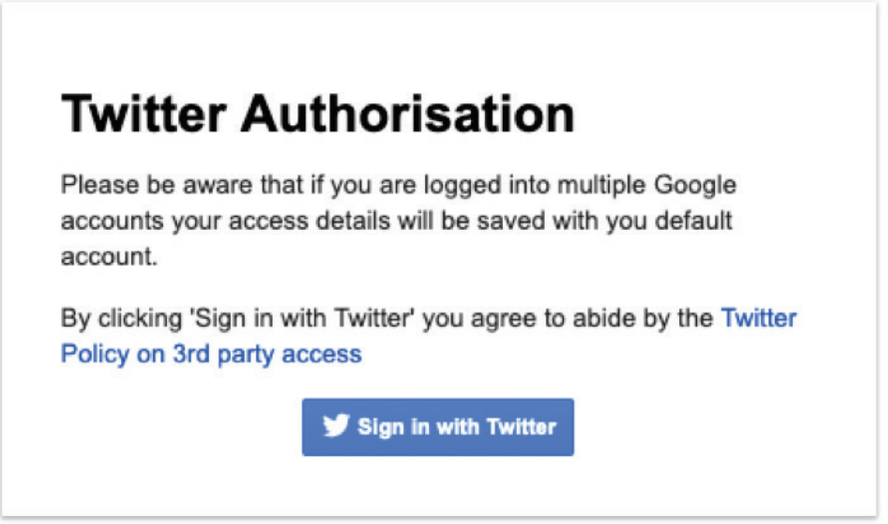
STEP 8:

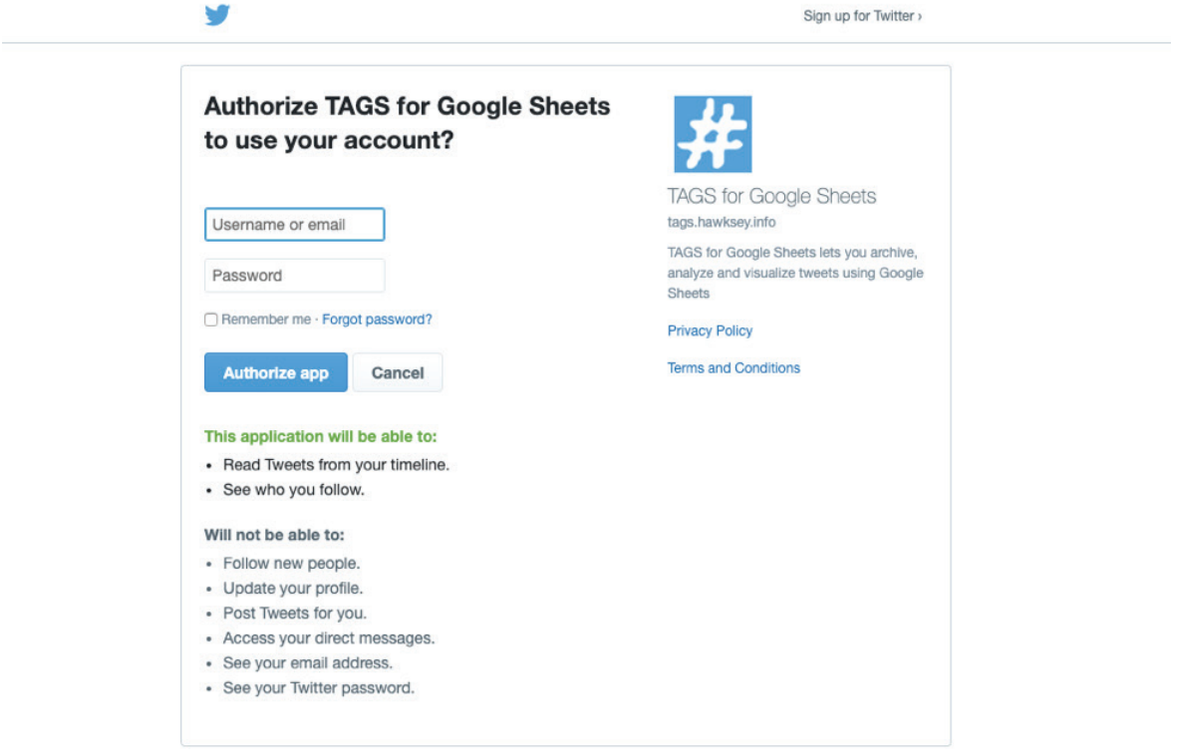
Twitter authorisation steps



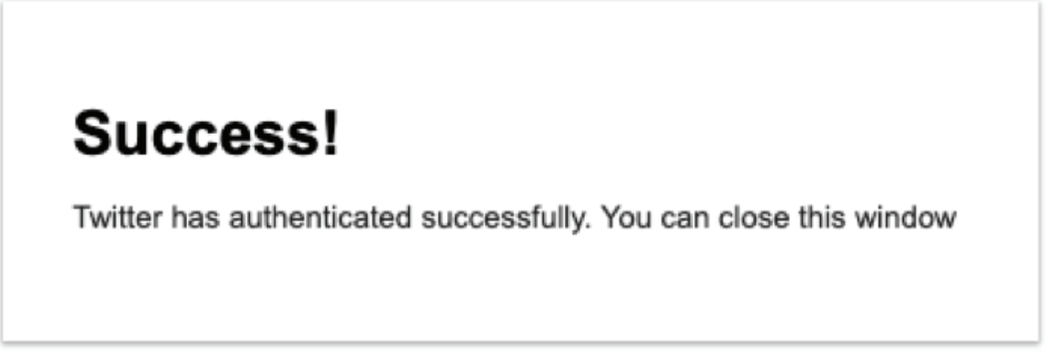
STEP 9:

Log into Twitter with your personal credentials



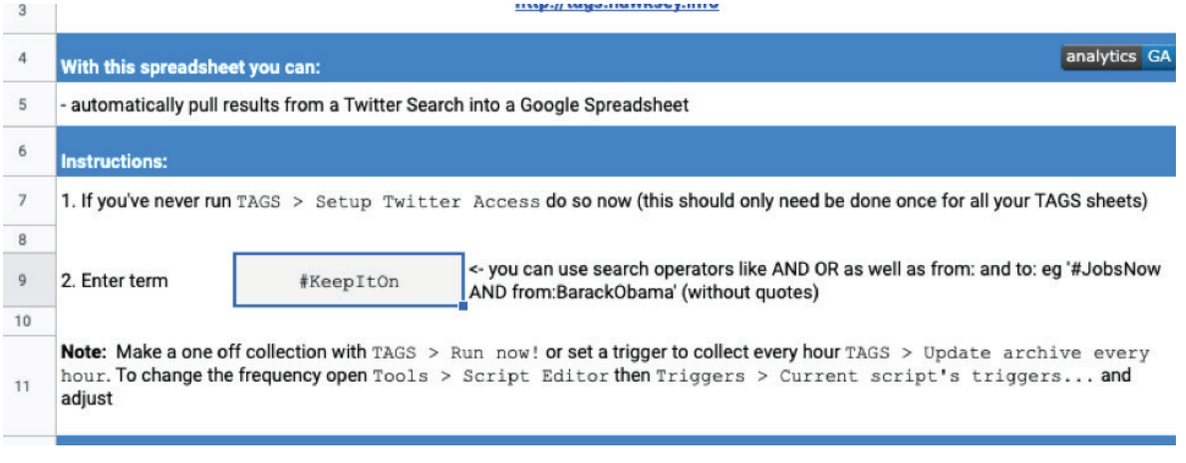


Finally, you should achieve the 'Success!' message pictured below, which indicates the setup is complete.



STEP 10:

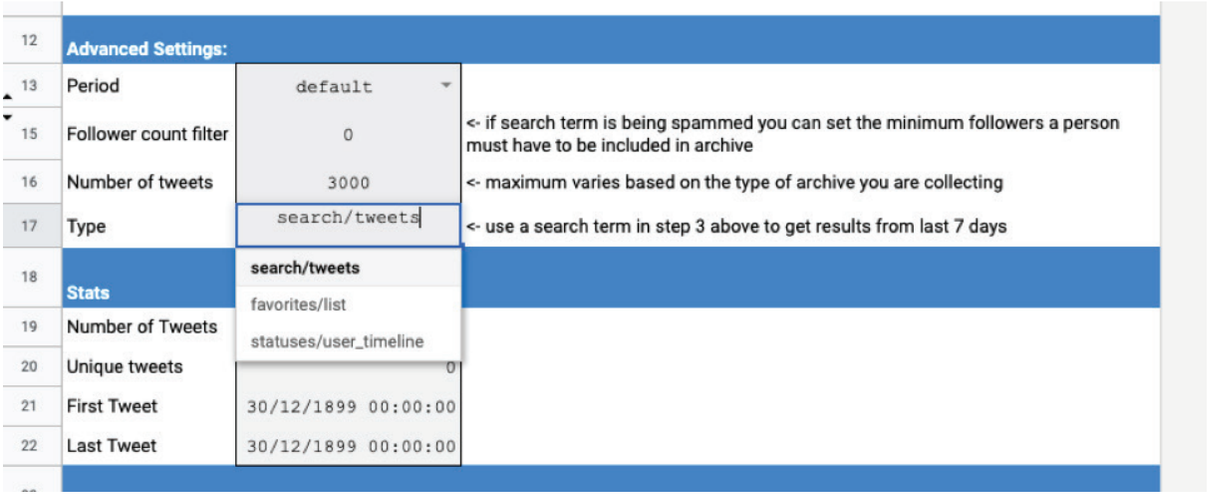
Set up the hashtag or search term to scrape



STEP 11:

Select 'search/tweets' from row 17: 'Type'

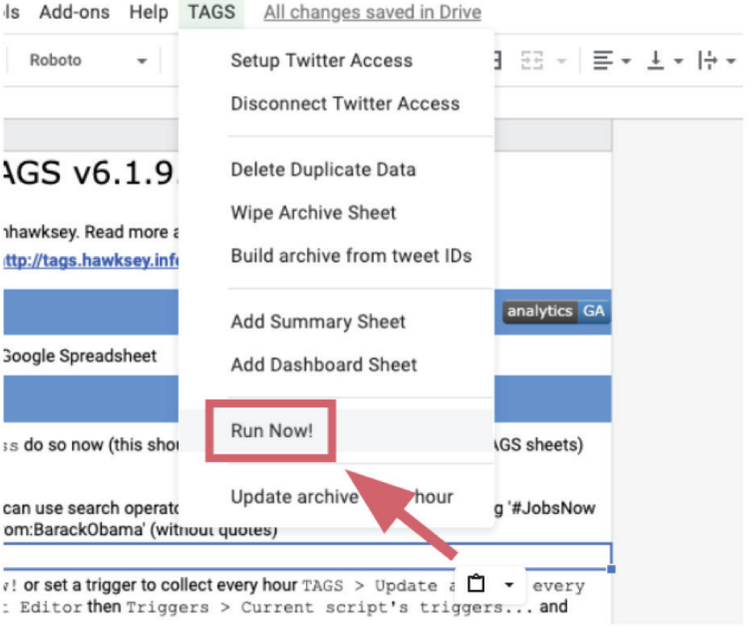
Here, we inform the tool that we wish to run a search for the keyword 'KeepItOn' in order to scrape related tweets across.

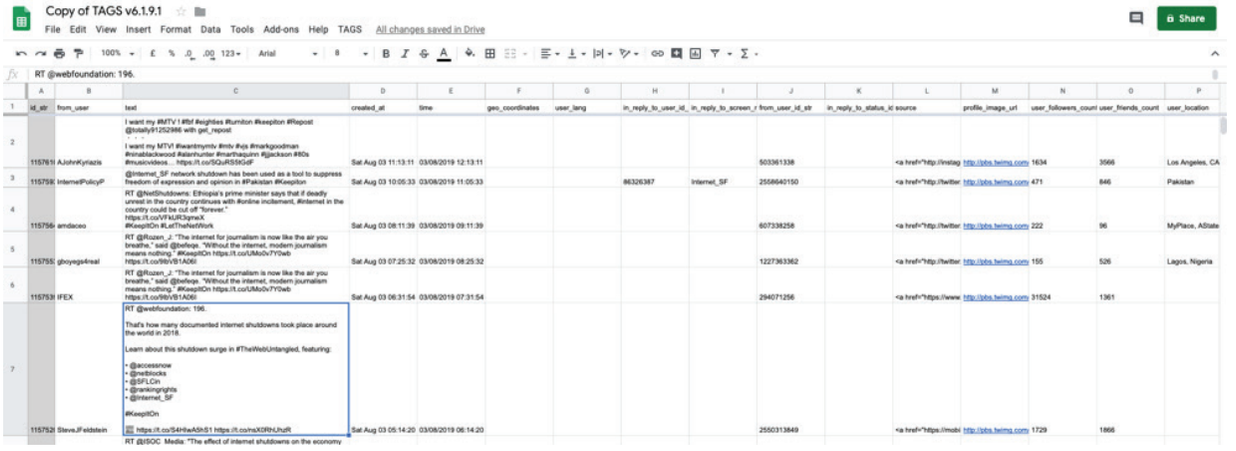


STEP 12:

Select 'Run Now!'

This selection will run the script in the back and start mining or downloading the Tweets under the 'Archive' sheet or tab.

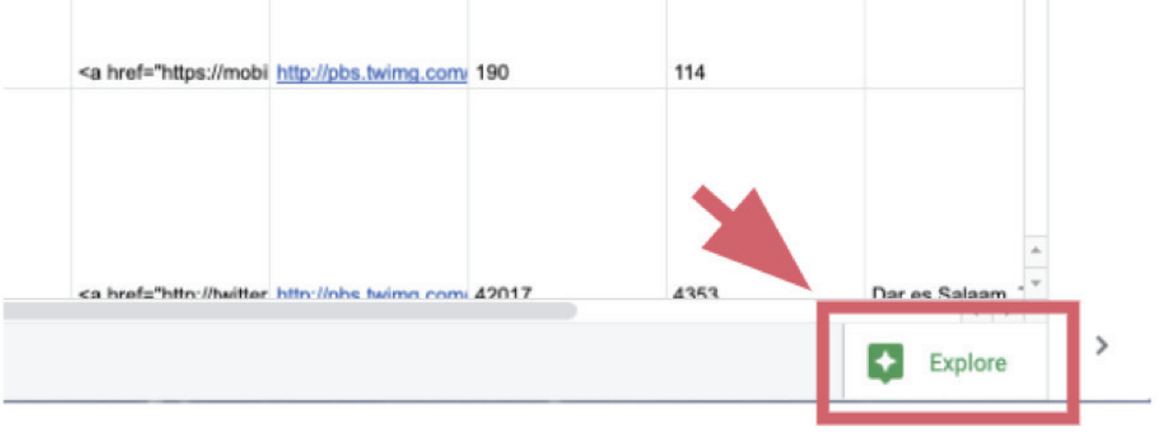




Explore the data

To further maximise and easily understand the Twitter social network data, we shall employ 'Explore' in Sheets, which is powered by machine learning and helps teams instantly gain insights from the data. Simply ask questions—in words, not formulas—to quickly analyse the Twitter data. For example, you can ask 'what is the follower count distribution?' or 'what is the user_friends_count vs time?' and Explore will help you find the desired answers.

Now, we're using the same powerful technology in Explore to make visualising data even more effortless. If you don't see the chart you need, just ask; rather than manually building charts, ask Explore to do so by typing in 'histogram of followers' or 'pie chart for source', as shown in the examples below.



Summary of findings from the data

Using the Explore tool, we can share some quick findings from the data by answering the few questions below:

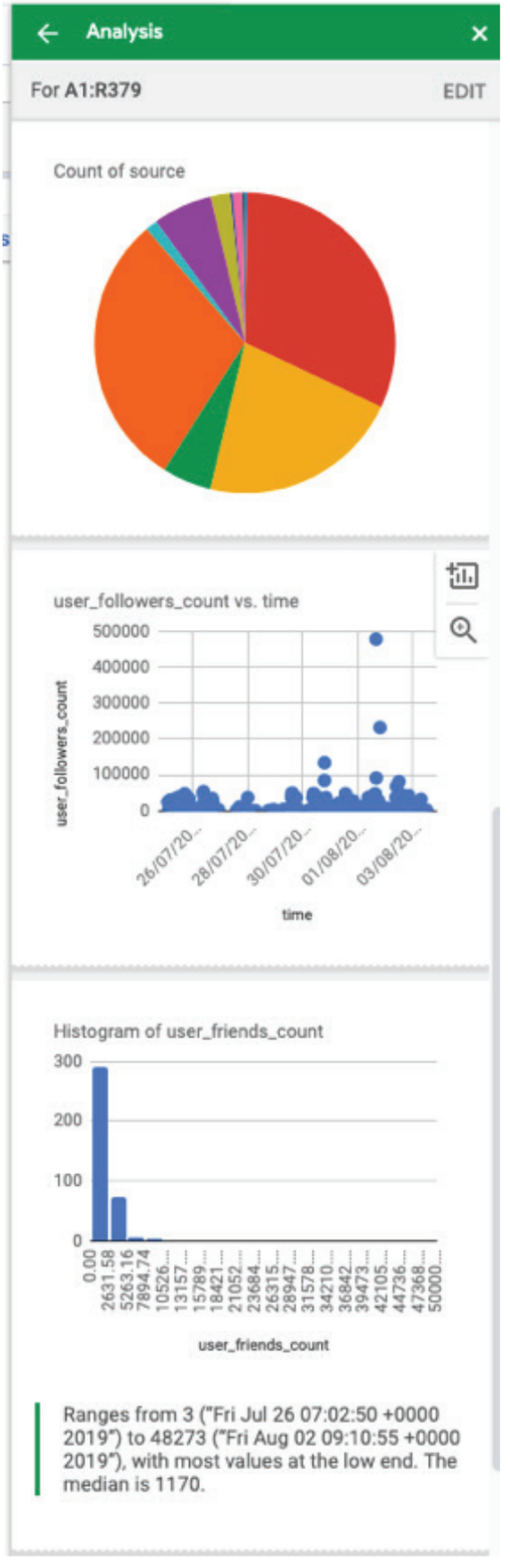
- **How many countries seem to have shutdowns?**
- **When do you think these shutdowns happened?**
- **Which accounts or users seem to have this information?**
- **How influential are the users according to their number of followers?**
- **Are the accounts made up of organisations or individuals?**

We observe the graph that displays the source and respective number of times the users Tweeted under 'COUNTA of id_str' for each source. Also, we have a quick view of a correlation between the source 'user_ friends_count vs. user_followers_count', which displays a few outliers that would prove interesting to further investigate.

Finally, this graph illustrates the number of sources across the downloaded Tweets. Here, we share a few graphs that were created without any effort in order for the advocates to start understanding the data. They will of course need to further analyse these graphs to answer additional questions. We can see from the data that the organisation handles shared below seem to be the main sources mentioning the #KeepItOn hashtag. Please note that the results may vary depending on when advocates download their data.

- **@accessnow**
- **@netblocks**
- **@SFLCin**
- **@rankingrights**
- **@Internet_SF**

In addition, the frequent mentions of countries such as Tanzania, Ethiopia, and Spain would be indicators of recent shutdowns related to the time the data were downloaded.



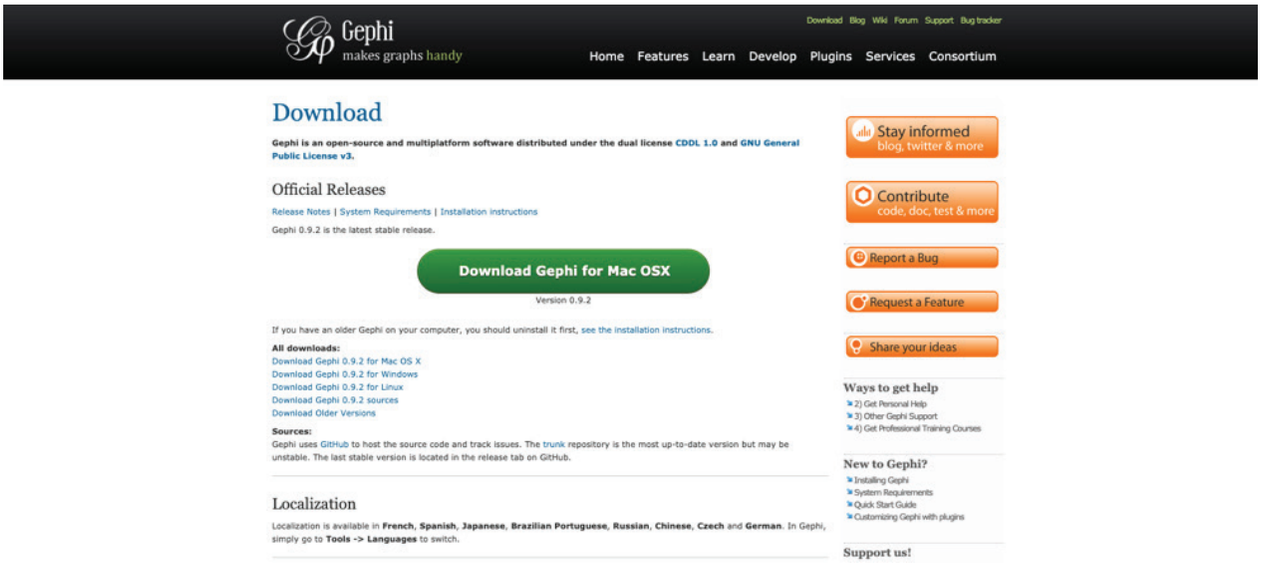
Twitter streaming or real-time importation – trending topics

As an additional step, we would like to analyse all Tweets that are streaming in real time or are currently trending. This process can come in handy on the exact days or time periods precisely when these shutdowns occurred. Here, we shall use the Gephi tool alongside an additional plugin called 'Twitter Streamer' to help collect the data in question.

Downloading and installing Gephi:

Download guide:
https://gephi.org/users/download

Installation guide:
https://gephi.org/users/install



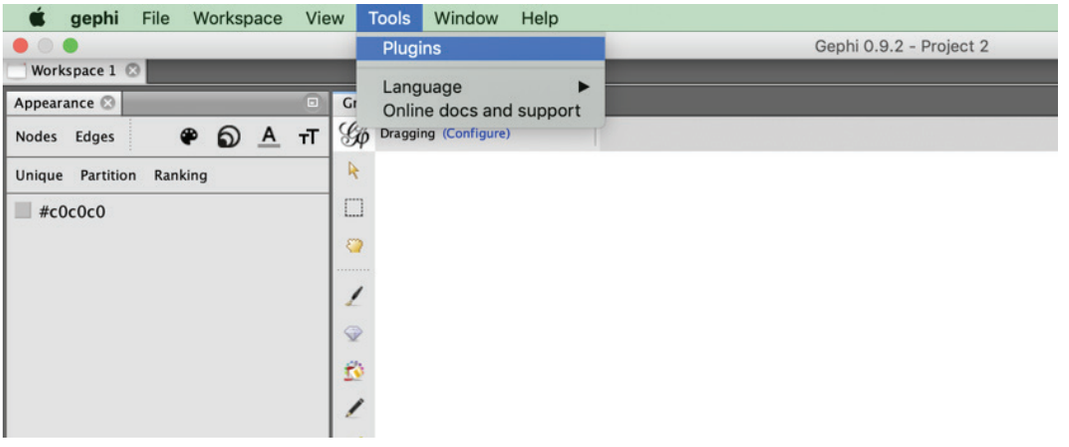
Gephi:

For a detailed video that explains the process advocates are encouraged to follow, review this YouTube Twitter plugin video: **https://bit.ly/2NhB3jI**

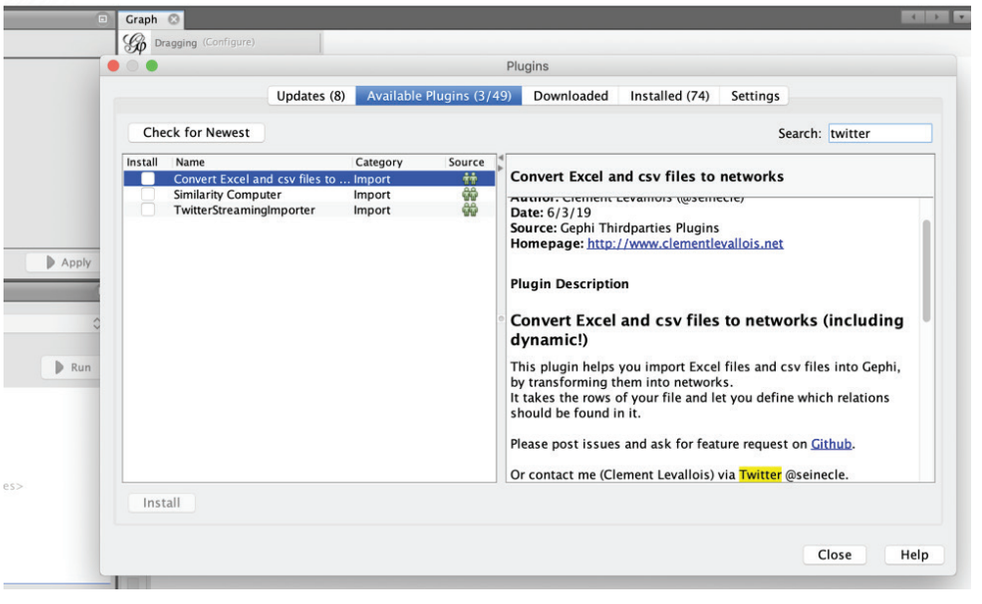
This plugin enables users to:

- **Collect Tweets in real time on their topic of choice**
- **Retrieve the connections between the users mentioned in these Tweets**
- **Visualise these connections in Gephi or simply export all the Tweets to Excel**

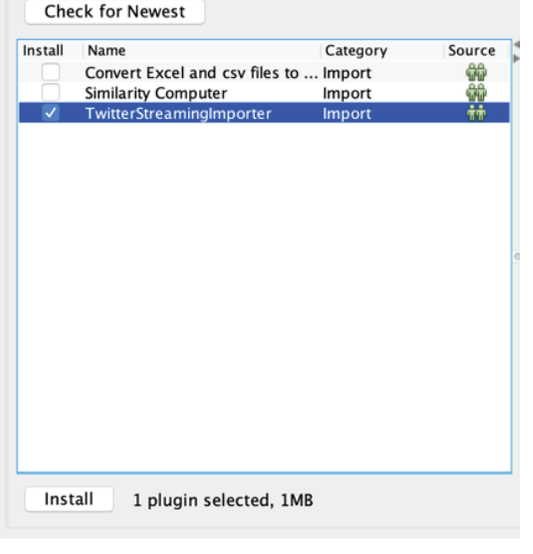
After installing Gephi, click on Tools > Plugins, as shown below.



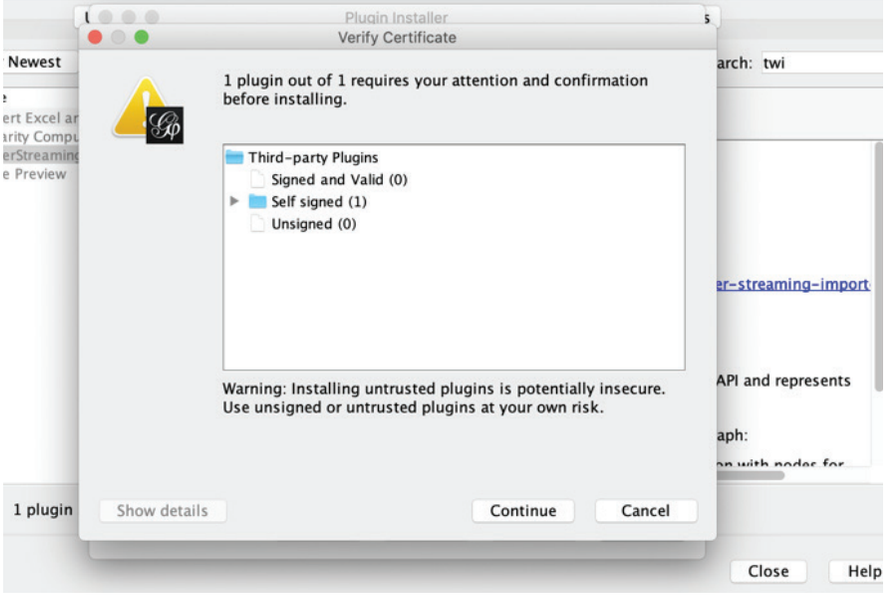
Select 'TwitterStreamingImporter', as shown below.



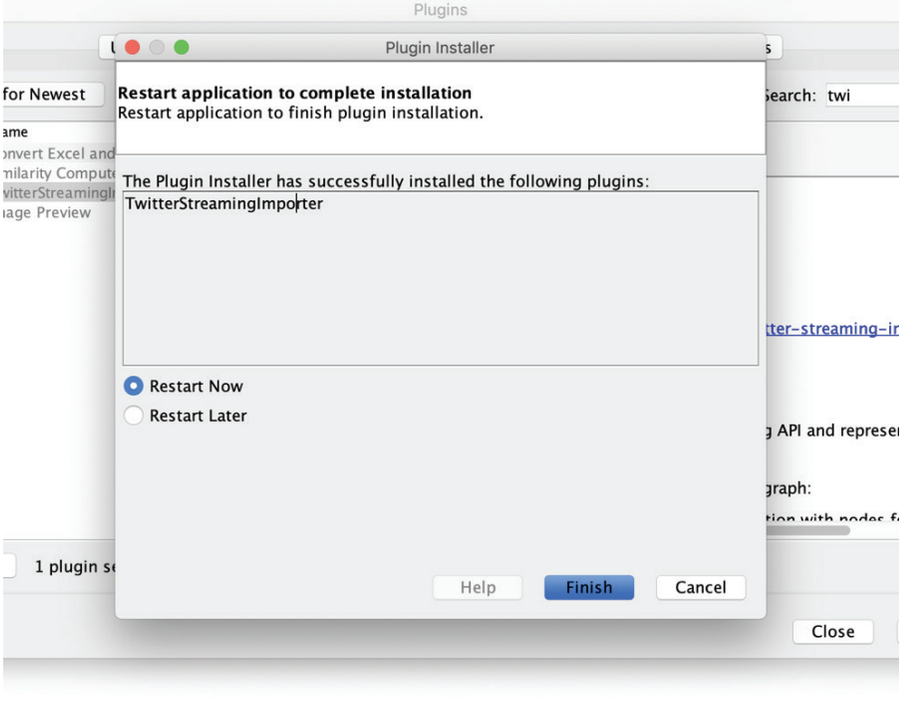
Click 'Install' to have the plugin installed locally.



Ignore the prompted messages and merely click 'Continue', as shown in the dialog box pictured below.



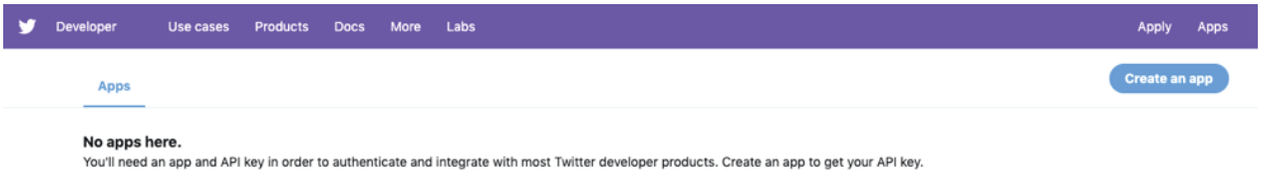
Doing so should install the toll. Next, click 'Finish. Make sure you have the 'Restart Now' radio button selected to restart the Gephi tool with the new plugin just installed.



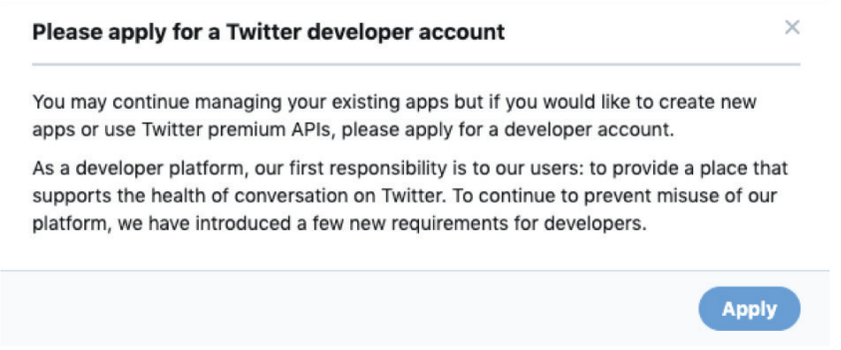
Applying and setting up Twitter API data access

https://developer.twitter.com/en/apps

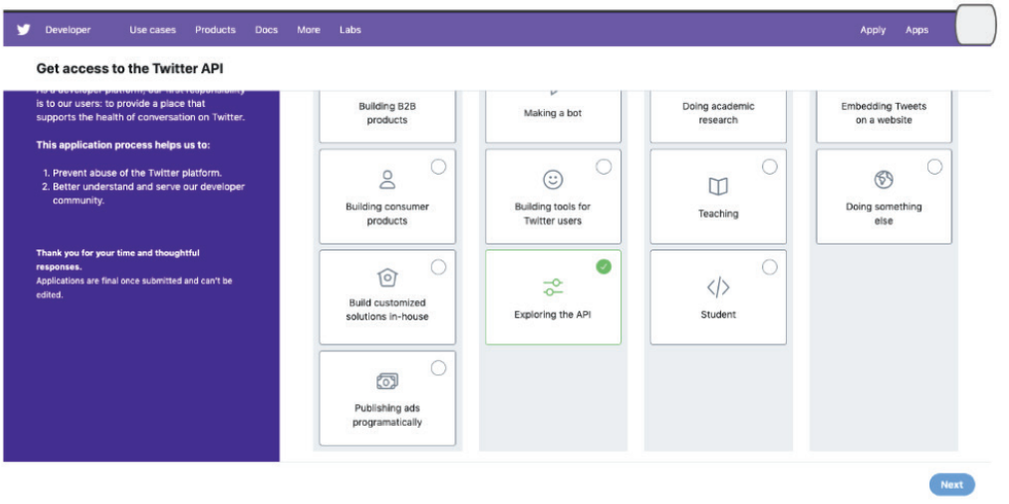
As soon as you access the link above, you'll be redirected to this 'Create an app' process that will be reviewed by Twitter, from whom you should receive a response within 24–48 hours.



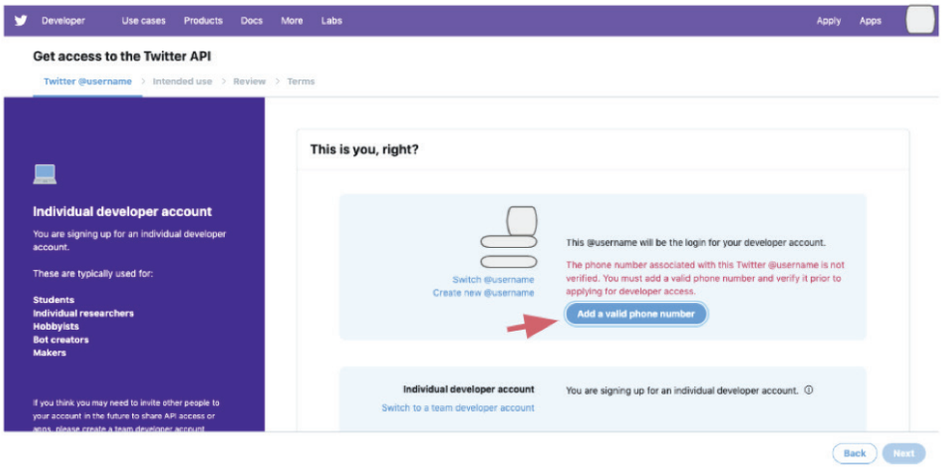
Click the 'Apply' button indicated below.



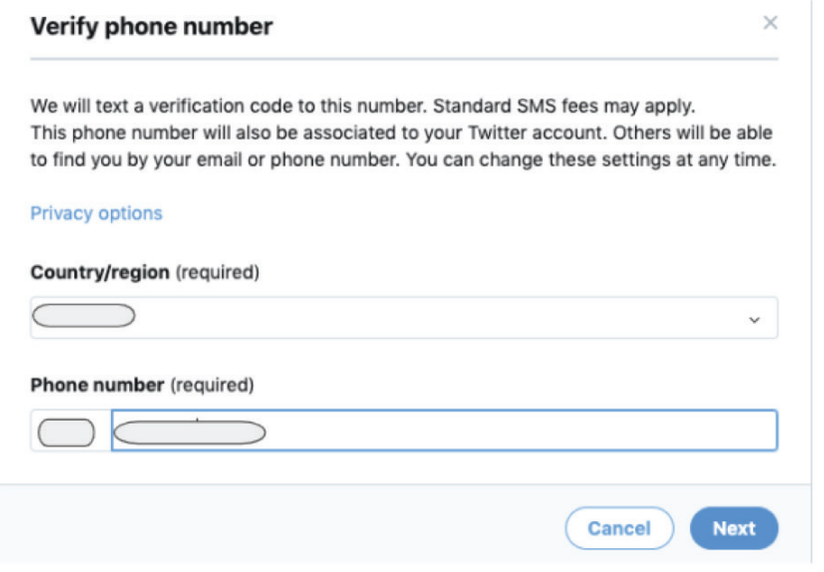
Select 'Exploring the API' as the reason for installing the application, which applies to individual researchers.



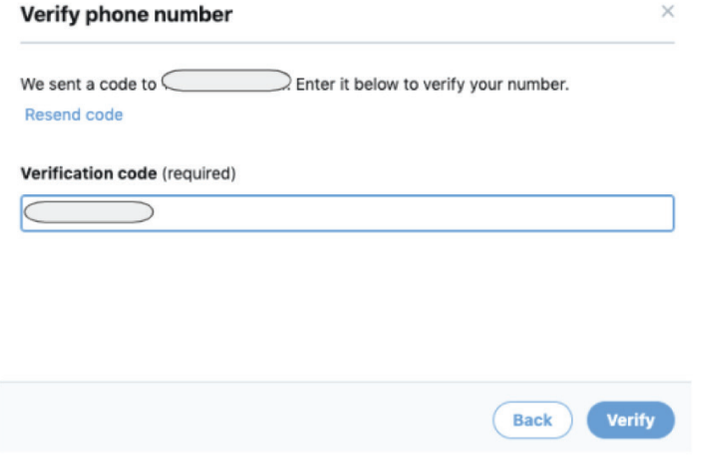
Add a phone number, make sure the corresponding phone is nearby, and you will shortly thereafter receive an SMS containing a verification code.



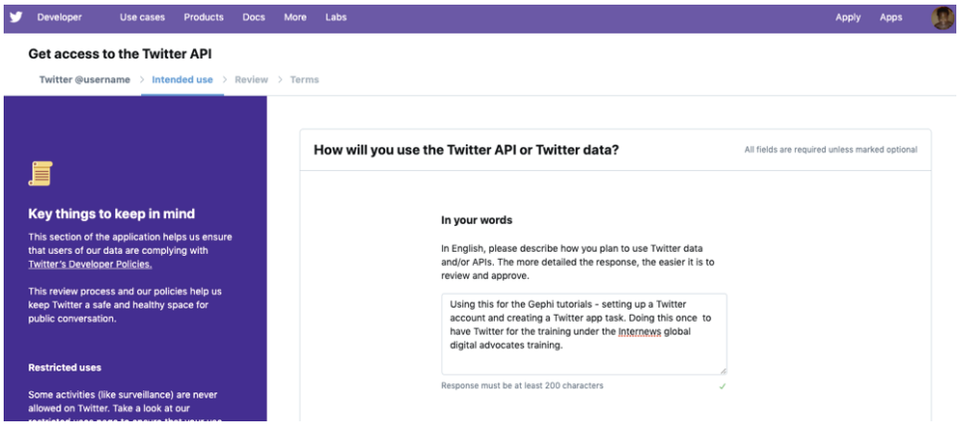
Verifying the phone number requires that you enter the country and phone number details as shown on the right.



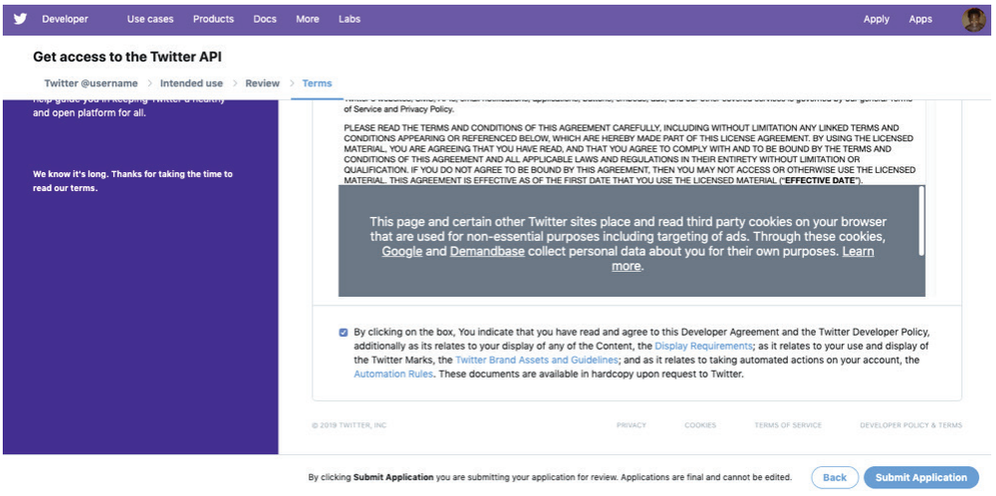
Enter the code sent to your phone and click 'Verify'.



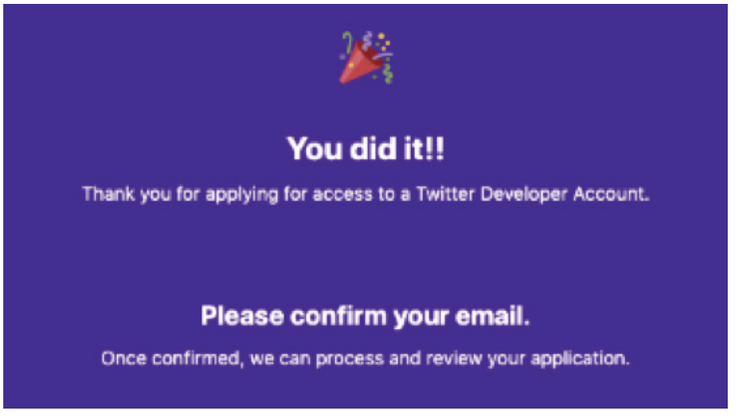
Briefly share or complete the questions asking how you plan to use the Twitter data. Write this explanation in your own way, but keep your reasoning limited to personal research, training, and learning purposes related to the Internews DRA program.



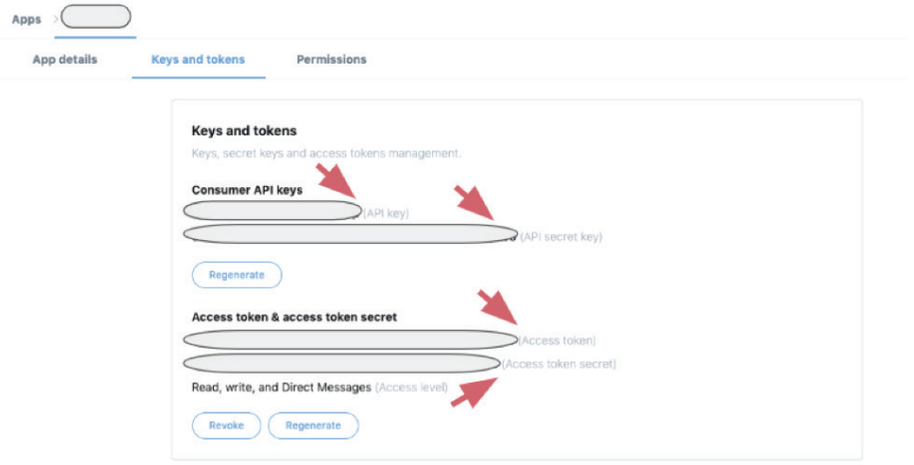
Confirm your agreement to the prompted policy by selecting the box shown below and click 'Submit Application'.



Check your email for confirmation to complete the process.

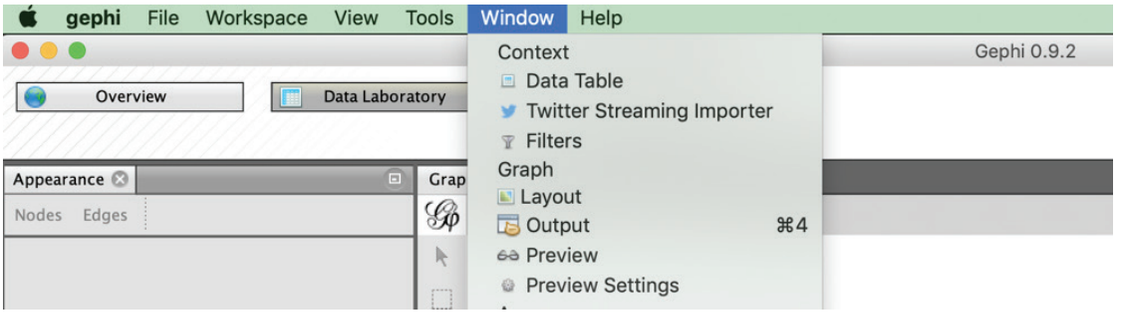


If approved, go through the creation of Twitter app, and you should then be able to view the following information that will be needed for the next steps detailed below.



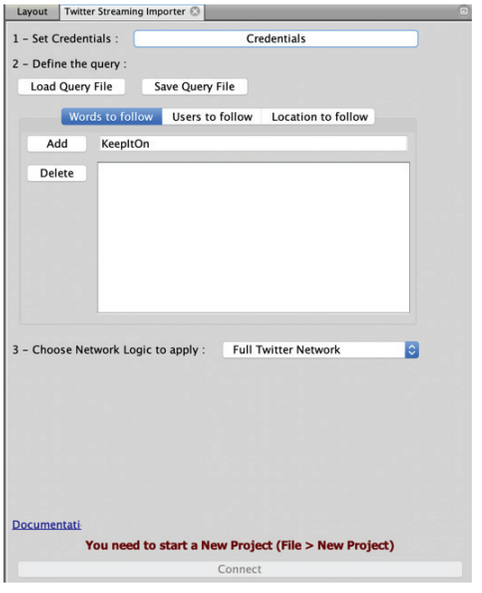
Back to Gephi:

To access the Twitter connection for downloading, go to Window > Twitter Streaming Importer.



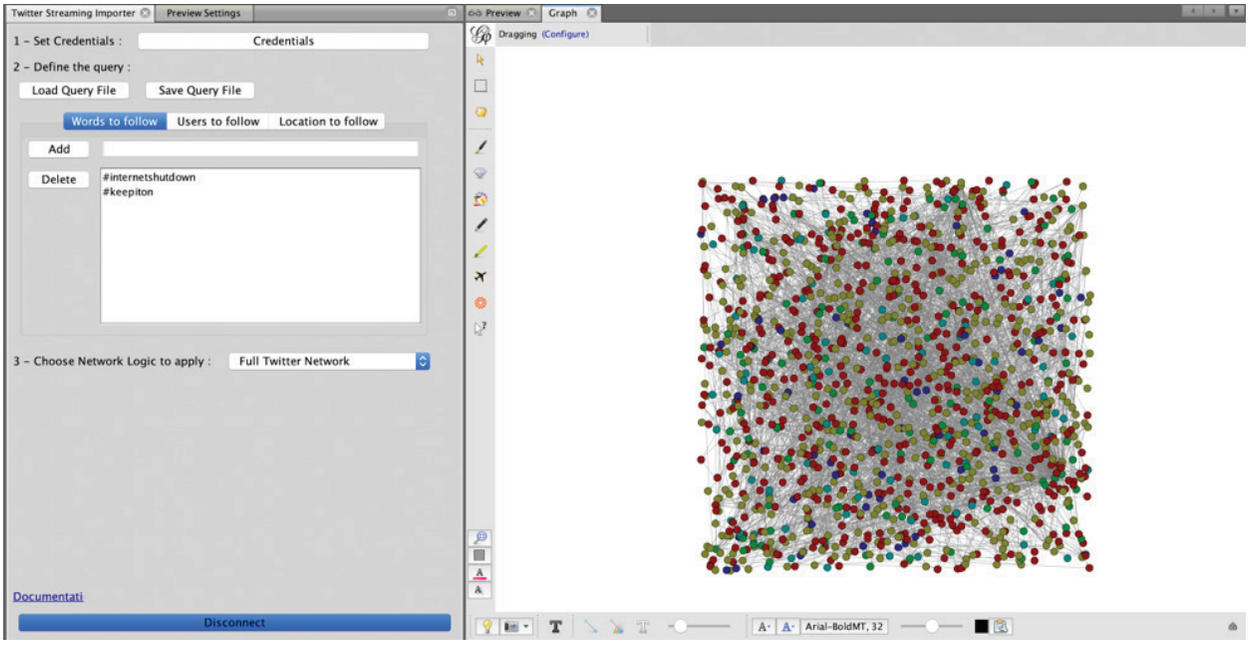
Click on Credentials >

You will be presented with the screen depicted below and asked to enter your Twitter credentials to access the data through the Twitter API.

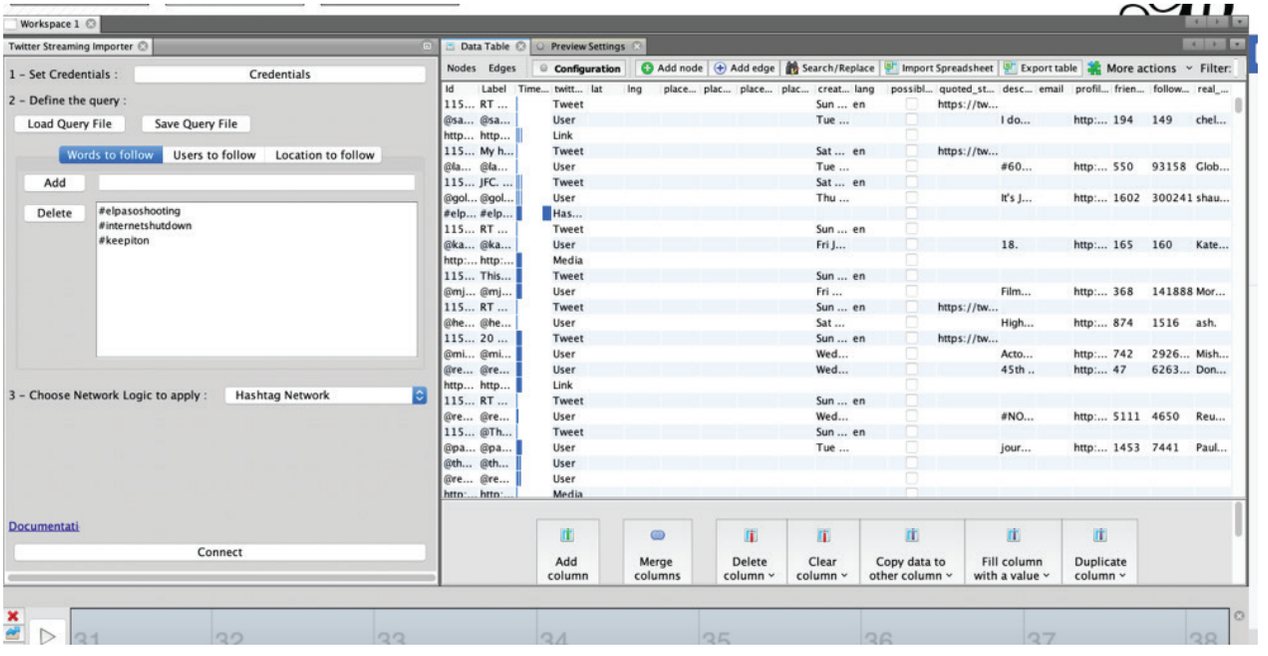


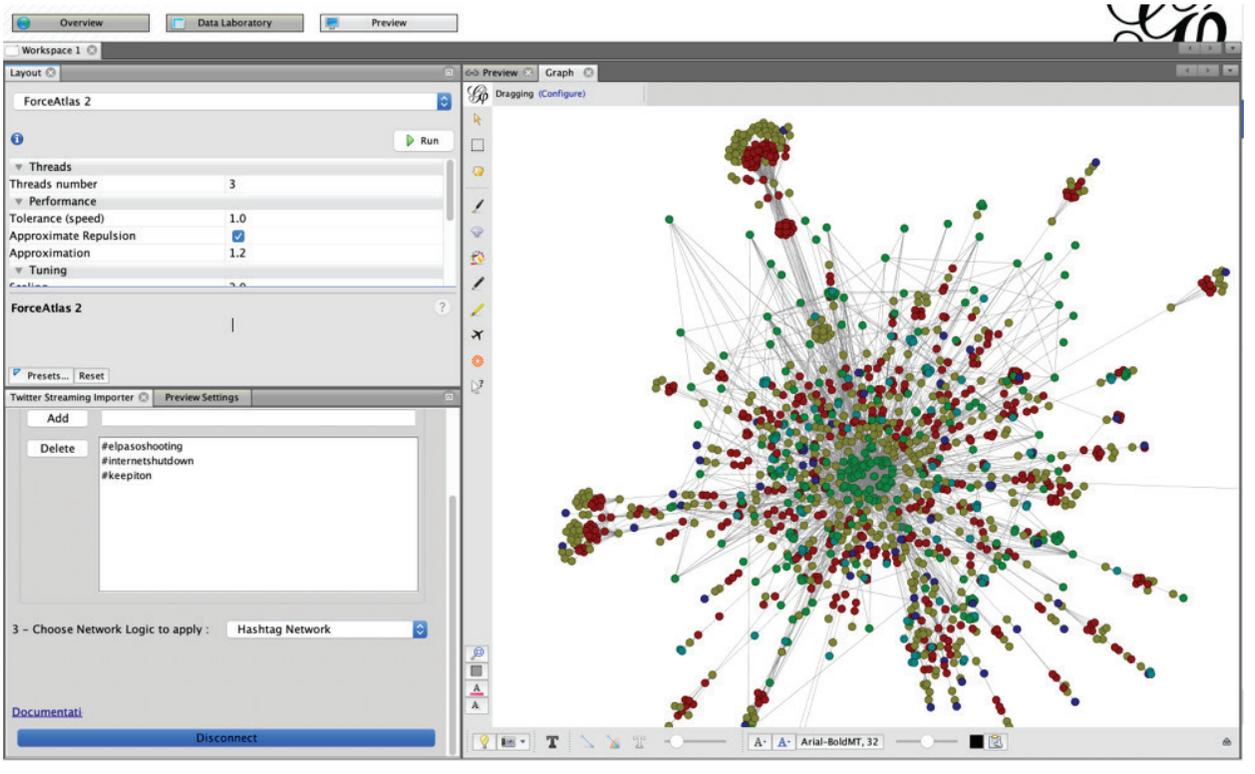
When all the necessary credentials are entered, click Add > Words to follow to see hashtags that are currently trending, then click 'Connect'.

You should start observing the network map formulate according to the incoming data, which are dependent upon how highly trending the data currently are. We worked with two hashtags: #elpasoshootings and #KeepItOn; for test purposes, we encourage that advocates look at their Twitter and pick any currently trending hashtags to test this process.



Clicking on the data table reveals the raw data being downloaded in real time.





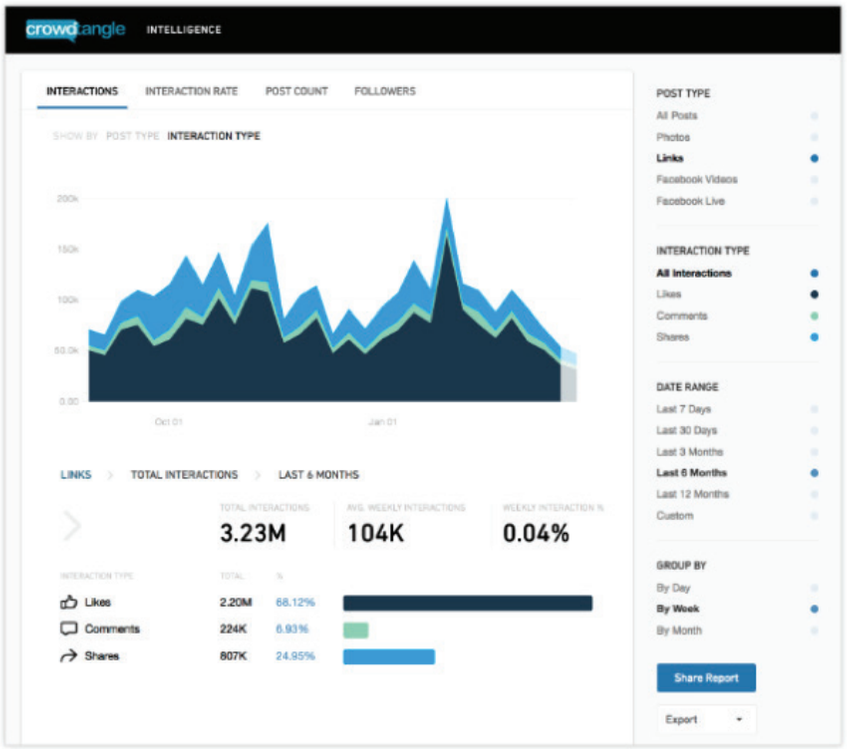
SOCIAL NETWORK DATA ACCESS.

During the process of creating this handbook, we were informed via a tweet by Professor Bernhard Rieder, the creator of the #Netvizz tool, that Facebook had finally taken down their app. This left researchers, advocates, academics and activists without a platform on which they could analyse and download Facebook's social graph data for research purposes.

In the short time since the Cambridge Analytica scandal, many of the social network analysis tools that were previously available have become obsolete or have ceased to exist. Facebook removed Graph Search, Twitter removed geotagged Tweets, IntelTechniques lost its tools and Pipl put up a paywall. Although these setbacks may be frustrating, it is important for advocates to remember that the foundation of good social network analysis is not simply using the right tools but also a sound method. Tools are transient, but good analysis methods stay the same.

CrowdTangle is the future!

In early 2019, Facebook and Instagram announced that they would offer CrowdTangle access to academics and the research community. This will be the only way researchers and academics will be able to access Social Graph data.



CrowdTangle is a content discovery and social analytics tool that was acquired by Facebook in November 2016. Content from more than five million public social accounts can be accessed in real time. There are several key features of CrowdTangle, such as real time dashboards that can be used to track accounts, keywords, links and historical data. Users can create custom reports and track topics across platforms. To date, CrowdTangle has been used primarily to help newsrooms and media publishers understand what is happening on the platform.

Researchers will only have access to CrowdTangle through proposals. Successful proposals will allow researchers to obtain access to both a CrowdTangle API through Social Science One and training that will assist them with data access and analysis.

Researchers obtaining access to the CrowdTangle API will be required to sign a Research Data Agreement with Facebook, which enables approved researchers to obtain legal, trusted, secure and privacy-protected access to Facebook data under the Social Science One framework. Researchers will also be required to agree to the CrowdTangle terms of service. For more details on this, advocates are advised to review this Facebook Newsroom post:

https://www.facebook.com/facebookmedia/blog/crowdtangle-for-academics-and-researchers

STORAGE & PRESERVING DATA.

Hosted on the platform Github, we have a Facebook page dataset that may be downloaded at a later date. Before getting into the case studies' details, we will cover Github's basics features, including repositories, commits, branches, pull requests and issues, as additional content for advocates. The goal is for advocates to be able to use GitHub independently for their own projects. It is an amazing tool for any advocate who wishes to collaborate on SNA, work on general data analysis, back up their work or share their methodology with partners.

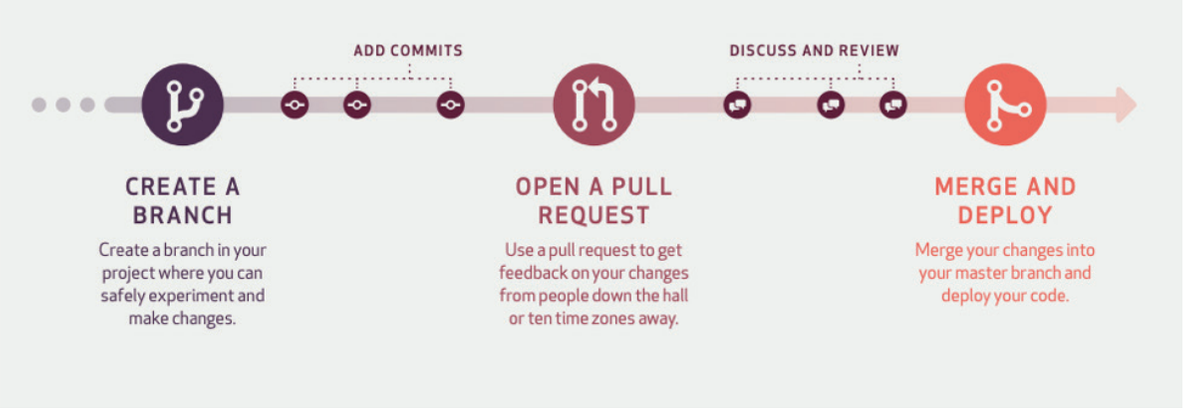
What is GitHub?

GitHub is a platform for hosting and collaborating on projects.

Say an advocate wishes to store a specific dataset and create a handmade visualisation or report template. They will need to create the whole thing themselves, and they will want to store all the data and documentation online in one place. That's GitHub. Not only does the platform store files, it can also remember all changes made through a program called version control. Therefore, it is always possible to return to an earlier stage and correct errors.

GitHub is also a cool platform that advocates, academics and even activists are encouraged to use because it allows their projects to be open source. In other words, their projects will be free for others to share, use and improve. In return, they will be able to use the gems others put on GitHub. They will be able to search for them, copy ('fork') and reuse (if they can). Or, they may 'star' them for later.

It is also important to note that GitHub is far from being a Content Management System (like WordPress); however, advocates can host a blog on GitHub or publish their projects as standalone GitHub Pages.



SOME VOCABULARY.

Before diving into GitHub, get a leg up by understanding its litany of technical terms. Later, advocates will also want to download Git and the GitHub native app for either Mac, Linux or Windows. GitHub explains every step on their website.

1. COMMITS:

A commit is simply a change you make to a piece of code or a set of related changes made on multiple files.

2. FORKING:

Forking is the copying of an existing code repository at a specific point so you can add to it as you see fit. The original code will still exist for the creator, but your fork may spin off into a lively new project. This means that any project publicly hosted on GitHub can be edited by anyone. For example, if you find a typo in the Internews Digital Rights handbook, you may change it.

3. PULL REQUESTS:

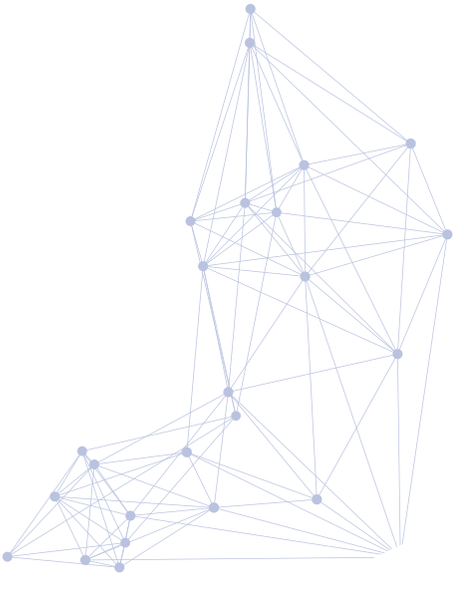
Pull requests let users know what changes you may have made to a particular repository. Once you send a pull request, other users can review your changes and incorporate them into their own repositories.

4. REPOSITORY:

A collection of commits and the branches and tags used to identify commits.

5. REMOTE:

The ‘remote’ locations of your repository, normally on a central server.



WHY BOTHER?

‘Why should advocates bother going the hard route if they aren’t technical (yet)?’

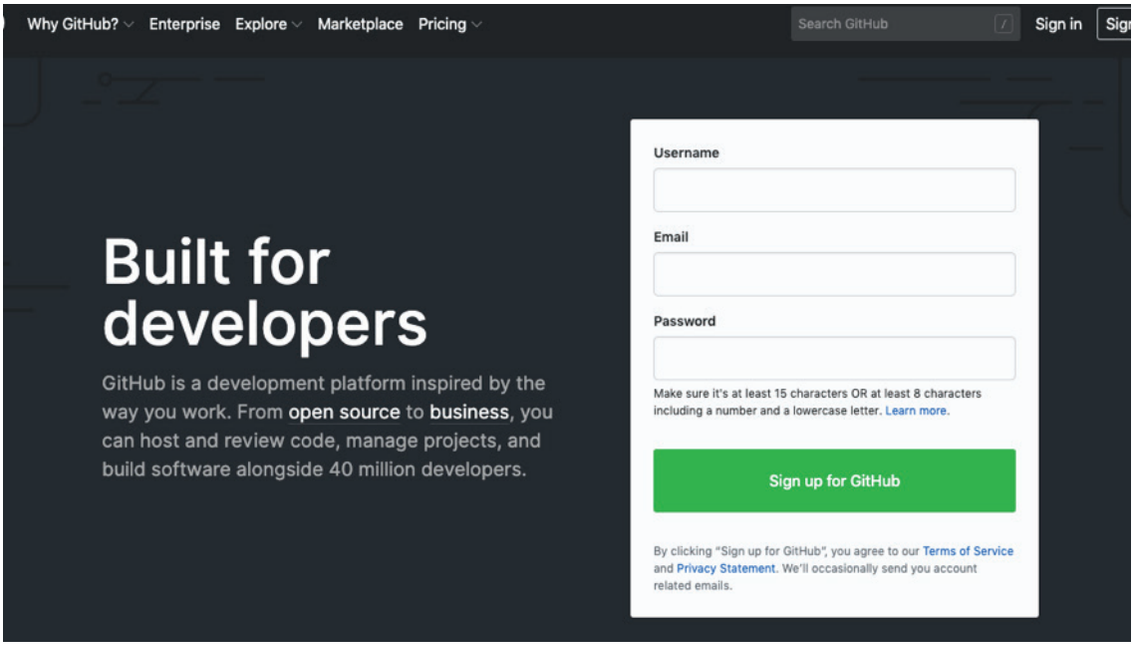
For a smooth collaboration with partners and to track management changes across projects, we encourage advocates to invest some time learning how to use Github. We know that it’s much more fun (and better motivation) to learn things when you need them for real projects rather than to start from the very (boring and theoretical) beginning. Therefore, let’s get started with Github’s practical application.

Let’s begin by setting it up

1. Go to GitHub.com and sign up.

This is as easy as registering for any site on the web. Choose a free plan, obviously. (Later on, don’t forget to add a nice pic of yourself and to edit your ‘About Me’ section).

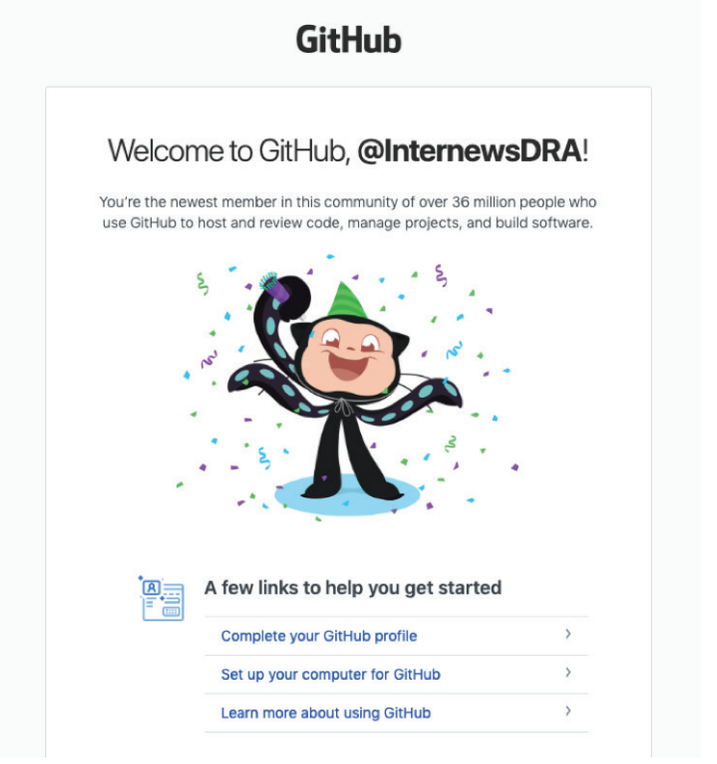
You’ll arrive at this start page:



2. Complete the registration.

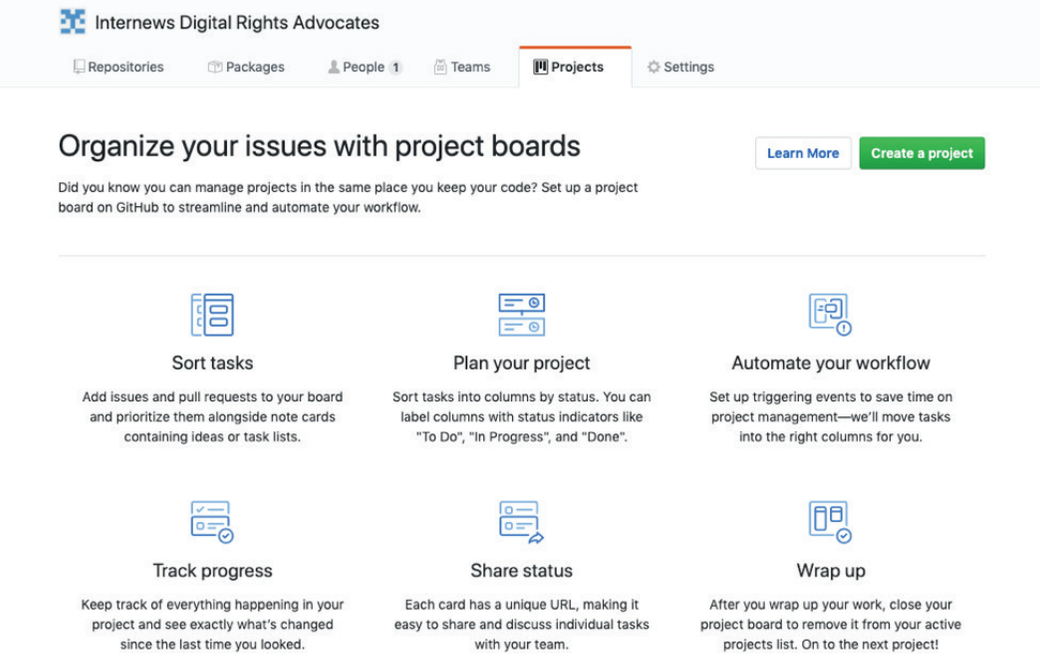
You should receive a confirmation email and the messages shown on the right.

Advocates should choose a username relevant to themselves or to a project they are currently working on.

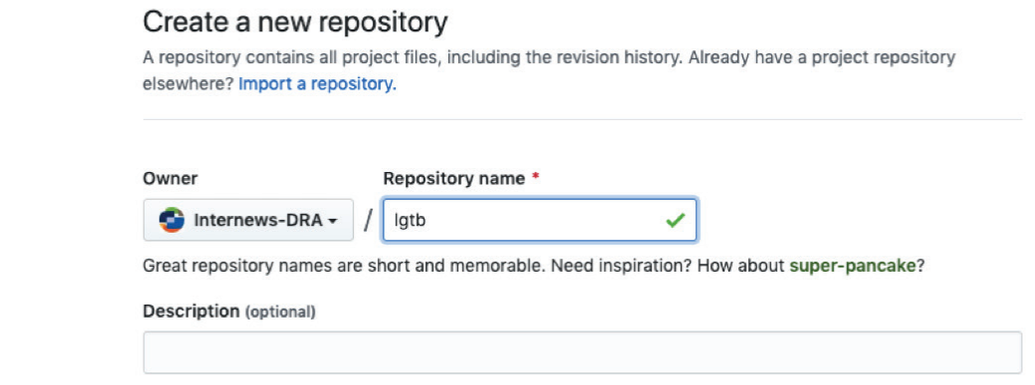


3. Create a project.

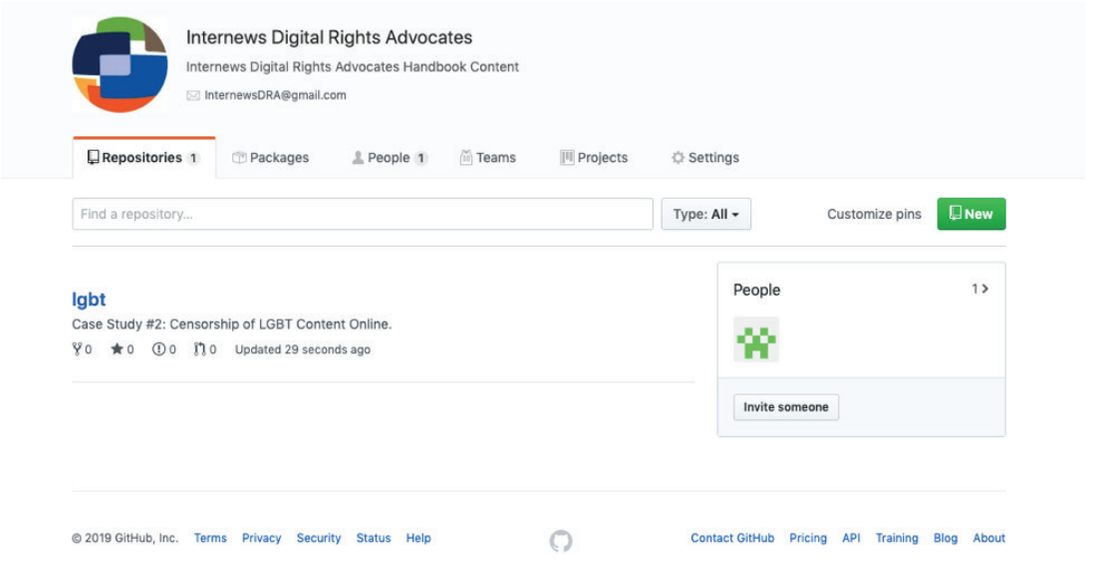
Within this project, let's go ahead and create a repository in which data can be uploaded.



4. Create the repository as an area in which project data files can be uploaded.



You should have the repository, as can be seen below, sitting somewhere within your personal Github account.

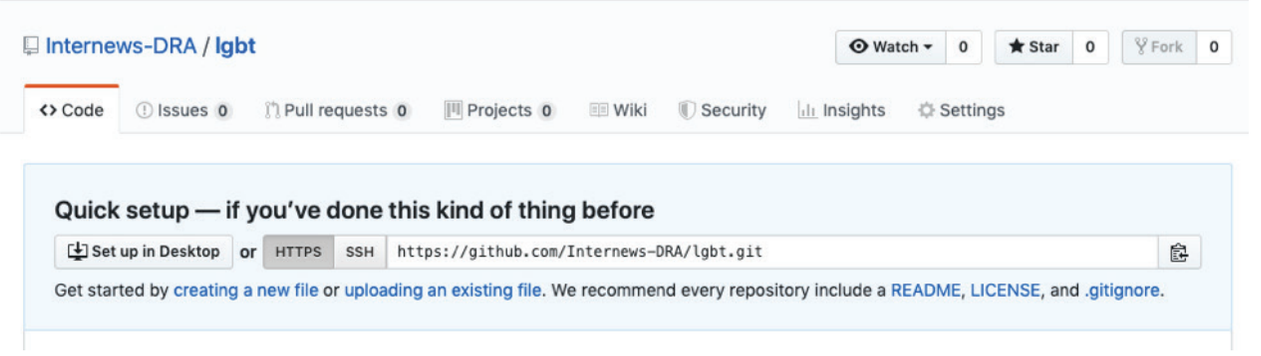


Adding a file to a repository

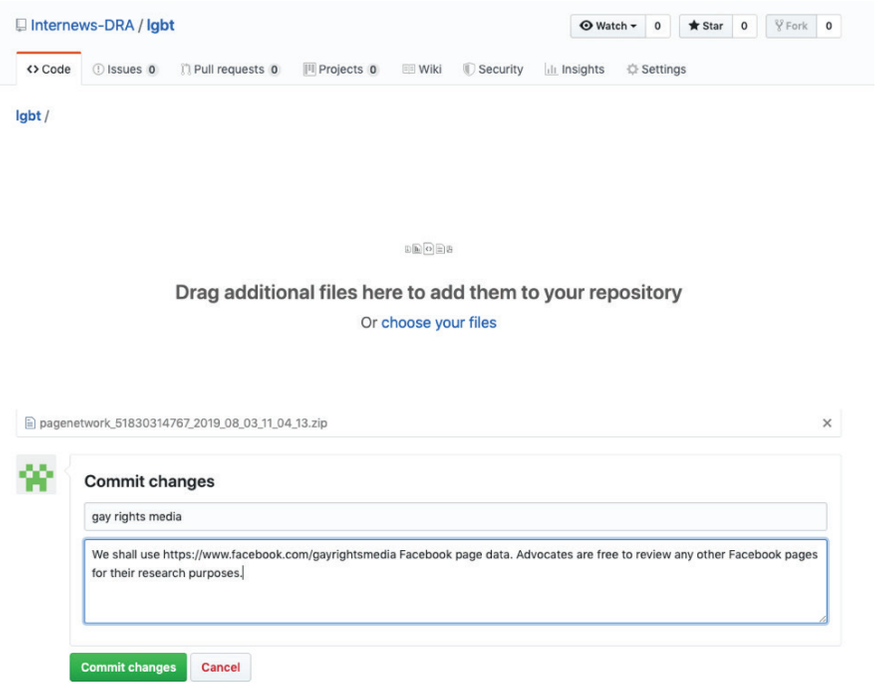
Advocates can upload and commit an existing file to a GitHub repository. Drag and drop any file to any directory in the file tree, or upload files from the repository's main page.

Files that you add to a repository via a browser are limited to 25 MB per file. You can add larger files, up to 100 MB each, via the command line; however, that is not within the scope of this handbook.

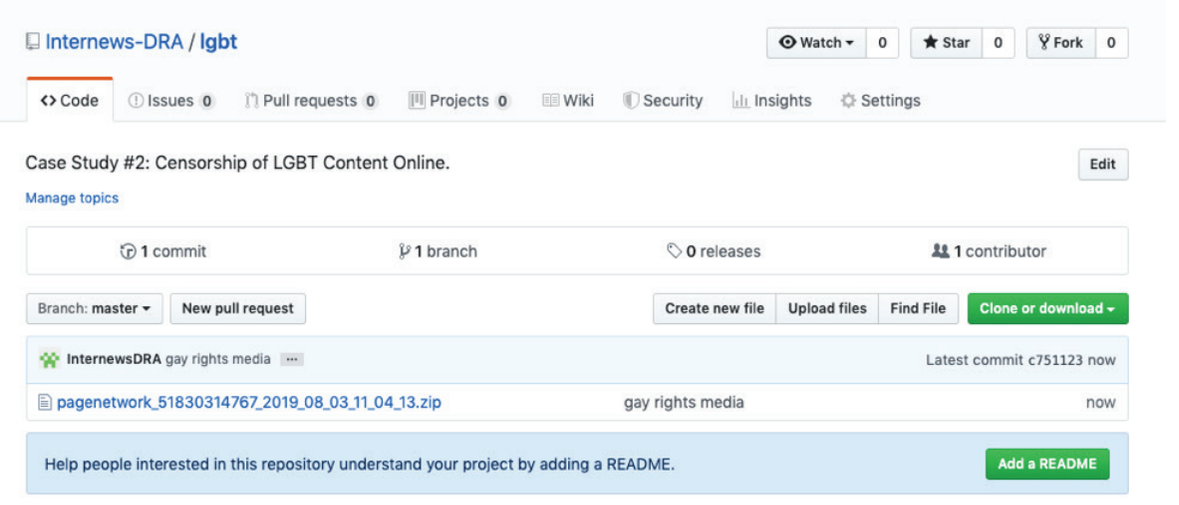
Drag and drop the file or folder you would like to upload to your repository onto the file tree.



At the bottom of the page, type a short, meaningful commit message that describes the changes made to the file. You can attribute the commit and commit message to more than one author.



After clicking 'commit changes', advocates should now have a file within their Github Repositories, as shown below, that is ready to be shared or used in collaboration with other partners.



CASE STUDY #2

Censorship of LGBT Content Online.

We will leverage the data already downloaded from Facebook's social 'Graph API' to perform various analyses. User likes, pages and posts are data points related to the many people who actively use Facebook on a daily basis and both produce and consume content.

Here in particular, we shall find connections between people, pages and organisations based on the sample dataset obtained from Facebook in August 2019 before they shut down the #Netvizz tool. We shall download this data from the Internews Github account https://github.com/Internews-DRA. The data is related to the 'Gay Rights Media' Facebook page on which several organisations discuss topics related to the censorship of LGBT online content. We shall use the Gephi tool to, for example, find links between people or pages involved in the discussions, advocacy or development of projects or who are members of various groups with strong connections or activities within these discussions.

As previously mentioned, after the Cambridge Analytica scandal in 2018, Facebook received a great deal of criticism for not providing an easy way to export personal data to be used for research. However, the #Netvizz Facebook app created by Bernhard Rieder (professor at the University of Amsterdam) provided a tool that could do just that and even more. Unfortunately, as this handbook was being created in August 2019, this app was also shut down by Facebook, which left almost no other app that allowed access to this data.

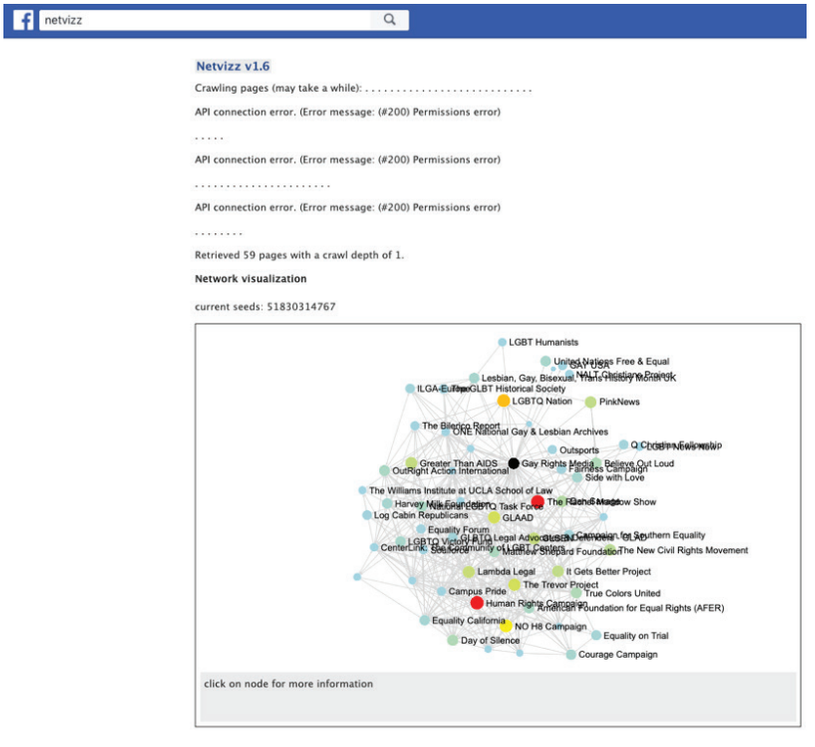
For this case study, we shall use the downloaded data that is now hosted on the Internews DRA Github account. #Netvizz was a data collection and extraction application that allowed researchers to export data in standard file formats from different sections of the Facebook social networking service. Friendship networks, groups and pages could be analysed quantitatively and qualitatively with regard to their demographics, post demographics and relational characteristics. This handbook provides an overview of the directions this data made available. We will not discuss data extraction via the official Application Programming Interface (API), as the tool was shut down and no longer exists.

We shall go through the details on how advocates can obtain the sample network data now hosted on Github, import it into Gephi and analyse it there. Once they have downloaded the network and imported it into Gephi, they will be able to perform the following activities to answer some of the previously posed questions.

- **Compute the density and centrality measures of the network**
- **Identify the brokers and central nodes in the network**
- **Visualize the network to identify its brokers**
- **Perform a modularity analysis and identify the communities**

The 'Gay Rights Media' Facebook page used in this analysis can be accessed as shown below.

A preview of how the #Netvizz tool visualized the network [Note: no longer accessible].



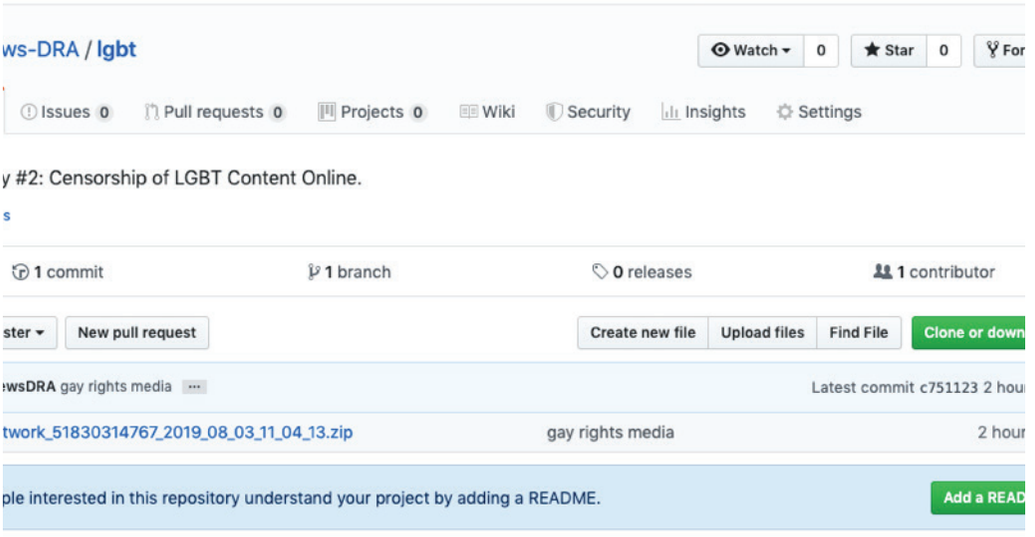
Download the raw GDF data that will be used later or use the import to Gephi tool.

Note: We're hosting the raw datasets on Github. This data is a zip file, also known as a GDF file (which is a variation of the CSV format), in which all Facebook page friendships, relationships and even some meta-data are encoded. Advocates can open this file using a network visualization package (for example, Gephi) and obtain a neat visual representation of the Facebook network and the communities that comprise it.

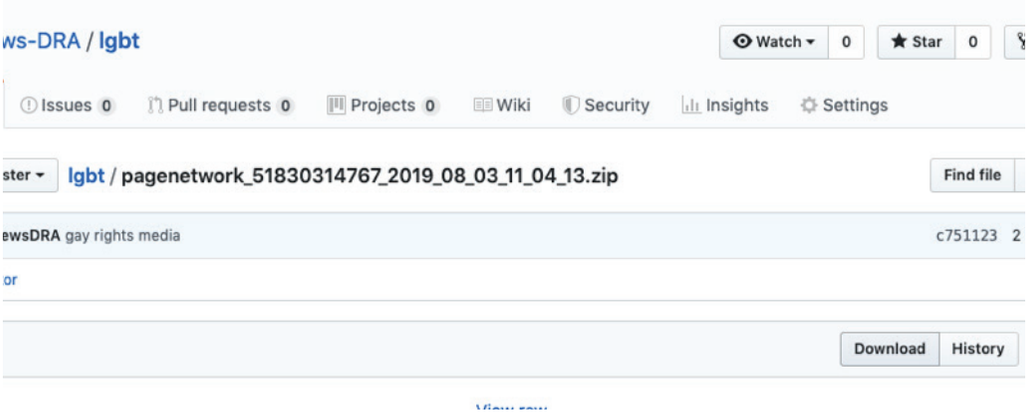
Go to https://github.com/Internews-DRA/

Click on the repository named 'lgbt'.

Click on the .zip file that contains the data.



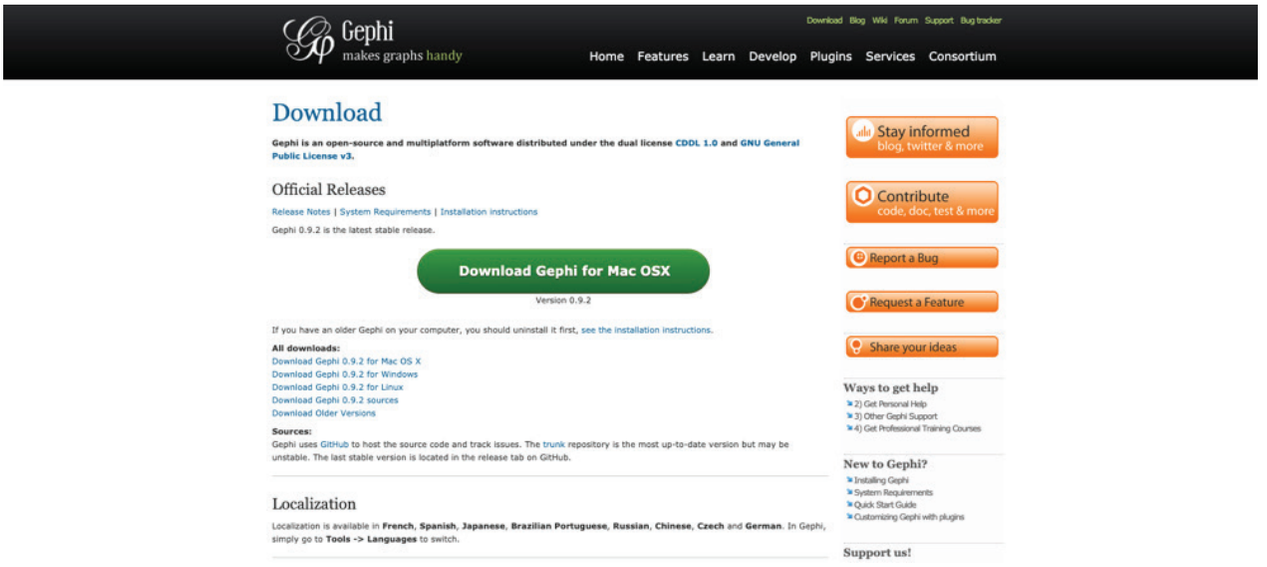
Click on the 'download' button to download the file to your computer.



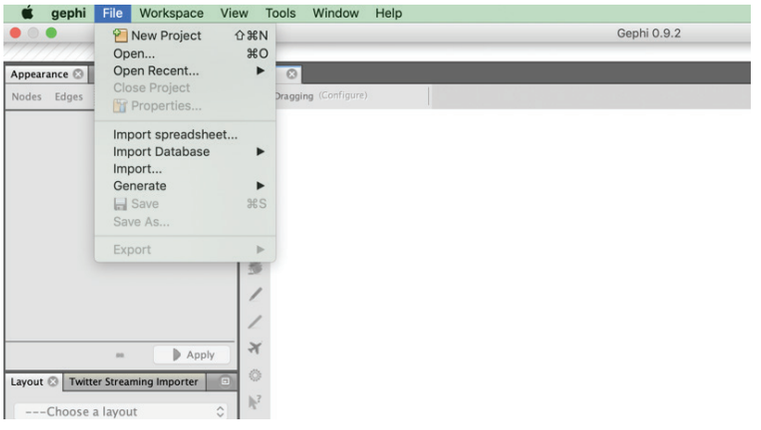
Downloading and installing Gephi:

Download guide:
https://gephi.org/users/download

Installation guide:
https://gephi.org/users/install

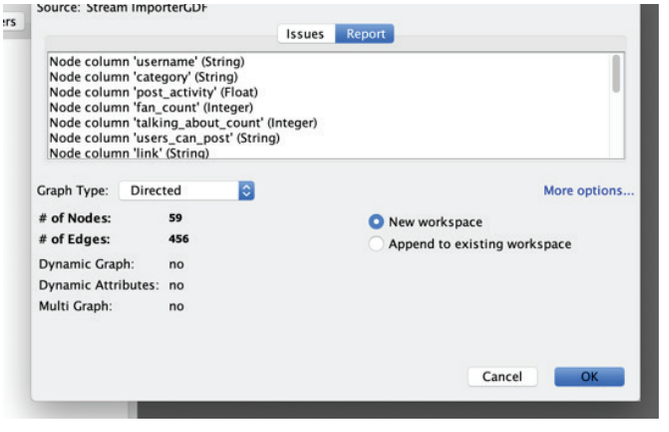


On the Gephi menu bar, go to 'File' and 'Open' the .gdf file.

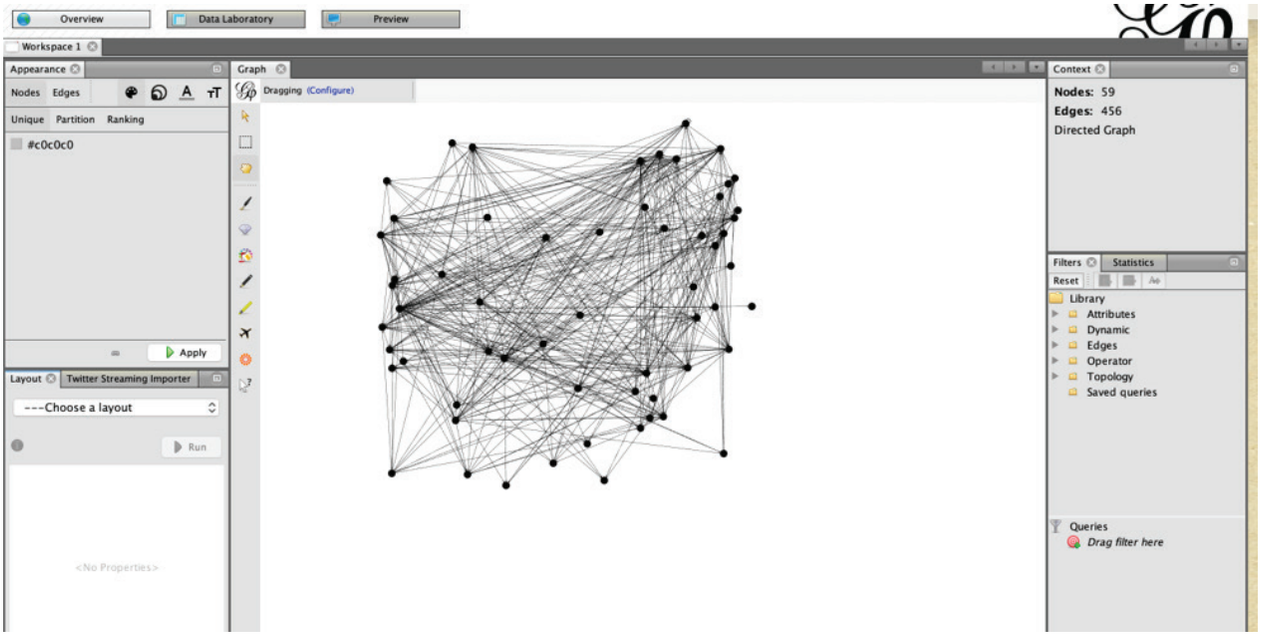


Opening the file.

When your file has been opened, the import report will summarize the data found and any issues. Click 'OK'.



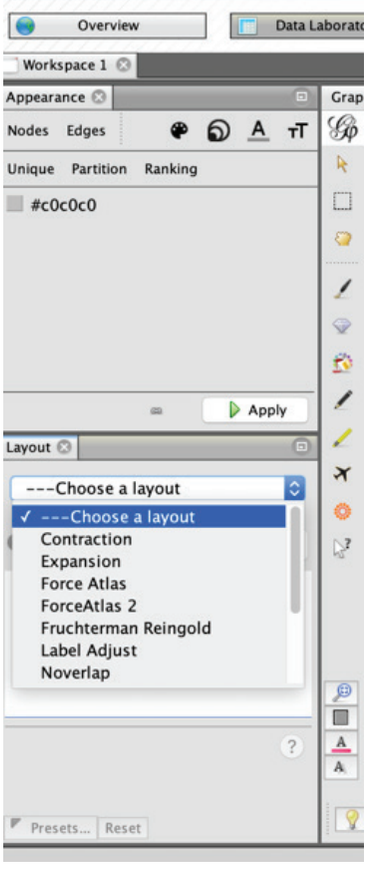
When your file has been opened, you can see a hairball like this:



Locate the layout module on the left panel. Choose 'Force Atlas'. You can see the layout properties on the right. Click 'Run' and 'Stop'.

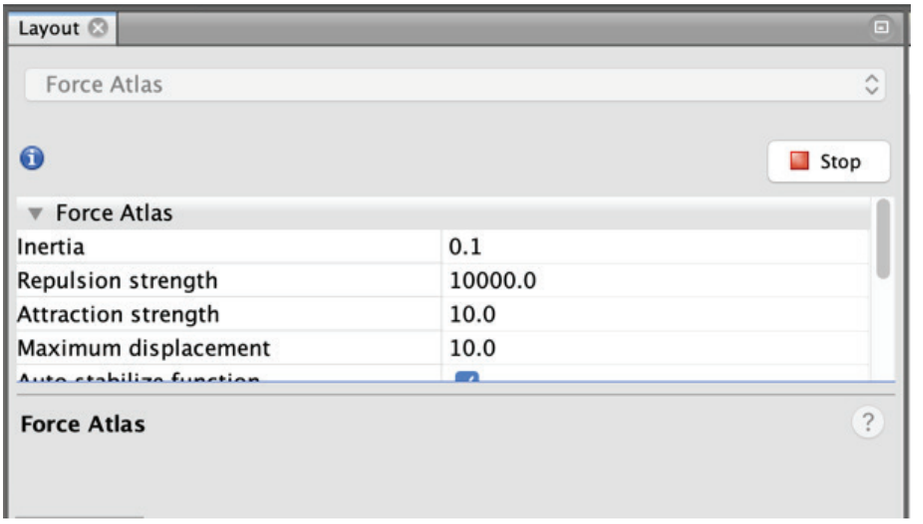
Control the layout

The purpose of the layout properties are to allow you to control the algorithm and create a readable representation. 'Force Atlas' causes the connected nodes to be attracted to each other and pushes the unconnected nodes apart to create clusters of connections.



Set the repulsion strength at 10,000 to expand the graph.

- **Click 'Enter' to validate the value.**
- **Click 'Stop' when it seems as though you have some different clusters.**



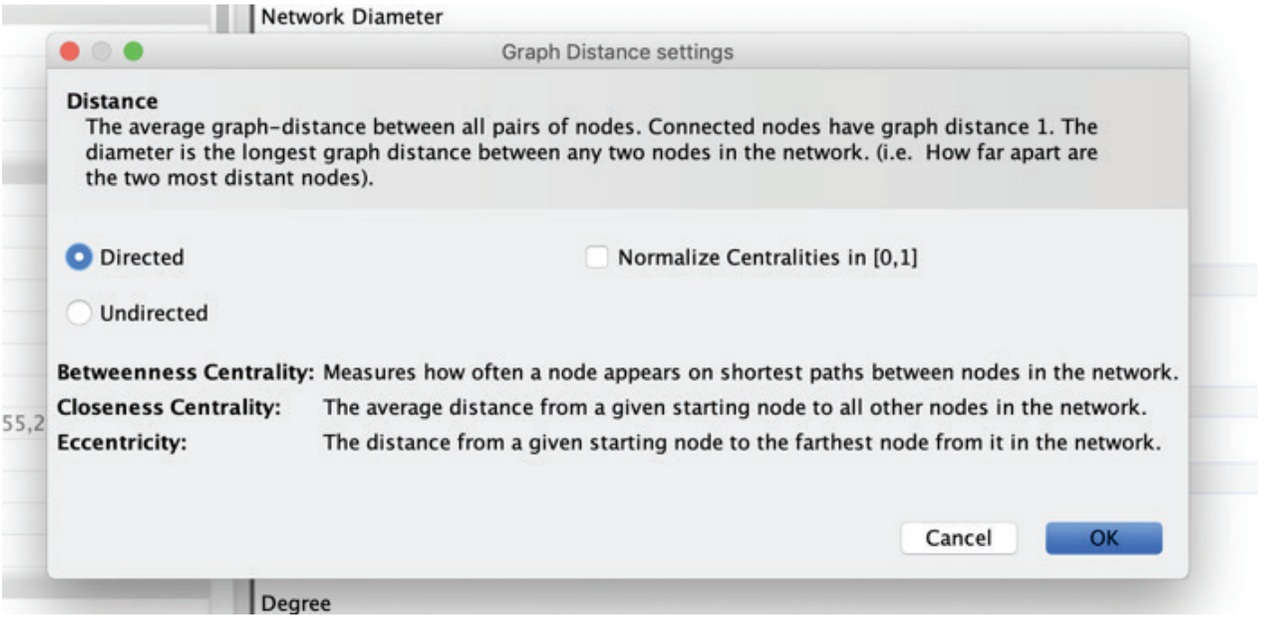
Ranking (degree)

The ranking module allows you configure the nodes' colour and size.

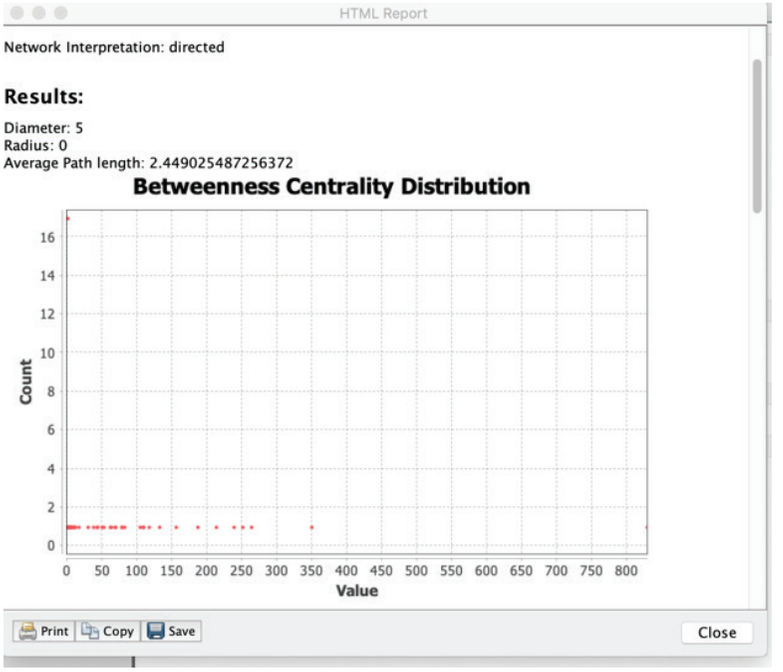
- **Choose the ranking tab on the top left module and choose 'Degree' from the menu.**
- **Click 'Apply'.**

Statistics

- **Click the statistics tab on the top right module.**
- **Click 'Run' next to average path length.**
- **Select 'Directed' and click 'OK'.**



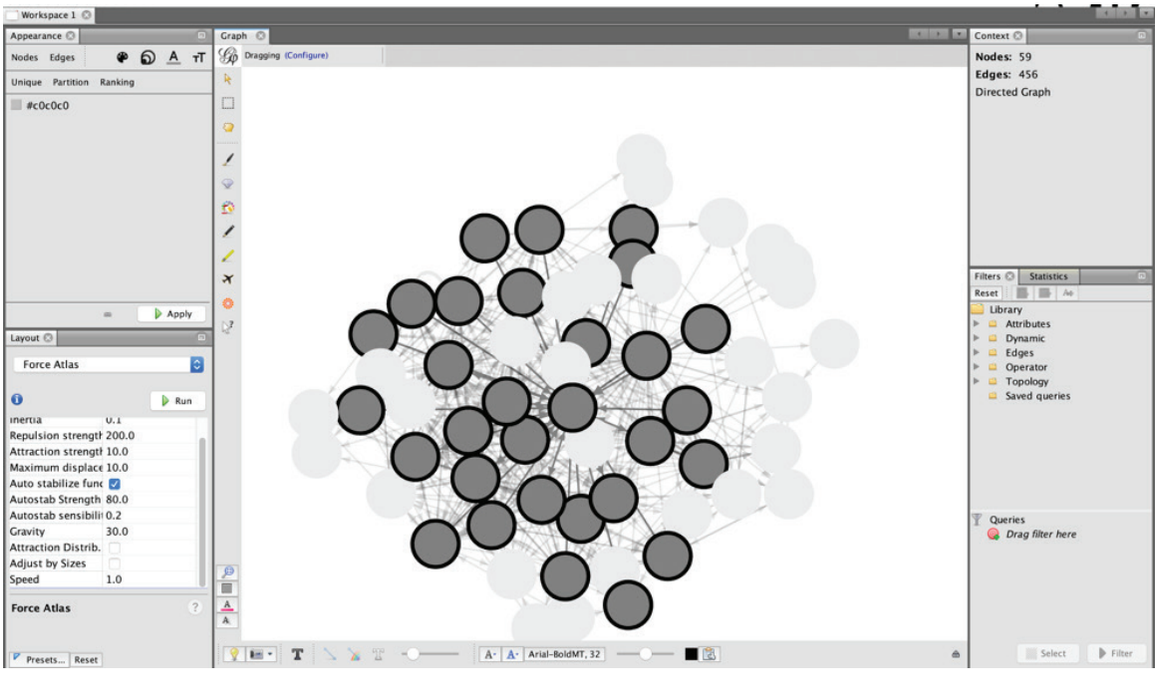
When finished, the metric displays its results in a report like this (betweenness, closeness and eccentricity):



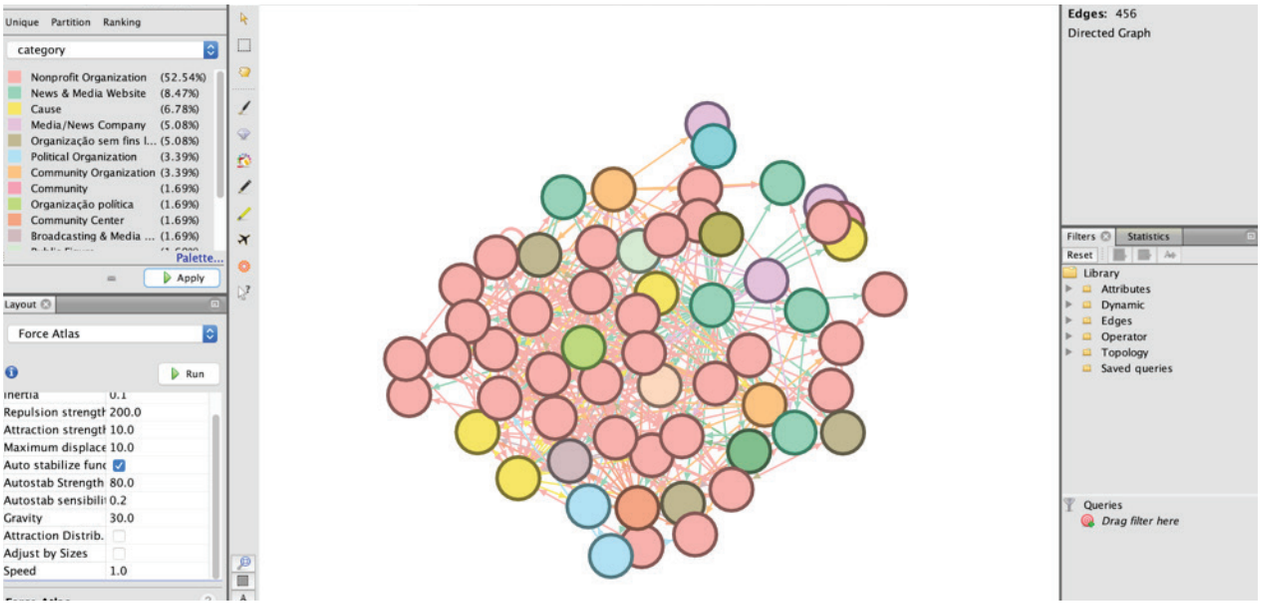
Ranking (Betweenness)

- **Return to ranking on the top left module and choose a rank parameter from the dropdown menu 'Betweenness Centrality'.**

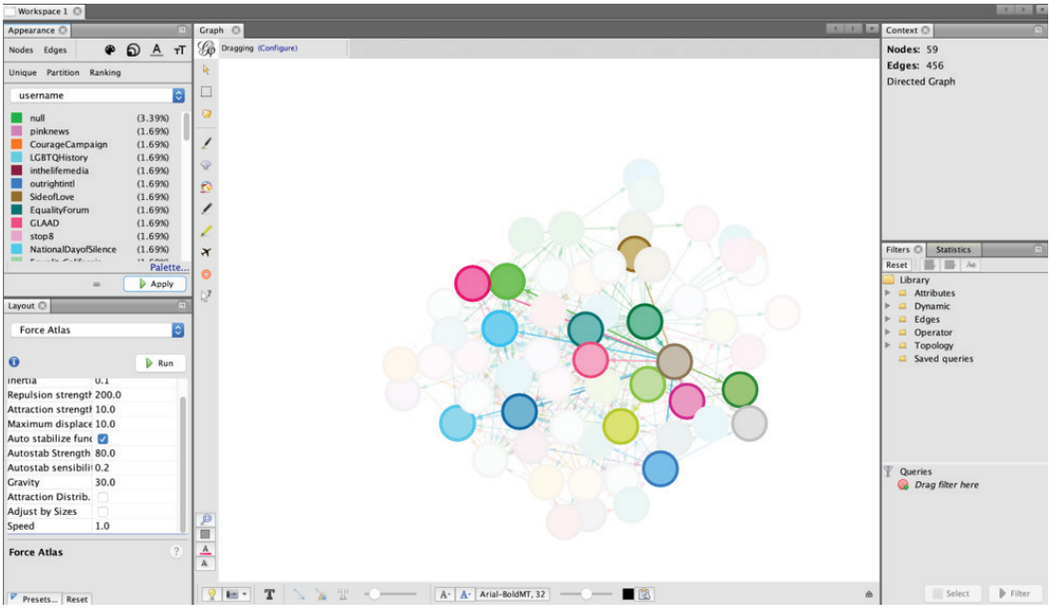
We can closely see the network below formulating. At this point, advocates should be able to begin analysing and explaining the network as related to the definitions explained in the beginning of this handbook.



Advocates may now notice the different actors of interest categories, such as Non-Profits, Cause, Political, Government and News & Media.
Which category has the highest number of actors?

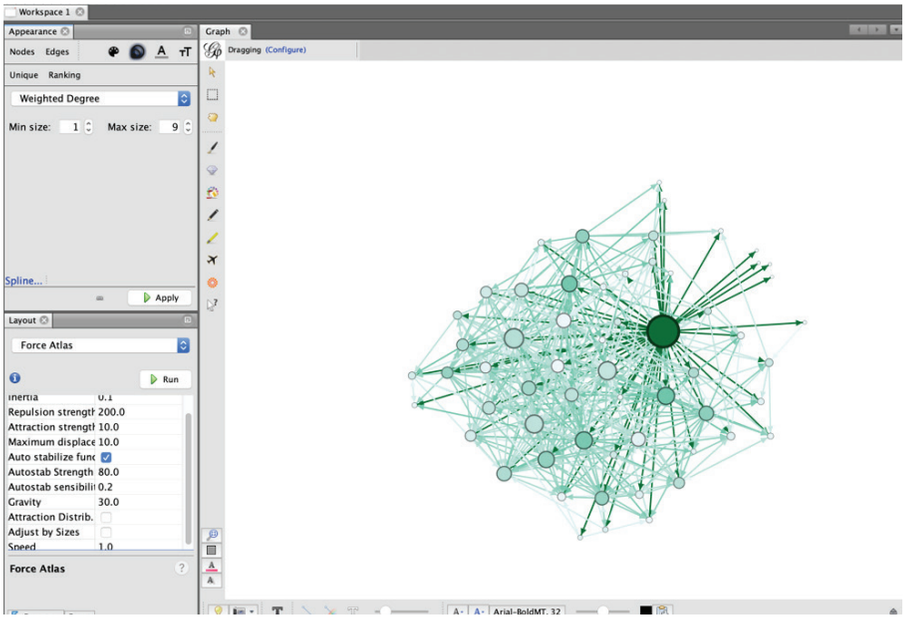


Advocates should be able to see individual page networks and explain this in detail here.
Which actors or pages seem more influential than others?
Roll over the actors to see the arrows pointing across.



Weighted Degree.

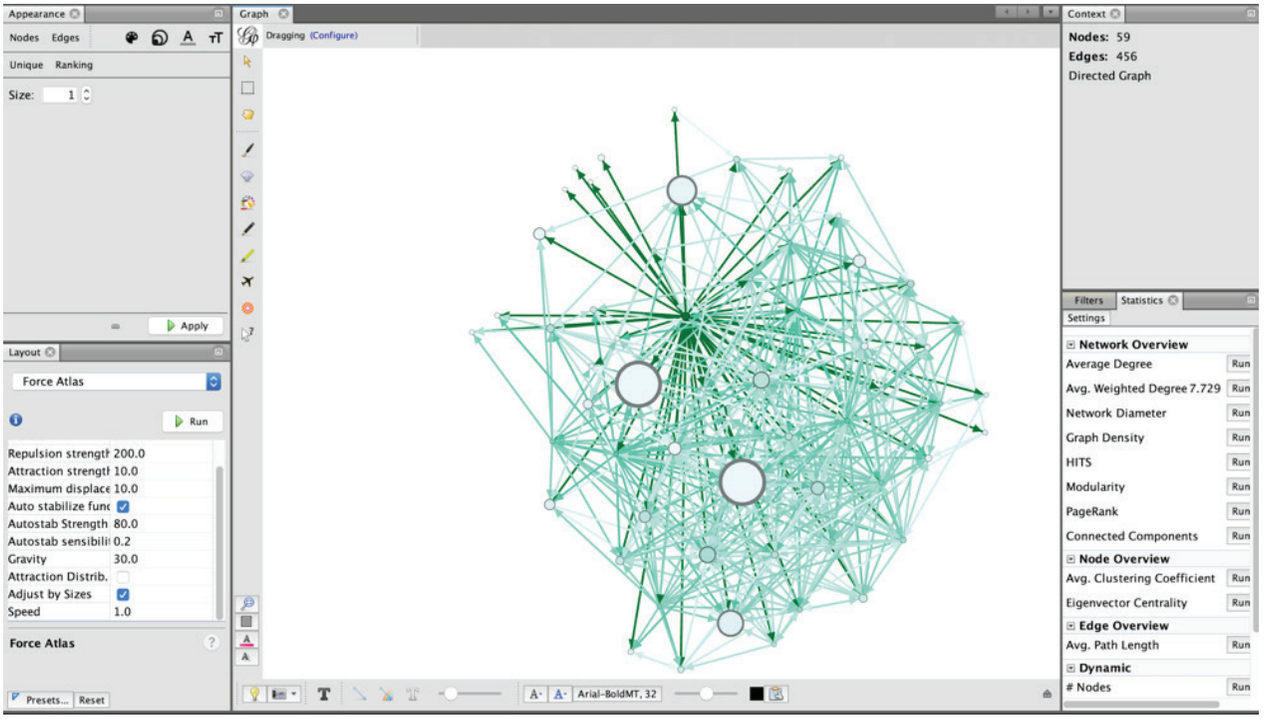
Look into the weighted degree and see which pages or actors have the highest number of relationships. This can be done by going into the statistics tab and running the weighted degree. To view the different sizes, go to the ranking tab > select symbol > select weighted degree / run. Finally, you may wish to go to the spatial tab > select force atlas > check parameter 'fit by size' > run.



Centrality.

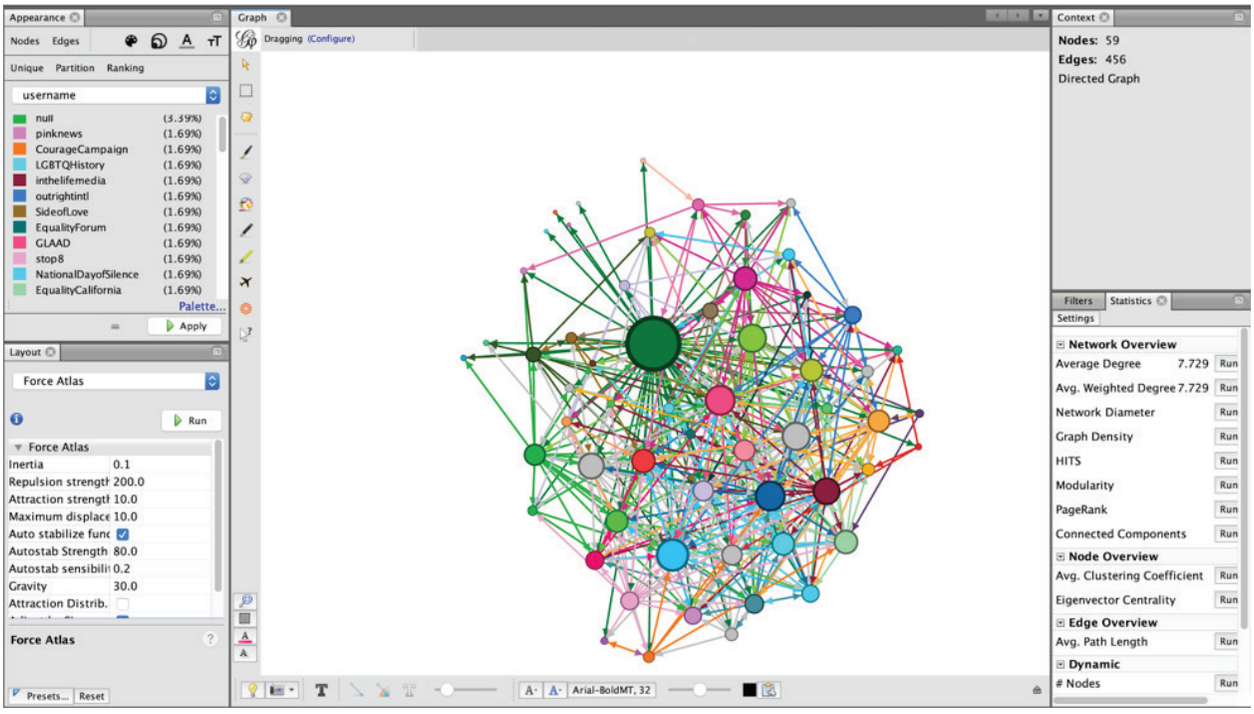
Check how central or peripheral a specific node is to the network. You might wish to go further and look into the degree-centrality. Examine which nodes have many connections or the number of ties each specific node has.

Do you see in-degree or out-degree in the directed networks?



Network Distances.

Look at the shortest paths (geodesics). The diameter of the network is the length of its longest geodesic. What is the length of the network below? What is the reachability of connectedness? In other words, can every node reach every other node and how would this be important for digital advocates examining the censorship of LGBT online content?



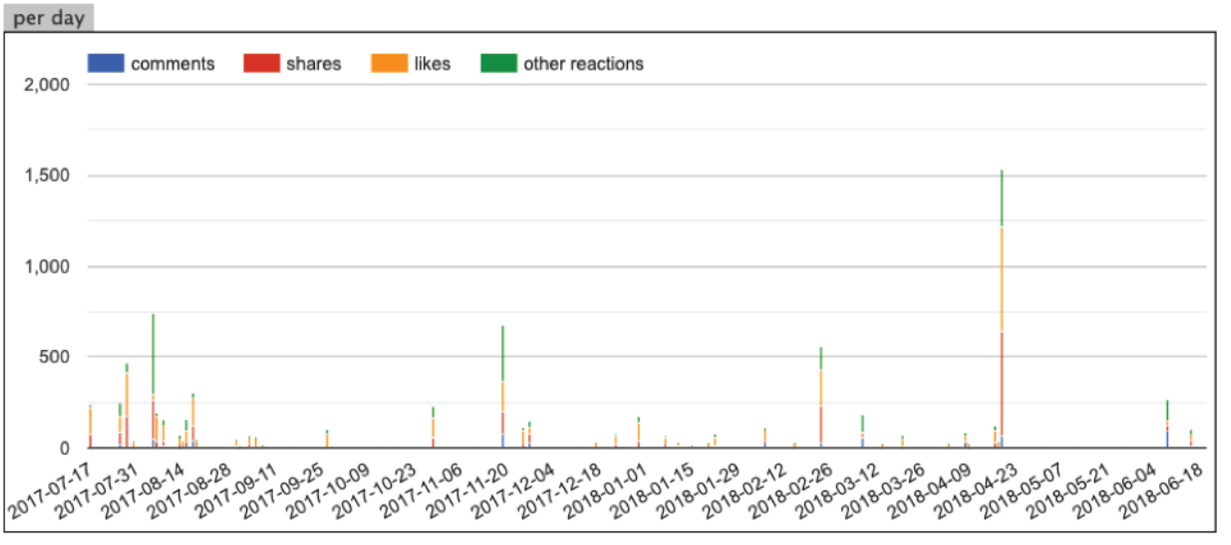
In the above case study, we demonstrate how advocates can analyse or perform SNA on any Facebook page or profile using a combination of freely available tools. We can identify the structure and key influencers of the 'Gay Media Rights' Facebook page. In addition, we are able to answer questions, such as 1) how does the structure of this page affect the ability to successfully interact with the censorship of LGBT content online and 2) what kinds of people or actors exert the most influence within this page.

Visualizing the page as a network and identifying the most prominent communities and most influential members can answer both questions. It can also highlight any possible strategies for strengthening or weakening these groups and find the main points of entry: the people and clusters within the groups that should be addressed to convey a message to the groups in the quickest and most efficient manner possible.

Below are screenshots from additional analyses that were possible using #Netvizz before the program was taken down by Facebook. We can see the specific page trends through its comments, shares, likes and other reactions over the past few months. We can see the dates on which posts had the most activity and correlate those with any censorship discussions or collaborations from around the same time.

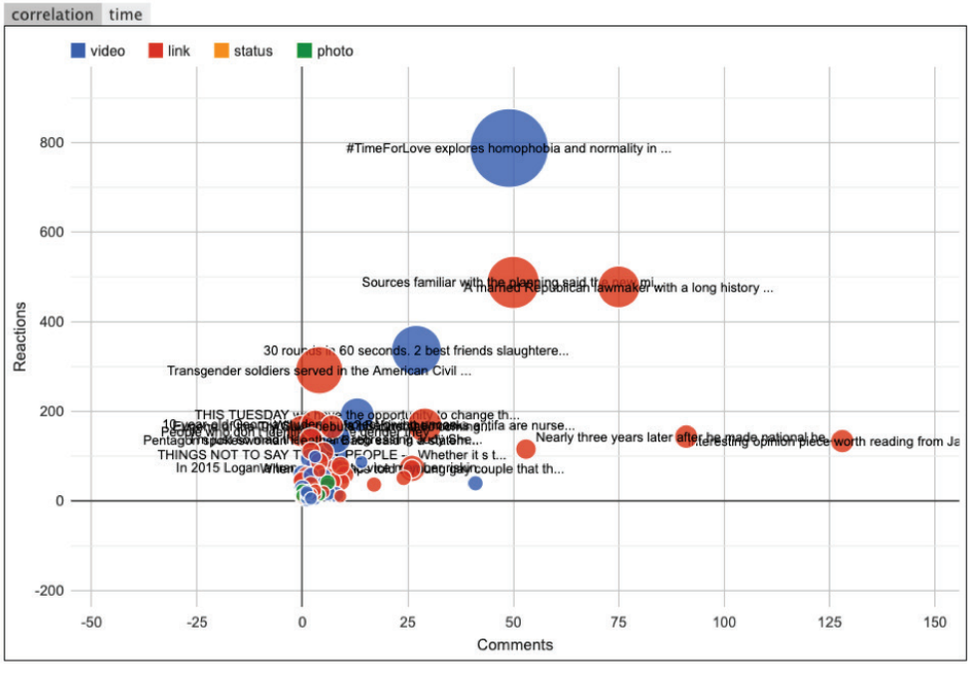
Aggregate Analysis

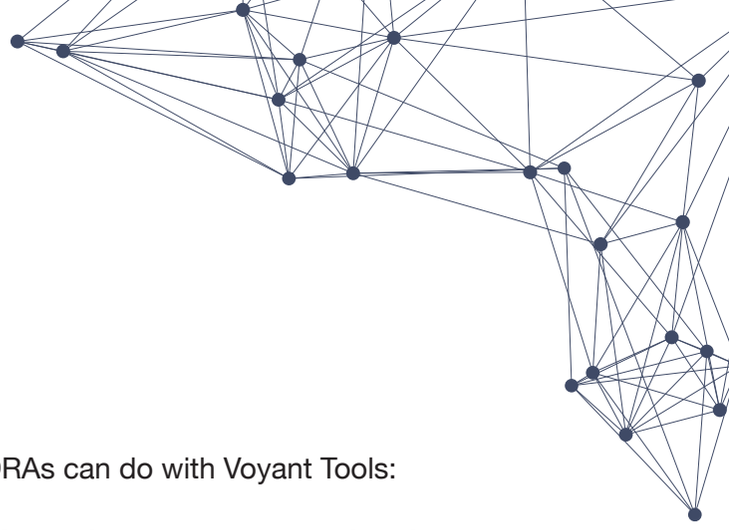
100 posts covering the period from 2017-07-17 20:19:23 to 2019-07-26 13:40:31
895 comments (8.95 average)
6692 reactions (66.92 average)



In addition, with the #Netvizz tool, we could use the Post Explorer tool, which examined high-activity posts across multiple surfaces, such as videos, links, statuses and photos. We could review what type of content had the highest reactions and over what surfaces to help advocates modify their content and discussions based on the available data.

Post Explorer





CASE STUDY #3

Digital Rights and Freedom Bill, 2016

In this final case study, we shall refer to the Nigeria Digital Rights and Freedom Bill, which can be accessed at https://bit.ly/33GtrNG. With an internet penetration rate of 46%, Nigeria has the largest number of internet users in Africa as well as the seventh largest in the world. The need for a law that governs, protects, administers, and enforces the digital human rights was a priority because as technology continues to share and disrupt the current global landscape, certain measures and standards are required to ensure the sanctity of each and every citizen's rights—particularly, their online rights related to digital freedom.

To visualise the data collected through bill reports with significant amounts of text in formats such as PDFs or web HTML, we will need to look at different visualisations and techniques to make sense of this unstructured data. We must work with the data to guide advocates towards determining which keywords or topics seem most important to the bill regarding its frequency or size.

You may find the chosen tool at:
https://voyant-tools.org

Voyant Tools is a web-based text-reading and analysis environment. It is a scholarly project that is designed to facilitate reading and interpretive practices among researchers, digital humanities scholars, as well as the general public.

What DRAs can do with Voyant Tools:

- **Learn how computer-assisted text analysis works; check out our 'next steps' examples that describe how text analysis may be conducted on the Digital Rights and Freedom Bill in the Nigeria case study;**
- **Study texts they find on the web or texts they have carefully edited and saved on their computers;**
- **Add functionality to their online collections, journals, blogs, or websites such that others can see through their texts with analytical tools;**
- **Add interactive evidence to the reports they will publish online and add interactive panels directly into their research essays (if they can be published online) such that their readers can recapitulate their results; and**
- **Develop their own tools using the functionality and code.**

For this handbook and case study related to Nigeria's digital rights bill, we present to you indicator tools that will offer a complete list of five autonomous information blocks related to the data.

- **Cloud of tags: a word cloud visualisation in the form their text; DRAs will be able to choose other forms of displaying the corpus of words and, if necessary, excluding others.**

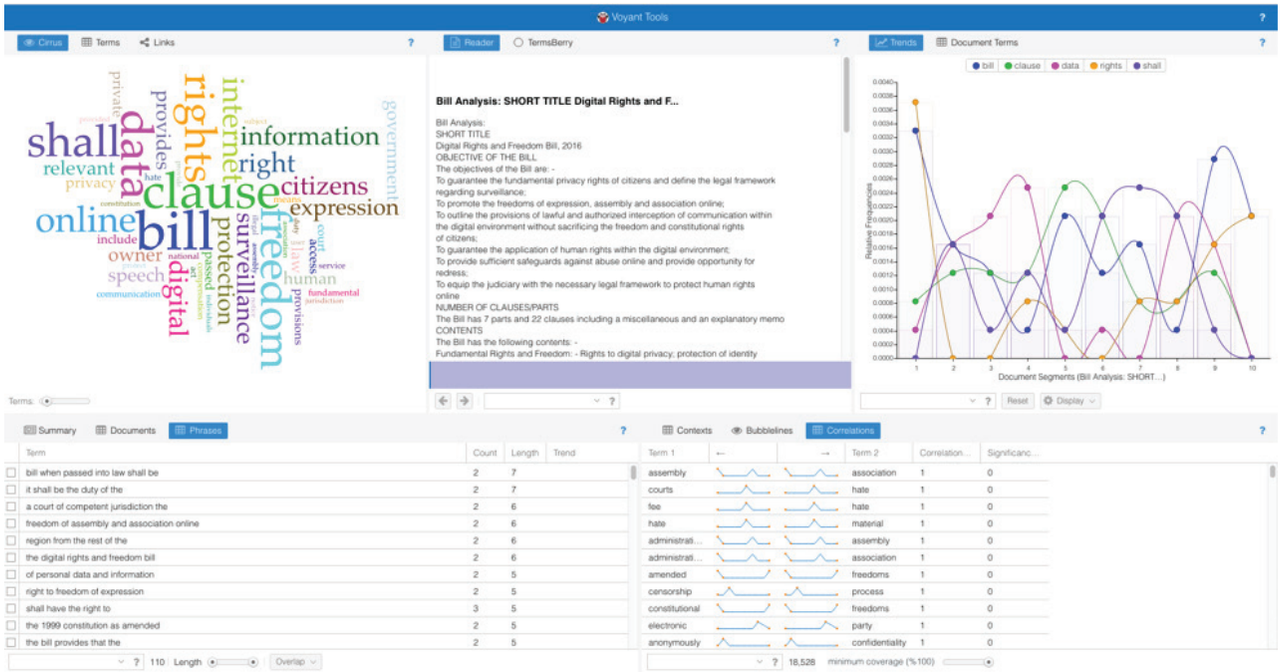
- **Reader: provides access to plain text. DRAs can search for a word via a search box. By clicking on one word in the text, a frequency histogram is generated beneath the text, and the selected word becomes the subject of analysis for all the blocks.**

- **Trends: a powerful tool that identifies and compares the most frequently used keywords as well as studies their proximity within the text.**

- **Summary: This block result of the text's automatic analysis provides statistical information regarding the bill's text under study, including number of words, number of documents, and most frequently used words, among other variables.**

- **Contexts: When selecting a word in the reader, phrases and documents in which the word is mentioned within the bill will emerge.**

These five blocks are, again, merely part of the information extracted from the automatic text analysis by Voyant Tools. By walking through the menus, advocates will be able to modify or even replace the data blocks with others.

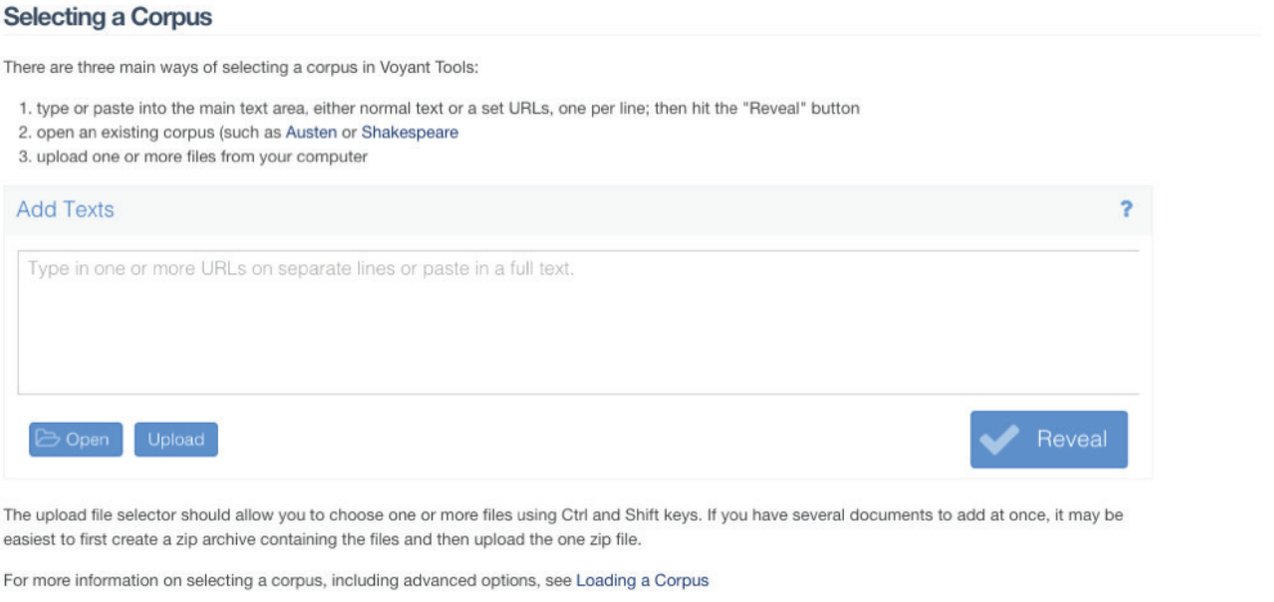


Get started with a corpus or text from the Digital Rights and Freedom Bill:

https://bit.ly/33GtrNG

The text may be downloaded as a document and uploaded or copied and pasted directly into the text area, as shown in the image below.

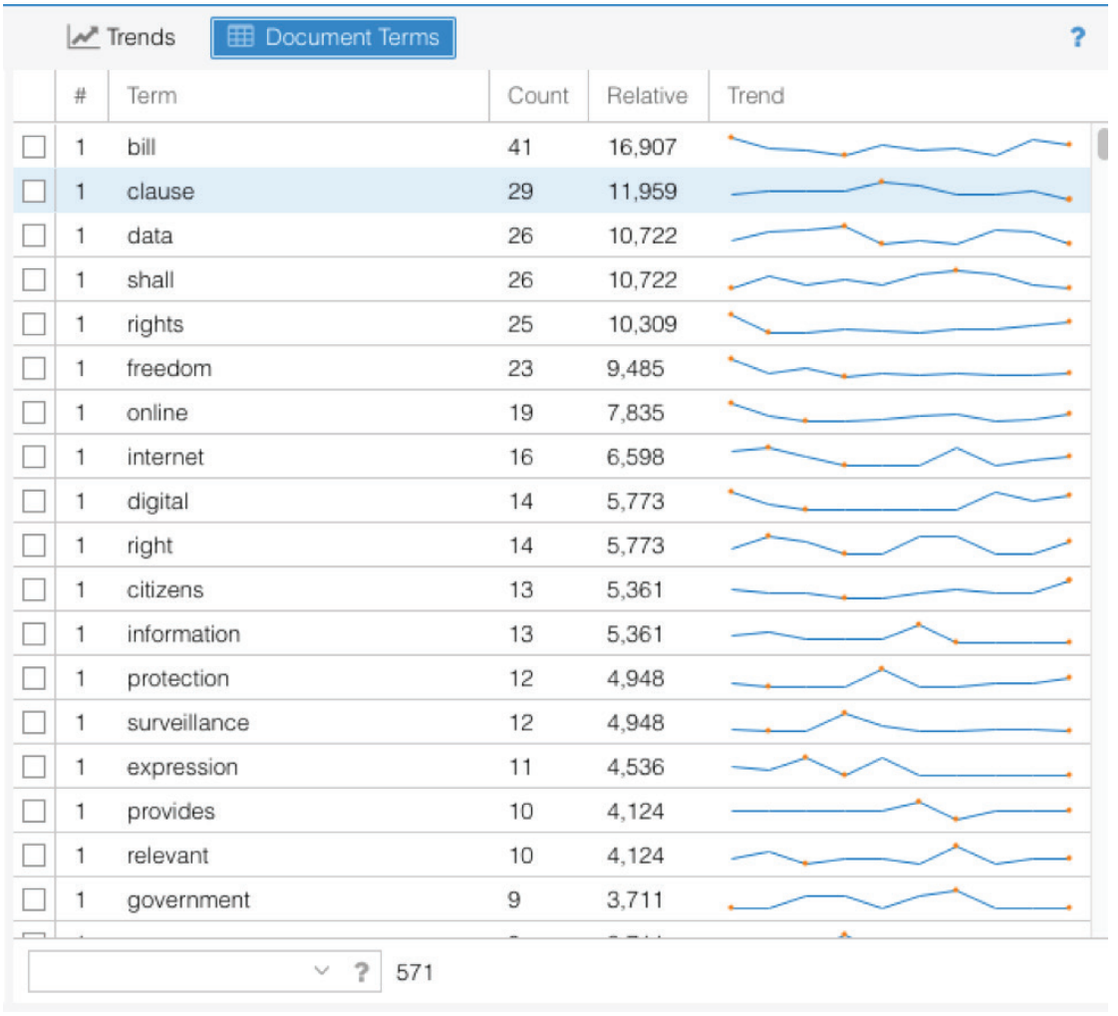
Go to: https://voyant-tools.org



Document Terms

Here, we are presented with several data points from the bill that offer some insights. We notice that 'bill', 'clause', 'data', 'shall', and 'rights' are the most frequently used keywords, and we can observe the document number—that is, the position of the term's document in the corpus.

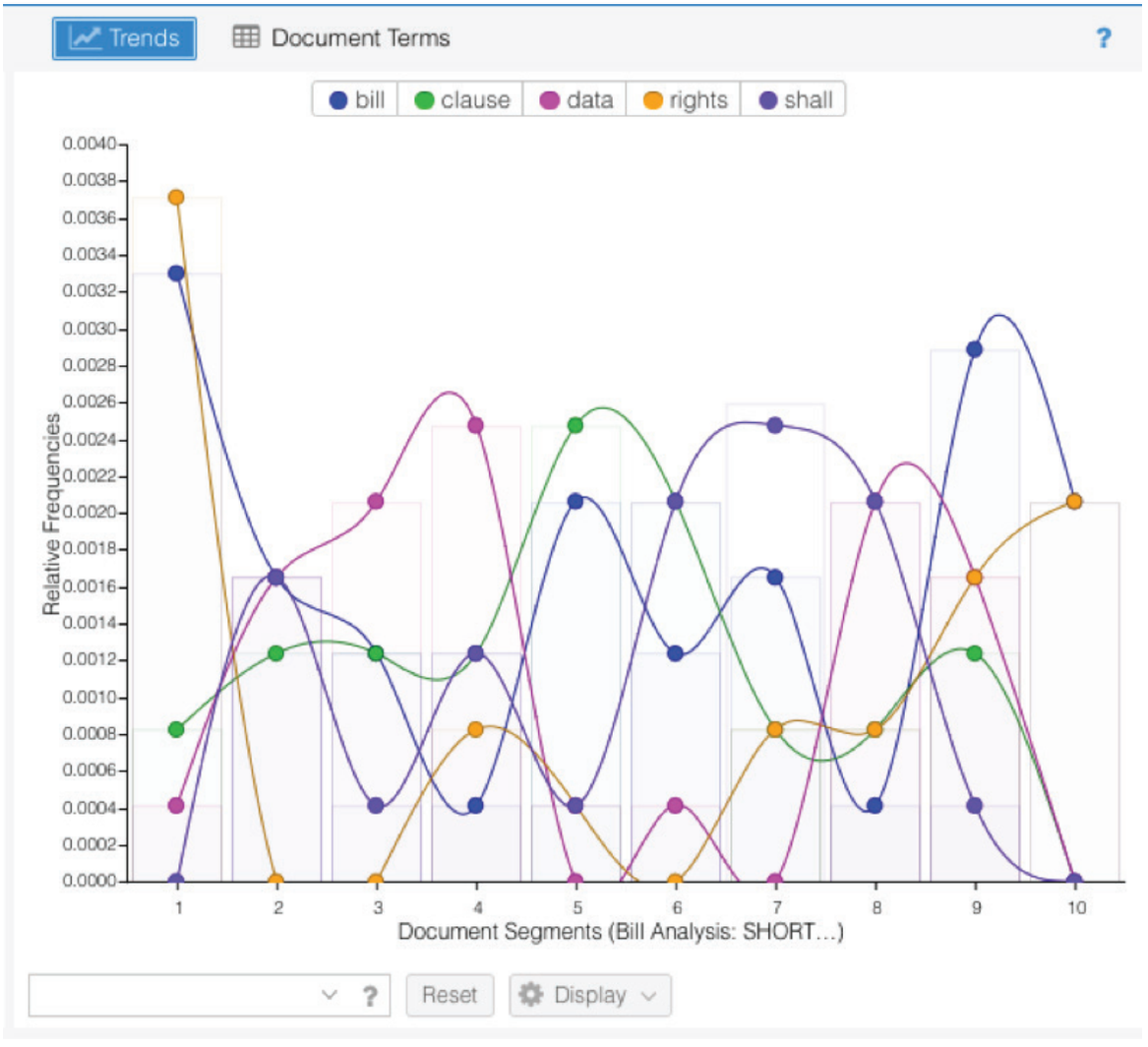
'Term' represents the document term used, 'Count' is the raw frequency of each term, 'Relative' is the relative frequency (per ten million words) of each term, and finally, 'Trend' is a sparkline graph that illustrates each term's distribution within the document segments; by hovering over the sparkline, one can view finer-grained results.



Trends

Similar to document terms, we shall analyse the trends, which present a line graph depicting the distribution of a word's occurrence across a corpus or document depending on the mode.

Each series in the graph is coloured according to the word it represents, and at the top of the graph a legend displays which words are associated with which colours. Advocates can click on words in the legend to toggle their visibility. Hovering over any point on the graph causes a callout box to appear with information about the point, including the word, the frequency (raw or relative depending on the mode), and the document or document segment.



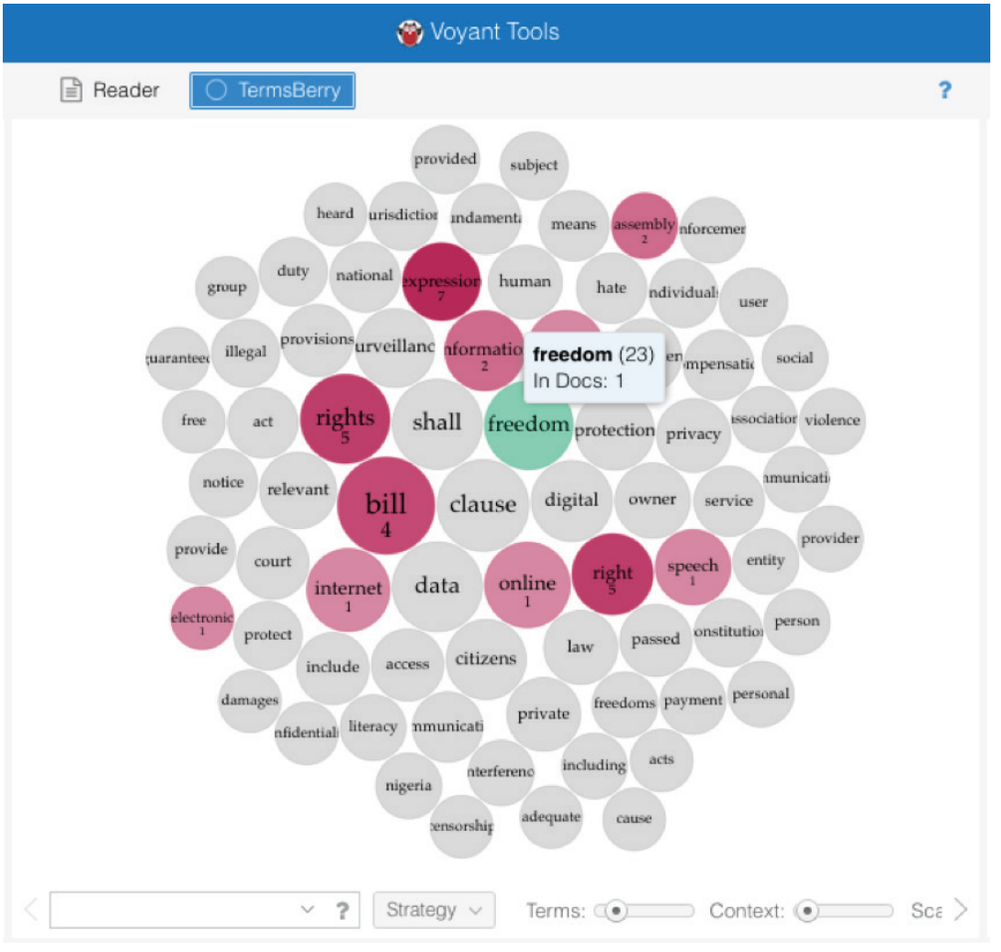
TermsBerry

The TermsBerry tool provides a method for exploring high-frequency terms and their collocates (i.e., words that occur in proximity) that are used in the bill.

The TermsBerry tool is intended to mix the power of visualising high-frequency terms with the utility of exploring how those same terms co-occur—that is, to what extent they appear in proximity with one another. In some ways, this tool is similar to Cirrus (the word cloud) although even more useful due to the added collocates and corpus coverage information.

The highest-frequency terms (or most distinct terms if you change the options) appear towards the middle of the screen and in larger bubbles spiralling outwards. The darkness of each term represents the proportion of documents wherein it appears; darker colours mean the term appears in more documents, and no differentiation is observed if there exists merely one document in the corpus.

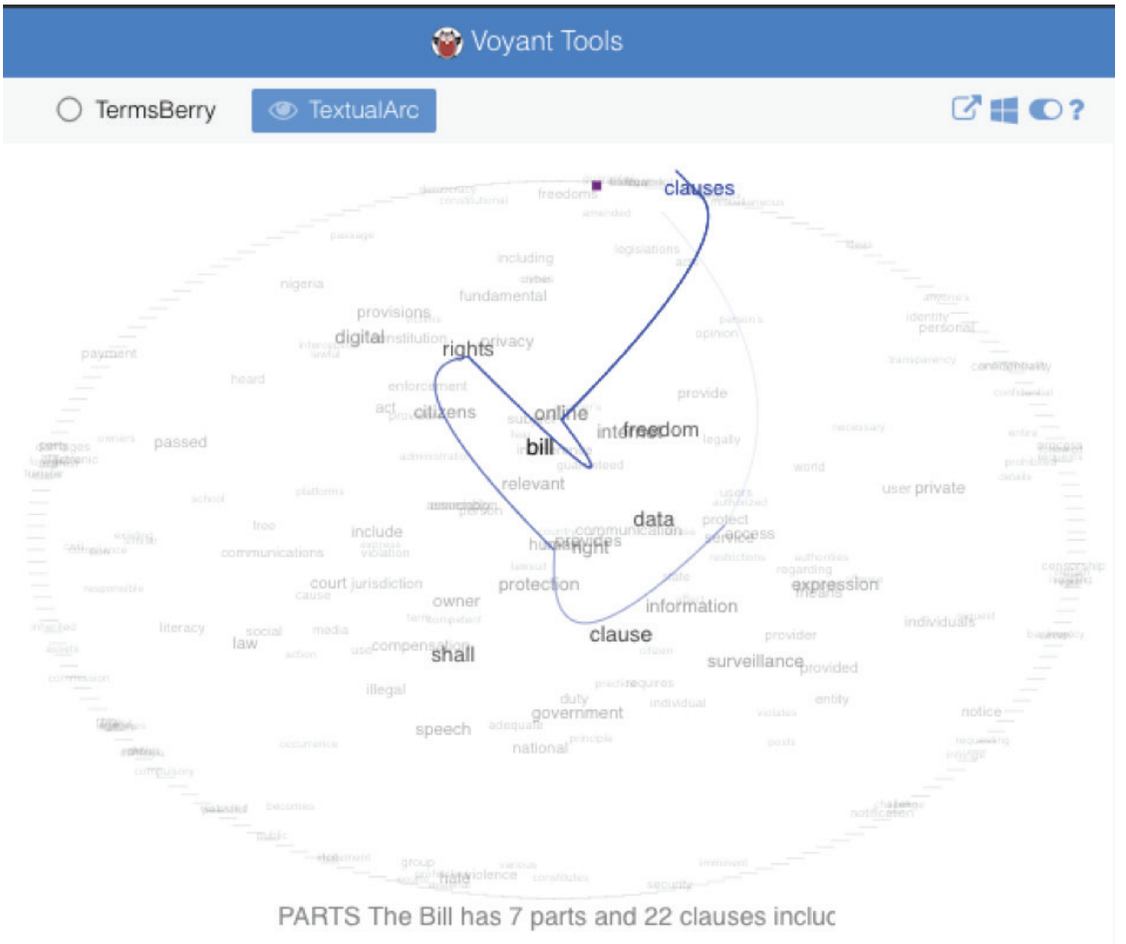
When you hover over a term, it becomes the keyword. Then, each of the other bubbles indicates the collocate frequency for that term (within the specified context; by default, two words to the left and two words to the right); the darker the colour, the higher the collocate frequency. The hovering term also has a tooltip that appears and then generates the term frequency as well as the number of documents in which that term appears.



TextualArc

Similar to TermsBerry, we look into TextualArc, which is a visualisation of the terms in a document that includes a weighted centroid of terms and an arc that follows the terms according to their order of appearance within the document.

The current text is represented on the circle's perimeter, starting at the top and looping around clockwise. Each occurrence of a term pulls the term towards its location on the perimeter, while the term label's position represents the mean of these forces (or weighted centroid). The text is 'read' from start to finish, with repeated non-stop words visited by the animated arc. The occurrences of the currently read term are indicated by lines to the perimeter. You can also hover over any term to see its occurrences on the perimeter.



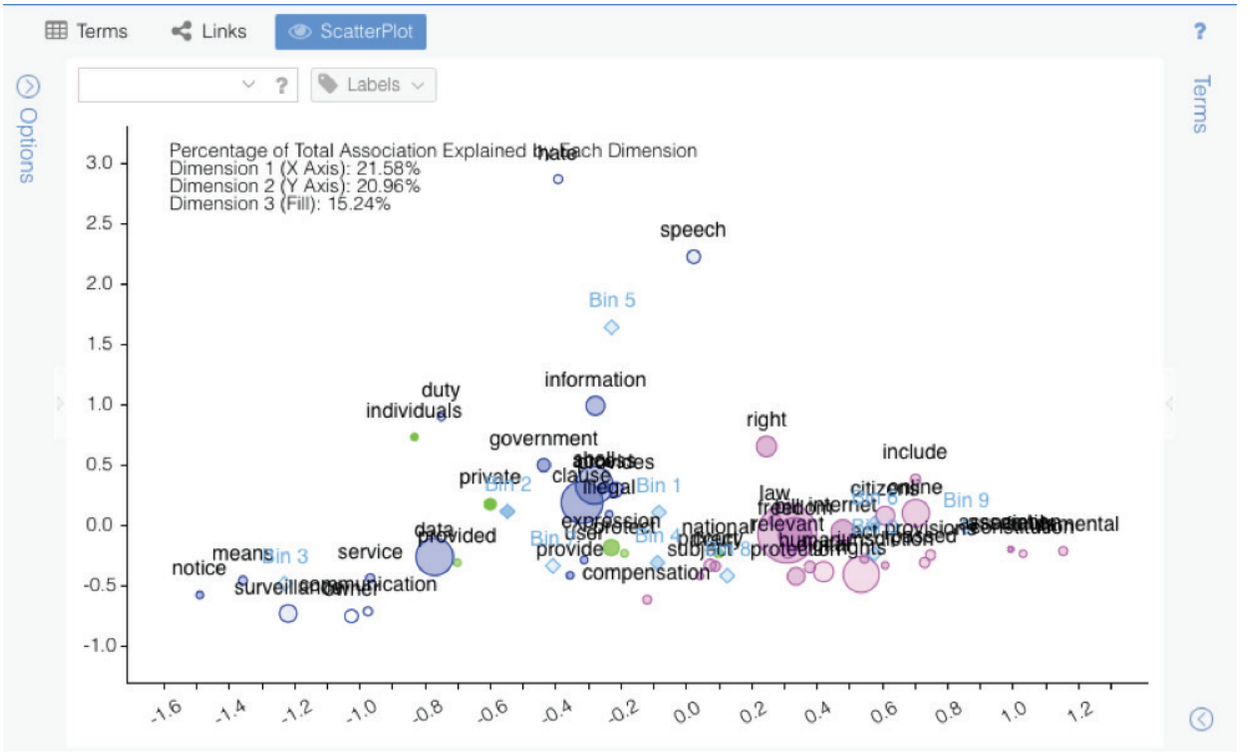
Scatter Plot

A scatter plot is a graph visualisation of how words cluster via a corpus document similarity, correspondence analysis, or principal component analysis (PCA).

PCA is a technique that takes and optimises data in a multidimensional space, thus reducing its dimensions to a manageable subset. By discarding all but the first two or three dimensions, we are left with a new data set that ideally contains most of the original set's information that is easy to visualise. In the resulting visualisation, words used in the bill that are grouped together are associated—that is, they follow a similar usage pattern in the corpus.

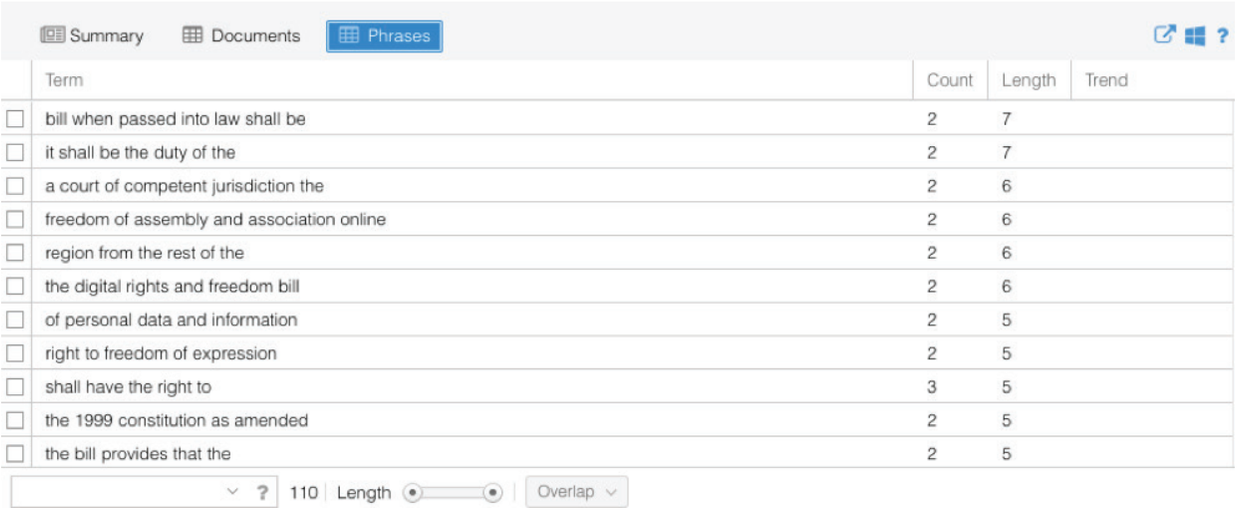
A scatter plot is presented in the tool's main display with a legend in the top left-hand corner, and hovering over a word in the graph will display additional information about that word's frequency of occurrence.

Above the main display is the primary toolbar, while to the right is a sub-panel of words that appear in the corpus as well as their frequencies.



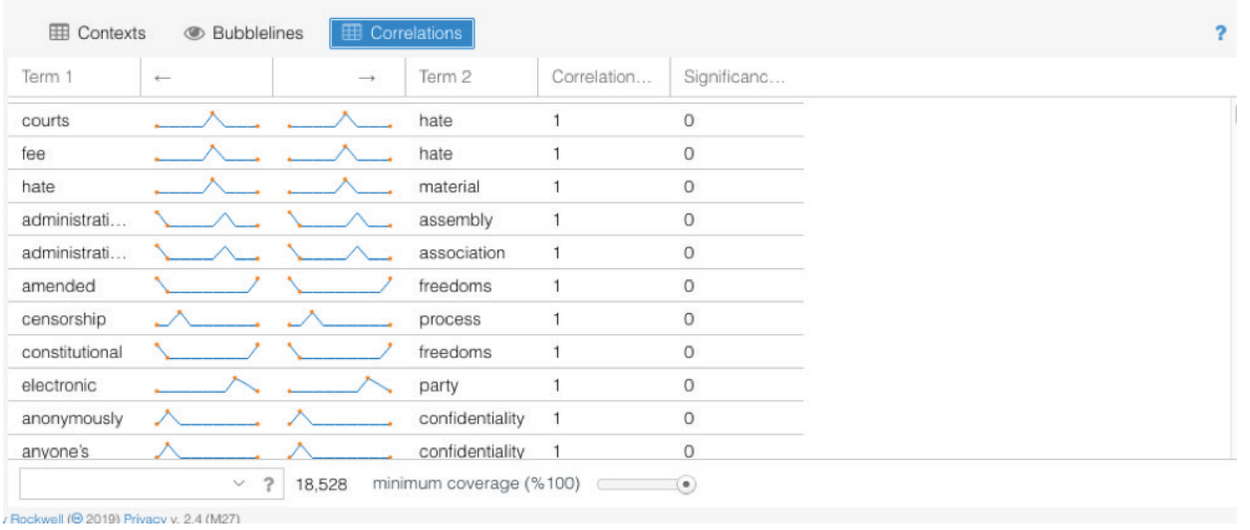
Phrases

The 'Phrases' tool indicates the bill's repeating sequences of words organised by repetition frequency or number of words in each repeated phrase. The lists below provide us with a strong indication of which key phrases describe the bill and may be analysed further.



Correlations

The 'Correlations' tool below enables an exploration of the extent to which the bill's term frequencies vary in sync (i.e., terms whose frequencies rise and fall together or inversely).



Advocates are encouraged to explore other tools against the same bill's data set such that various stakeholders may determine how to more thoroughly understand the Digital Rights and Freedom Bill. To access other tools, see:

https://voyant-tools.org/docs/#!/guide/tools

CONCLUSION

The design, implementation, and evaluation of development programs towards digital rights advocacy is intrinsically about people, organisations, institutions, and the relationships between them. From the case studies above, we have demonstrated that SNA can help one understand stakeholder relationships and answer questions such as: who is the most critical for sharing, spreading, or disseminating information? Who is the most influential or connected to the greatest number of individuals? Who acts as a bridge between different parts of the network?

We have also come to understand that SNA conceptualises a policy-making process as a network of actors. It can assess whether or not an interest group occupies a leading central position within this policy network and whether it belongs to various ad-hoc coalitions or plays a brokering role between different stakeholders. Such network variables are crucial for capturing how the various interest groups mobilise and gain access to policymakers as well as for explaining their goal achievements and potential policy influence.

The benefit of an approach such as SNA is that its methods can describe both simple and complex environments—from a local community to a nation state or even further to an international level. These methods may also range from an analogue (i.e., in-person mapping exercises through workshops/ conferences) to the most sophisticated (quantitative) measures of individuals and the network using sophisticated software programs and tools.

In conclusion, we hereby address the final factors advocates should consider for SNA's successful introduction in their programs:

- Much consideration should be expended towards learning questions about and 'use cases' of SNA. Advocates might choose to use SNA for a specific digital rights program design to, for example, help design a technical approach by selecting the appropriate audience/target beneficiaries. SNA may additionally be employed for digital rights program implementation to, for instance, help improve organisational capacity.

- SNA may be implemented in program monitoring and evaluation to measure relationship changes over time, generate base- and end-line data to understand what interventions contribute, and provide program learning information that adaptively manages their programs as they are implemented.

- It is critical that advocates clarify the types of questions they feel are most critical to their programs and how they plan to inform their interventions according to their SNA findings. Advocates should consider how they will ethically employ the information for program design and implementation, including monitoring and evaluation for learning purposes.

- The program context should influence the exact SNA approach and method that advocates will be considering; in many instances, simpler is better. Once they have determined their use cases, advocates should select the approaches,

methods, and tools that will capture the amount and type of data they can easily collect and analyse. A variety of techniques are available, ranging from low-cost analogue and collaboration mapping to more sophisticated analyses that use metrics, such as free or open-source platforms (e.g., Netvizz, Gephi, NodeXL, SocnetV, R). Their choice of platform will likely depend on the nature, frequency, and type of data they plan to collect.

- It is important that advocates understand that some digital rights programs will benefit from SNA more than will others. SNA can be applied across a range of sectors, such as politics, social science, economic growth, and governance, although it is more easily integrated in programs focused on network-level or system-level change. Interventions wherein SNA may be the most useful include external advocacy, policy, communications, messaging, and programming focused on organisational development.

- As stated within this handbook's introduction and as exemplified by the Facebook–Cambridge Analytica scandal in early 2018, security, privacy, and ethical considerations are paramount. SNA data pose specific privacy concerns that other types of statistical survey data do not; for instance, data about relationships may generally be sensitive, and additional analyses may reveal wider connections that may also be sensitive. Although they can be aggregated to a higher level, SNA data are often difficult to anonymise. It is therefore critical that efficient data practices are ensured by including fully informed consent and an opt-out option when collecting potentially sensitive data. Advocates should also be transparent about how they will collect, analyse, and report on their collected data. In addition, given their sensitive nature, maintaining appropriate data security collection, storage, and deletion practices is paramount. Before deciding to apply any SNA approach, advocates should assess the situation, the alternatives, and the implications for the participants to ensure they are taking their approaches in an appropriate and ethical way.

- It is necessary that data be supplemented, co-interpreted with context, and further analysed. SNA findings are only as good as their corresponding data. It is important to take stock of and interpret the findings, especially if the analysis is quantitative. For example, if advocates employ SNA to understand changes in a particular network over time, it is often critical that they always use the same hashtag across the network or interview the same people; this detail is often less important for larger quantitative surveys. It is important that any SNA analysis be supplemented with qualitative data to ensure that the understanding and interpretation of the data is accurate. Often, the most effective way to do so is to let people from within the network explain how they would interpret and utilise certain aspects of this information. The appropriate context for interpreting data is paramount and may help connect the dots.

SNA is a powerful tool. When wielded appropriately and ethically, it can be incredibly illuminating in helping advocates understand the contexts in which they work. Advocates should confidently refer to the knowledge shared in this handbook while building on their creativity and paving the path to success within their digital rights programs.

REFERENCES AND RESOURCES

Free Software!

- TAGS: a free Google Sheet template that lets users set up and run the automated collection of search results from Twitter: **https://tags.hawksey.info**
- NodeXL: the easiest visualisation tool I've used, but (1) it only works with Microsoft Excel in Windows and (2) its advanced features are no longer free. This tool also calculates some metrics: **https://nodexl.codeplex.com**
- Gephi: another nice (open-source) visualisation tool that works with Macs, Linux, and Windows alike: **https://gephi.org**
- Pajek: once the favourite visualisation tool in SNA—and still popular today: **http://vlado.fmf.uni-lj.si/pub/networks/pajek/**
- SoNIA: a visualisation software package for dynamic data (network movies!) **http://web.stanford.edu/group/sonia**
- Voyant Tools: a web-based text reading and analysis environment designed to make it easy for users to work with their own text or collection of texts in a variety of formats, including plain text, HTML, XML, PDF, RTF, and MS Word: **https://voyant-tools.org/**

Aral, Sinan, Lev Muchnik, and Arun Sundararajan. Engineering social contagions: Optimal network seeding in the presence of homophily. *Network Science,* pp. 0–43, 2013.

Cranmer, Skyler J. and Bruce A. Desmarais. Inferential network analysis with exponential random graph models. *Political Analysis,* 19(1), p. 66, 2011. http://people.cs.umass.edu/~wallach/courses/s11/cmpsci791ss/readings/ cranmer11inferential.pdf.

Denny, Matthew J. Graph *compartmentalization,* 2014.

Gomez Rodriguez, Manuel, Jure Leskovec, and Andreas Krause. Inferring networks of diffusion and influence. *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,* p. 1019, 2010. http://dl.acm.org/citation.cfm?doid=1835804.1835933

Goodreau, Steven M., James A. Kitts, and Martina Morris. Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks. *Demography,* 46(1), pp. 103–125, 2009. http://link.springer.com/ article/10.1353/dem.0.0045

Hoff, Peter D., Adrian E. Raftery, and Mark S. Handcock. Latent space approaches to social network analysis. *Journal of the American Statistical Association,* 97(460), pp. 1090–1098, 2002. http://www.tandfonline.com/doi/abs/10.1198/ 016214502388618906

International Rescue Committee. *Social network analysis handbook: connecting the dots in humanitarian programs,* 2016. https://www.rescue.org/resource/social-network-analysis-handbook-connecting-dots-humanitarian-programs

Digital Rights Advocates Social Network Analysis Handbook

Prepared by Richard Ngamita

This report was funded by Internews 2019

internews.org